A CASE OF INTENSIVE LEXICAL DIFFUSION: ARNHEM LAND, AUSTRALIA

JEFFREY HEATH

Harvard University

The languages of Arnhem Land, Australia, form a contiguous block and Sprachbund, despite a sharp genetic boundary. Thus in Ngandi and Ritharngu, which belong to different major genetic divisions, about half the total vocabulary is shared, virtually all because of recent diffusion. 'Core' vocabulary and highly inflected stems (verbs) are among the borrowed items, but linguistic structural factors affect the extent of diffusion. Extensive over-all borrowing is attributed to local demographic and cultural features: small size of ethnolinguistic group, absence of strong tribal organization, and extensive cross-linguistic marriage.*

1. BACKGROUND. In eastern Arnhem Land, along the coast of the Gulf of Carpentaria, several ethnolinguistic groups (linguistic 'tribes') were in close contact with each other prior to the recent intrusion of whites. I contend that some of these languages came into contact at a fairly recent time (let us say a thousand years ago), and that lexical diffusion among them has reached a level not reported elsewhere for stable multilingual zones. The plausibility of this claim is increased by my previous documentation (Heath 1978a) of extensive phonological and grammatical diffusion involving the same languages. In addition to structural convergence, involving inherited materials, I showed that direct borrowing of many bound affixes had occurred: noun-class prefix systems, negative suffix, denominative verbalizing suffix, pre-inflectional thematicizing verbal suffix, comitative affixes (with nouns and verbs), several case suffixes (ergative-instrumental, genitive-dative, ablative, instrumental), dyadic dual suffix with kin terms, and various others.

Before proceeding, I will review the evidence for the genetic relationships posited among the languages. The crucial division is between the Yuulngu group, bottled up in northeastern Arnhem Land, and the so-called 'prefixing' languages which surround the Yuu. group to the south and west. The prefixing languages, or at least those involved in this paper, are a genetic group, but the fine-grained subgrouping has not been worked out.

The Yuu. group consists of approximately ten languages; and except for one or two Yuu. offshoots in the west which do not concern us, the group is internally quite homogeneous. That the Yuu. languages are closely related has been obvious on lexical and grammatical grounds to all observers, and some of the 'languages' grade into each other in dialect-chain fashion. Adjoining Yuu. languages, or at least those involved in this paper, are a genetic group, but the fine-grained subgrouping has not been worked out.

The Yuu. group consists of approximately ten languages; and except for one or two Yuu. offshoots in the west which do not concern us, the group is internally quite homogeneous. That the Yuu. languages are closely related has been obvious on lexical and grammatical grounds to all observers, and some of the 'languages' grade into each other in dialect-chain fashion. Adjoining Yuu. languages, or at least those involved in this paper, are a genetic group, but the fine-grained subgrouping has not been worked out.

Before proceeding, I will review the evidence for the genetic relationships posited among the languages. The crucial division is between the Yuulngu group, bottled up in northeastern Arnhem Land, and the so-called 'prefixing' languages which surround the Yuu. group to the south and west. The prefixing languages, or at least those involved in this paper, are a genetic group, but the fine-grained subgrouping has not been worked out.

The Yuu. group consists of approximately ten languages; and except for one or two Yuu. offshoots in the west which do not concern us, the group is internally quite homogeneous. That the Yuu. languages are closely related has been obvious on lexical and grammatical grounds to all observers, and some of the 'languages' grade into each other in dialect-chain fashion. Adjoining Yuu. languages, or at least those involved in this paper, are a genetic group, but the fine-grained subgrouping has not been worked out.

The Yuu. group consists of approximately ten languages; and except for one or two Yuu. offshoots in the west which do not concern us, the group is internally quite homogeneous. That the Yuu. languages are closely related has been obvious on lexical and grammatical grounds to all observers, and some of the 'languages' grade into each other in dialect-chain fashion. Adjoining Yuu. languages, or at least those involved in this paper, are a genetic group, but the fine-grained subgrouping has not been worked out.

The Yuu. group consists of approximately ten languages; and except for one or two Yuu. offshoots in the west which do not concern us, the group is internally quite homogeneous. That the Yuu. languages are closely related has been obvious on lexical and grammatical grounds to all observers, and some of the 'languages' grade into each other in dialect-chain fashion. Adjoining Yuu. languages, or at least those involved in this paper, are a genetic group, but the fine-grained subgrouping has not been worked out.

The wider genetic connections of the Yuu. languages are not to the adjoining prefixing languages, but rather to languages far to the south—in the southern

* Fieldwork on several languages was supported by the Australian Institute of Aboriginal Studies between 1973 and 1976. I am indebted to Paul Black and Ken Hale for comments on an earlier draft.
part of the Northern Territory and adjoining areas. This conclusion was drawn on lexicostatistical grounds by O’Grady et al. 1966; and since the groups in question are not contiguous, the lexical sharings cannot be explained as diffusional. For example, compare Warramunga wayga- ‘to speak’ with Yuu. wayga- ‘to speak’, and Warr. wanja ‘which?, where?’ with Yuu. wana ‘where?’ Here we see a regular correspondence of Warr. homorganic nasal + stop cluster to Yuu. simple nasal; the correspondence of alveo-palatal ñ to (laminal) interdental ñ is no problem (Dixon 1970). Further striking similarities occur in stance verbs (also used as auxiliaries): Warr. yu- (fut. yu:-nu), Yuu. yu:ra- (fut. yu:ri) ‘to lie down, to sleep’; Warr. ja- (fut. ja:-ni), Yuu. da:ra- (fut. da:ri) ‘to stand’; Warr. ni- (fut. ni:-ni), Yuu. ni:na- (fut. ni:n-i) ‘to sit’. Cognate verb forms normally occur in inflectional classes which are themselves historically related (see Dixon 1980 on conjugation markers and inflectional endings). The future tense forms shown above are exact matches, except for progressive assimilation in Warr. yu:-nu (*yu:-ni) and some interchange of *-r- and *-n- conjugation markers (still a productive process in Warramunga). Even from these few examples, we can see regular sound correspondences beginning to emerge; I may add that the Warr. stance verbs have present tense suffix -nda (e.g. ni-nda ‘sits’) matching Yuu. forms despite distinct morpheme boundaries (e.g. ni:na-∅ ‘sits’), with nd vs. n paralleling ng vs. η and nj vs. n in examples above. Since the connection between Yuu. languages and the geographically non-contiguous southern languages (exemplified here by Warramunga) represents the consensus of Australianists with opinions on the matter, it can be taken for granted in this paper.\(^1\)

The Yuu. languages and their southern relatives are part of the Pama-Nyungan family (O’Grady et al.), which covers about 90% of Australia. The remaining 10% is occupied by the prefixing languages, of which only a small number are relevant here. Considerable genetic diversity exists within this group, and it may eventually be divided into two or more groups roughly coordinate to Pama-Nyungan. Fortunately, the crucial background point is simply that the prefixing languages are not genetically close to the Yuu. languages; this point is not controversial at all.

Two subgroups of prefixing languages are relevant here. Most important is a group containing Ngandi (A1), Nunggubuyu (A2), and Anindhilyagwa (A3). At first glance, the three differ sufficiently in phonology and morphology to make such a close genetic relationship questionable. However, further analysis

---

\(^1\) Human biological evidence can be appropriately applied in contexts like this one. Selected populations are tested for various genetically-coded traits, and correlation matrices are computed for each pair of populations to determine relative biological distances. (For a general discussion in the Australian context, see Balakrishnan et al. 1975.) More concretely relevant to this paper is White & Parsons 1973, showing that Yuulngu (their ‘Murngin’) people have more biological affinity to groups well to the south, speaking other Pama-Nyungan languages (represented by Aranda), than do speakers of certain prefixing languages in the Arnhem Land area (pp. 9–10). However, human biological results cannot be expected to match historical linguistic classifications in fine detail; extensive localized gene flows across the Yuulngu/prefixing linguistic boundary have obviously occurred, and nearby prefixing languages such as Ngandi undoubtedly now show substantial biological affinities to Yuu. groups.
A CASE OF INTENSIVE LEXICAL DIFFUSION

shows that the complex morphological structures of all three can be derived from a single proto-system, assuming regular sound changes and reasonable analogical changes. Since I am concerned mainly with A1 and A2 here, I will set A3 aside temporarily; but it can be shown, using similar arguments, that A3 separated from Proto-A at about the same time as A1 and A2.

Sticking mainly to the complex verb paradigms, note the precise correspondence (morpheme for morpheme) of A1 yə-na-ŋa-ŋa-, A2 ya-ŋa-ŋa- ‘I saw him’ (continuous aspect) from *ŋə-ŋa-ŋa- (1sg.-3m.sg.-see-PastContinuous). In both languages -ŋu- is a special allomorph of the masculine 3sg. morpheme, used only as direct object marker in certain combinations. Similarly, in both languages ‘to sit’ belongs to a special, unproductive verb class with its own characteristic suffixes. Reconstructable forms include past punctual *-na-y (A1 -ŋa-ŋa-, A2 -na-ŋa-), past continuous *-na-ŋi (see above), present *-na-ŋi (A1 -ŋa-ŋa- ‘I saw ... see’, A2 -niŋ ‘will not see’) and *na-ŋa (A1 -na-ŋa- ‘would have seen’, A2 -na-yan ‘lest ... see’). In both languages, the other verbs in this unproductive inflectional class are *ŋa- ‘to hear’ (A1 -ŋa-ŋa-, A2 -yaŋa- from compound *yaŋ-ya- ‘to hear voice’), *ŋu- ‘to eat’ (A1 -ŋu-ŋa-, A2 -u-), and *-wo- ‘to give’ (A1 -wo-, A2 -u/-i). The languages show various phonological changes (and a small number of analogical restructurings and semantic shifts); but these are to be expected in languages whose connection is mainly genetic rather than diffusional. We can cite similar detailed lexical and affixal similarities involving nearly all verb classes. For example, both languages have a verb class whose suffixes are essentially identical to those of ‘to see’ (above) but involve an augment *n- between stem and suffix in certain forms. In both languages, the past continuous (expected *-n-ŋa) is irregularly *-n-ŋi (A1 A2 -n-ŋi), while the present is expectably *-n-ŋi (A1 -n-ŋi, A2 -n-ŋi). In both languages, this class consists mainly of *-ga- ‘to carry’ and compounds based on it (-ga- by itself survives in A1, but not A2).

Since many complete verb forms can be reconstructed for the ancestor of A1 and A2 (some also with direct matches in A3), there is no question that Proto-A was a moderately recent proto-language, perhaps roughly contemporaneous to Proto-Yuulngu. I find no close matches between reconstructed Proto-A and Proto-Yuu paradigms (but cf. below). However, just south of the A group is a second group of prefixing languages which shows affinities to it. It consists of Warndarang (B1), Mara (B2), Alawa (B3), and perhaps other languages farther west.

Proto-B is more difficult to reconstruct than Proto-A; and since the B group is of limited importance here, I will be brief. To the Proto-A forms for ‘to see’ cited above, we may compare B2 -na-, durative stem -mi-na-. Past and future continuous -mi-na- and present -mi-na-ŋi nicely match Proto-A *-na-ŋi, *-na-ŋi, and point to a Proto-AB paradigm quite distinct from Proto-Yuulngu (or Proto-Pama-Nyungan). B1 lacks this stem, but does show -wa- ‘to give’ (Proto-A *-wo-), past irrealis -wa-ŋa (Proto-A *-wo-ŋa). It is difficult to reconstruct complete forms for Proto-B, mainly because the B languages appear to have separately reduced their stock of directly inflectable verbs, shifting toward
more complex auxiliary constructions. In doing so they happen to have retained different sets of Proto-B verb stems; thus a stem in a given B language must sometimes be interpreted historically by going outside the B group (e.g. to the A group) for cognates.

Perhaps the best evidence for relating the A and B groups comes from analysis of the pronominal prefix systems. Here we find good, substantive matches involving all A and B languages, partly disguised by analogical re-structuring. For example, B₃ shows reflexes of *-n- and *-gu- as allomorphs of a morpheme indicating that the preceding pronominal morpheme is direct object, with *-n- after vowel and *-gu- after consonant. B₁ and B₂ show traces of this system with some adjustments. A₁ and A₂ have developed the system into Algonquian-type direct/inverse systems, with A₁ generalizing *-gu- and A₂ generalizing *-n-. Thus we can show that Proto-A and Proto-B must have had very similar pronominal prefix systems, involving a large number of identical or near-identical morphemes in the same complex configurations. (For a more detailed demonstration, see Heath 1976.) So we have no reason to doubt that the Yuulngu and prefixing languages are distinctive genetic units; remaining questions involve fine-grained subgrouping, and are largely irrelevant here.²

At a very remote time depth, it is probable that the Yuulngu (and other Pama-Nyungan) languages are genetically related to the prefixing languages. Certain archaic elements in the prefixing groups, such as pronominal elements like 1sg.*qa-, old case endings *-n- and *-gu- (now frozen in the pronominal prefix systems), and some irregular monosyllabic verbs like *-bu- ‘to hit’, *-na- ‘to see’, *-wa- ‘to give’ have cognates in several Pama-Nyungan languages

² Classification of A₃ has previously been problematic because of its high lexical divergence, but this resulted mainly from the absence of extensive published material on the morphology. Late in my fieldwork, I obtained sufficient morphological data to permit recognition of a close genetic relationship with A₂ and A₁. Since A₃ does not figure prominently in this paper, I omit an extended discussion of its genetic relationship. However, the fact that it is morphologically quite obviously derived from Proto-A, but now shares relatively few lexical items with A₁, A₂—or, indeed, any other language—points to a problem in the historical study of lexicon distinct from that of diffusion dealt with here: namely, the question why certain Australian languages seem to show highly accelerated lexical (and grammatical) turn-over, while others remain more stable. A₃ also seems to have undergone various grammatical changes at a rapid rate, over and above those directly attributable to diffusion (e.g. from B₁), and its lexicon is rather idiosyncratic (though not so much as that of A₁). Since A₃, spoken on two islands, and Tiwi (another Australian language spoken on islands off the Top End of the Northern Territory) are noted for lexical divergences, we must face the question of effects which physical location on islands (and peninsulas) may have on lexical replacement (including internal, non-diffusional processes). It is possible that this ties in with the study of diffusion: inland languages are obviously subject to more diffusional pressures from several directions than insular or peninsular ones, and diffusional reinforcement of already shared vocabulary may inhibit lexical turn-over. However, the problem requires a separate study involving data which are not yet fully available.

Descriptive data used in this paper are chiefly from my publications in the references and from a forthcoming dictionary of Nunggubuyu (A₂). Of other languages mentioned, grammatical studies are available for Rembangga (McKay 1975) and Alawa (Sharpe 1972). I have a limited amount of unpublished material on Dhay’yi (Y₃) and Ngalakan. More substantial investigation of Ngalakan and Mangarayi has been undertaken by Francesca Merlan, but her results are not available at the time of this writing.
which cannot be easily attributed to borrowing. As yet we are not able to
reconstruct much in the way of complete words; but work in progress on verb
systematics (particularly by R. M. W. Dixon) may well lead to at least a core
of reconstructible formations for Proto-Australian. What we have in Arnhem
Land is therefore roughly similar to the Sprachbund in the Balkans, where
several Indo-European languages from highly diverse subgroups (Romance,
Slavic, Greek, Albanian) have come together in comparatively recent times.
In general, it is not difficult in such regions to distinguish those sharings caused
by independent retention from those resulting from recent diffusion.

The genetic groupings of the languages dealt with here are shown in Figure
1. Note that Y₁, Y₂ etc. are the symbols used for Yuu. languages, in contrast
to A and B for prefixing languages. Mnemonically, note that A and B are closer
to each other than either is to Y. However, the spatial arrangement of the
languages is shown schematically in Figure 2. The most significant pair of
languages is Y₁ and A₁, which are far removed from each other genetically,
but are contiguous and have undergone very extensive diffusional interaction.
A₂ and B₁ are another diffusional pair, but of somewhat less significance to us.
A₁ and A₂, though genetically close, have tended to diverge from each other,
partly because of these different regional influences. The effects of genetic and
spatial relationships on lexical diffusion and sharing are dealt with in the next
several sections (each focusing on a particular kind of vocabulary). I will then
provide basic sociolinguistic information, and suggest ways in which this case
study can be used in more general theoretical studies of lexical diffusion.

2. FLORA AND FAUNA NOUNS. Most of the local flora and fauna are endemic
to Arnhem Land rather than pan-Australian. Moreover, the areas occupied by
the relevant ethnolinguistic groups were closely bunched geographically, and
there was substantial mobility across ‘tribal’ boundaries. We can therefore
expect a substantial amount of diffused, regional flora-fauna vocabulary.

However, we face technical problems in generating meaningful statistics for
linguistic sharing. Species named in one language may be without names in others—because of local environmental variation, differential cultural elaboration, or gaps in my data. Coastal groups like A2 and B1 naturally have many more terms for marine, estuarine, and littoral species than do inland groups like A1. Finally there is the question of delimiting the domain of flora-fauna terminology for my purposes.

To evaluate A1–Y1 and A1–A2 sharings, I will construct a sharing index based on lexical lists from A1. This choice is reasonable since A2 and Y1 have lexical items for most of the glosses represented in the A1 list (A1 is an inland language, while A2 and Y1 were spoken on the coast and in adjacent inland areas; the data are more complete for A2 and Y1).

For the flora I will deal only with terms for trees and shrubs, since terms for smaller plants (including yams and water lilies) pose problems of demarcating flora terms from body-part and other kinds of terms. A list of 106 A1 tree/shrub terms (Heath 1978b:171–2) will be used.

By a conservative count, sixty-nine (65%) of these terms also occur in Y1, while forty-seven (44%) occur in A2. Thirty-six (34%) of the A1 items occur in both Y1 and A2. The A1–Y1 matches almost all show identical consonants; the vocalic differences are usually minor, resulting from differences in vowel systems. Similarly, almost all show exact or near-exact semantic identity. I am unaware of a single item here for which a plausible Proto-Australian etymology can be constructed, and I regard all the A1–Y1 matches as caused by recent diffusion.

In A1–A2 matches, it is much more difficult to determine whether the forms are retentions from Proto-A or recently diffused regional terms. Perhaps in some cases the term was in Proto-A, but continuing diffusional interaction prevented regular sound changes (especially in A2) from occurring. In some cases, we can be fairly sure that the forms are from Proto-A via regular sound
changes (retentionist explanation); in others, some type of diffusional interaction is fairly certain; in still others, we cannot decide.

I omit extensive lists of matching sets. In my volume on A1 (Heath 1978b) I mention A2 and Y1 matches for A1 lexical items, so complete documentation is not necessary here. Some examples are: A1 Y1 dulgu vs. A2 lulwu ‘riverine form of paperbark, Melaleuca leucadendron’ (lulwu from Proto-A *dulgu via regular changes); A1 and Y1 dalpi? vs. A2 lalbij ‘short fan-palm, Livistona humilis’ (lalbij from Proto-A *dalpi? via regular changes); A1 duku? vs. Y1 duku? ‘craggy sand-dune form of paperbark, Melaleuca leucadendron’ (vs. unrelated A2 midi); A1 Y1 banar? vs. A2 banar ‘marble tree, Owenia vernicosa’ (Proto-A *banar? should give A2 *wanar, so initial b in A2 suggests recent diffusion of lexical item, or diffusional interference with regular lenition process for this item). Note that the phonological and semantic correspondences are highly precise.

The area of fauna includes several small subdomains (kangaroos and wallabies, snakes, birds, insects etc.) I will simply take insects here as an illustration. Omitting larvae, there are forty A1 terms (Heath 1978b:173). Y1 and A2 each share fourteen (35%), though only four terms are found in all three languages. As in the tree/shrub case, the A1–Y1 matches show phonological and semantic (near-)identity, strongly suggesting diffusion; the A1–A2 matches are in several cases compatible with diffusional explanations, but in other cases show sound changes suggesting retention from Proto-A (as in A1 ja?, A2 ya:g ‘meat ant’).

That the extent of A1–Y1 diffusional sharings (50%) is high even by normal standards in this part of Australia is suggested by comparing A2 and B1. As we have seen, these two have a fairly distant genetic relationship combined with an important recent diffusional relationship (e.g., B1 has borrowed two case suffixes and some noun-class prefixes from A2 or its recent ancestors). Despite this, flora-fauna lexical sharings between A2 and B1 are substantially fewer than those between A1 and Y1. In a basic list of 103 B1 tree/shrub terms, thirty-one items are shared by A2, for an A2–B1 sharing index of 30%. However, within the B group there is a B1–B2 sharing index of 73%. The A2–B1 figure of 30% is unimpressive, since half these sharings are widespread regional terms, found also in A1 and usually in other non-contiguous languages like Y1. Since almost all the B1 terms have glosses represented in A2 vocabulary (both are coastal, and the two groups had close relations), the low A2–B1 figure cannot be explained away as aberrant. Rather, it shows that lexical diffusion between these two has been more limited than that between A1 and Y1. (Figures for A2–B1 fauna vocabulary are comparable to those for flora just given.)

Although details on all subdomains of flora and fauna cannot be given here, in general the A1–Y1 index is approximately 50%, A1–A2 35%, and A2–B1 30%. Such pairs as Y1–A2 and Y1–B1, involving non-contiguous groups or groups with weak social relations, are of course lower and involve widely distributed regional vocabulary.

Note again that these figures are based on lists of A1 (or B1) terms; using other methods, different figures can be obtained. I have excluded life-form
terms like ‘snake’ and ‘tree’ and body-part terms like ‘immature root corm of water lily, Nymphaea ?gigantea’. By these criteria, we have 600 to 1400 items per language in my data (B₁ is low because fieldwork was brief, and involved the last good speaker.)

It may be useful to compare these findings with those of Newman 1974, who deals with lexical sharings in ‘ecological vocabulary’ among several Northwest Coast American Indian languages. His items are mostly flora-fauna terms, but also include a few like ‘fishhook’, ‘fin’, and ‘fat’. He found that Bella Coola, physically separated from genetically related Salishan languages, shares a maximum of 16% of this vocabulary with any single other Salishan language, and that ecological sharings are smaller than sharings of ‘basic’ lexicostatistic vocabulary for the same pairs. Noting that the Bella Coola have had close relations (including much intermarriage) with tribes of the Athabaskan and Wakashan families, Newman argues that Athabaskan ecological borrowings are negligible, but that ‘the ecological lexicon of Bella Coola has been massively stocked with Wakashan words’ (206). Bella Bella (Wakashan) has the highest percentage of sharings with Bella Coola, i.e. 30%. Newman believes that most of these items were borrowed from Bella Bella into Bella Coola (not vice versa).

Newman’s statistics differ from mine in several ways. His initial list of Bella Coola ecological vocabulary consists of only 279 entries; my lists average 1000 per language, even after excluding some types of vocabulary admitted by Newman. Second, his indices are based only on glosses for which terms are recorded in all relevant languages; I have not done this, though I have approximated it by taking A₁ and B₁ (with comparatively small recorded vocabularies) as basic.

Despite these differences, it appears to me that Newman’s figures and mine can be usefully compared. Newman suggests that borrowing into Bella Coola from Wakashan languages was unusually high in the American Indian context (recall that Wakashan borrowed little from Bella Coola, and Bella Coola and Athabaskan engaged in virtually no diffusion). Yet Newman’s percentages for the maximal case (Bella Coola and Bella Bella) are significantly lower than my figures for A₁–Y₁ (30% vs. 50%). Newman states further that ‘Bella Coola borrowed almost nothing of its basic stock of words from Wakashan’ (207); however, we shall see below that many A₁–Y₁ borrowings of such vocabulary have taken place. In comparative perspective, we can see that A₁–Y₁ diffusion has been unusually extensive, and that even the more modest figures for A₂–B₁ are fairly high.

3. PHYSICAL-FEATURE NOUNS. Let us now consider topographic terms like ‘hill’, ‘billabong’ (pond) etc., and terms for celestial bodies. I begin by taking a list of fifteen glosses representing important lexical categories in these lan-

---

3 A general discussion of the structure of life-form terminology for A₂ is given in Heath 1978c. Roughly, the basic life-form terms for A₁ mean ‘wood, tree’; ‘grass’ (two synonyms); ‘fish’; ‘honey (bee)’; and ‘snake’. None of these has cognates in A₂, showing that this vocabulary is not especially archaic. There are two sharings between A₁ and Y₁: A₁ and Y₁ mulmu ‘grass’ (one of two synonyms in A₁); and A₁ parač (with prefix a-parač from *ra-parač), Y₁ ra-parač ‘snake’ (but the Y₁ term is much less common than general Yuu. ba:pi ‘snake’).

Al has seventeen terms for these fifteen glosses, including two sets of synonymous pairs. Nine of the seventeen terms are also found in Y1 (about 50%); but only four clear matches (about 25%) and two possible ones are found in A2. I will give only the A1–Y1 matches here: A1 gudu, Y1 gu:du ‘bushland’ (sometimes ‘arid country’); A1 Y1 dawal ‘country’ (A2 la:l); A1 jolko, Y1 julka ‘ground’; A1 and Y1 miniča ‘jungle’ (A2 minija ‘tree sp. found in mangrove jungle, Cathormion umbellatum’, a possible match); A1 Y1 ɲałal ‘spring’ (A2 ɲaljal); A1 dapolk, Y1 dapalk ‘star’; A1 Y1 walir ‘sun’ (A2 alir). The diffusional A1–Y1 sharings are striking because these are important, high-frequency terms, most of which involve concepts which must have been lexically encoded even before the prefixing and Yuu languages came together.

Moving to A2 and B1, we find an equally striking lack of matches. Using the same glosses, there are sixteen B1 terms, including three sets of synonymous pairs, with two of the original fifteen glosses not represented. A2 shares only two of these sixteen terms: B1 mayamaya (with noun-class prefix, wu-mayamaya), A2 wumayamaya; and B1 lirjal ‘jungle’, A2 lirjal ‘sapling of mangrove, Rhizophora stylosa’. Note the semantic divergence in the second set.

My preliminary conclusions are that Al and Y1 have engaged in extensive lexical diffusion, affecting basic nominal vocabulary roughly as much as (non-basic) flora-fauna vocabulary; and that A2 and B1 have engaged in moderate diffusion with a sharp difference between basic vocabulary (very few sharings) and flora-fauna (about 30% sharings). We can keep these points in mind as we move into other domains.

4. **Body-part nouns.** I begin with a list of sixty-nine A1 terms for human body parts, including substance terms like ‘skin’ and a few terms for bodily secretions like ‘saliva’ (Heath 1978b:174). Nineteen items or 28% occur in Y1 with (near-)identity in sound and meaning, e.g. A1 Y1 binihi ‘nail’, A1 Y1 duyuru ‘ankle’, A1 Y1 dawarak ‘beard, whiskers’. Other glosses are ‘body hair/fur’, ‘scrotum’, ‘mouth’, ‘jaw’, ‘flesh’, ‘tongue’, ‘skin’, ‘sinew/vein’, ‘wrist’, ‘shoulder blade’, ‘rib’, ‘elbow’, ‘saliva’, ‘bone’, ‘navel’, and ‘sweat’. I consider all these sets to be essentially diffusional; however, for ‘mouth’ (A1 Y1 da:) and ‘tongue’ (A1 delh, Y1 di:ly) we find cognates in enough non-Arnhem Land languages so that a Proto-Australian reconstruction is plausible, and hence a retentionist explanation for the A1–Y1 sharing is conceivable. In a few other cases, possible cognates exist outside Arnhem Land, but they show significant phonological or semantic differences; and it is virtually inconceivable that just A1 and Y1 (and one or two neighbors), out of all Australian languages, could have kept unaltered Proto-Australian forms. Since only a small number of items are affected, in any event, extended philological investigations of individual etymologies will not be attempted here; at most, the 28% A1–Y1 sharing index could be lowered to about 25% if plausible Proto-Australian retentions are disregarded.
From the same sixty-nine A₁ items, an A₁–A₂ sharing index of 26% (eighteen items) is obtained. However, to get this figure I have counted all A₂ cognates for A₁ items, whether or not they are close semantic matches. Semantic divergence is seen in A₁ diguru ‘ankle’, A₂ Dguru ‘lower leg (bone)’; A₁ mury ‘backbone’, A₂ mūr ‘carapace of turtle’; and A₁ ĕlē ‘clitoris’, A₂ ĕlē– ‘spike (of spearthrower)’. In general, regular sound changes are seen, with little evidence of recent diffusional leveling; irregular changes are seen in A₁ mar vs. A₂ maray ‘hand’. In addition to the eighteen clear A₁–A₂ matches, five doubtful A₂ cognates involve phonological or semantic problems (A₁ lay ‘cheekbone’ vs. A₂ derivational prefix -li- ‘side’, problematic because of the small number of phonemes, but phonologically possible; A₁ dolo ‘stomach’ and A₂ rala ‘hipbone’, with sharp semantic difference).

In short, lexical sharing indices for this domain give roughly the same figures for A₁–Y₁ pairs (perhaps totally diffusional) and A₁–A₂ pairs (perhaps totally retentions from Proto-A), provided we start with A₁ lists and count all historically related items in the other languages. However, if we limit the sharing indices to cross-linguistic synonyms, the A₁–Y₁ figure is completely unaffected, while the A₁–A₂ figure is reduced to 20% or slightly less. This problem did not arise above, because most A₁–A₂ (as well as A₁–Y₁) sharings in those domains involved synonymous sets. In the body-part domain, the problem is more substantial. It seems to me that the question is not which kind of sharing index to retain, and which to discard; instead, it is how to use all this information in constructing comprehensive historical (including diffusional) analyses. The large number of exact A₁–Y₁ sharings suggests either recent genetic split, with few post-split phonological or semantic changes, or else massive recent diffusion; and since the former is ruled out by morphological differences, the only possibility is diffusion. However, the A₁–A₂ sharings are of the sort found in other language families where the lexical stock of the proto-language has been altered independently in each language by largely internal phonological and semantic shifts. Hence we find a few obvious cognates; others which show fairly substantial changes, but are still recognizable as cognates; and a number of possible but doubtful cognates where we are not sure, even with reasonable knowledge of the historical phonology and morphology. Despite the fact that the A₁–Y₁ sharing index is as high as the A₁–A₂ list (and higher if synonymy is required), the patterning of the sharings is such that diffusion is identified in the first case, and independent retention (for most items) in the second.

Moving to B₁, we start with a list of seventy-four body parts. Clear or probable cognates in the other languages lead to sharing figures of 11% with A₂, 15% with A₁, and 8% with non-contiguous Y₁. These figures are substantially lower than corresponding figures for flora-fauna vocabulary, but comparable to the figures for physical-feature terms. Most of the shared body-part sets involving B₁ have complex histories, and some probably go back to Proto-AB. The only B₁ terms for which recent diffusional intrusion is probable are yarbij ‘thigh’ (cf. A₂ lərbij, A₁ Y₁ darpic, where the B₁ form for phonological reasons must be a late borrowing from A₂) and warj ‘urine’ (cf. A₁ worc, where I suspect B₁ borrowed from A₁). In other examples, the B₁ form is too divergent
to have resulted from recent diffusion, as in B₁ (also B₂) ɲayigan ‘bone’ (A₁ Y₁ ɲaraka, A₂ metathesized ɲagara). In one clear case, A₂ has borrowed from B₁, namely A₂ B₁ ɲgurya ‘excrement’ (this is the only common noun in A₂ beginning with a nasal + stop cluster, though such clusters are common in B₁ nouns).

The body-part domain thus displays a fair amount of diffusion (notably in A₁–Y₁₁), but somewhat less than in other domains. In some language pairs, it appears that special factors are inhibiting massive diffusion in this area. Specifically, in the Yuu. languages, body-part terms are particularly prominent in the lexicon; e.g., many topographic terms like ‘hill’ and ‘river’ are merely secondary uses of body-part terms like ‘backbone’ and ‘neck’ (for an extensive study of such secondary and metaphorical uses, see Schebeck 1978). An important effect of this is that the Yuu. languages have been less favorable to accepting non-Yuu. borrowings in this domain than in some others, and it appears that most A₁–Y₁ sharings result from diffusion from Y₁ into A₁. Intuitively, body-part terms are part of the ‘core’ vocabulary of each language (and they are well represented in lexicostatistical lists); but they are more core-like in some languages than in others, and we should be sensitive to such local cultural variations in comparing data from different parts of the world.

5. Kin Terms. Terms for kin (especially for immediate familial relationships), along with personal pronouns and numerals, were long thought to be highly resistant to diffusion:

‘Dans une langue aussi profondément altérée à tous égards que l’est l’anglais, l’examen des pronoms me, we, us, you, des noms de nombre one, two, three, ten, des noms de parenté father, mother, brother, sister, son, daughter, des verbes be (is, was), come, eat, love, bear etc., suffirait à dénoncer le caractère indo-européen de l’anglais.’ (Meillet 1965:108)

In the present Australian languages, kinship terminology is complex grammatically as well as semantically; there are special suppletive dyadic forms, irregular number markers and possessive affixes, and other features not shared with other nouns. Particularly in the prefixing languages, such as A₂, the systems are complicated with extensive suppletion. One pattern is to use one stem in vocative (‘dad!’) and 1st person forms (‘my dad’), and another in 2nd (‘your father’) and 3rd person forms (‘his/her/their father’); but several other combinations (including as many as four distinct stems) are attested.

In general, we would expect that borrowed kin terms would enter a language initially in the vocative and 1st person forms, and then perhaps spread into other forms. Specialized forms such as dyadic duals (‘father-Dyad’ meaning ‘father-child pair’) might constitute the most archaic lexical layer, particularly in languages where the dyadic suffixes themselves are irregular and unproductive. Creole English kin terms appear to have recently entered local Aboriginal languages in this fashion; such creole terms as banga ‘sister’s child’ occur in my texts for most of these languages, but chiefly in vocative and 1st person form.

A₁ has 27 kin terms based on 21 kin categories (Heath 1978b:38–41). Suppletion is found in wawa?₁ and yawuyu₂₃ ‘elder Br’ and in ɲana₁ and ɲele₂₃
'Mo', where the subscripts specify co-occurrence with pronominal possessors (vocative uses the same stem as 1st person). Partial 'suppletion' occurs in mudi and gamuri 'FaFa' and in bijaja and gabijaja 'MoFa'; but these are simply different forms of the same etymological root, and in each pair the two forms are interchangeable. Alternative unrelated stems, also interchangeable, are found for 'MoBr' (gaykay or lambara) and 'FaMo' (memem or jam?qam).

Of these 27 A1 kin terms (25 if 'FaFa' and 'MoFa' are counted as just one stem each), six (22%) are also found in Y1; these are A1 mudi, Y1 mu:di 'FaFa'; A1 wawaʔi, Y1 wa:wa 'elder Br'; A1 Y1 gaykay 'MoBr'; A1 and Y1 guryŋ 'WiMo, MoMoBrDa'; A1 memem, Y1 mu:mu 'FaMo'; and A1 roydoy, Y1 ruyday 'WiFa'. Even though possible cognates can be found for one or two of these in Pama-Nyungan languages outside Arnhem Land (e.g. in Cape York Peninsula), independent retention from Proto-Australian by Y1 and A1 for these items is very unlikely. Diffusion is strongly suggested by considering local geographical distributions (shared kin terms tend strongly to occur in contiguous blocs of languages cutting across genetic divisions), and by the fact that (as usual for A1–Y1 sharings) we find exact or near-exact correspondences in sound and meaning, while the more distant cognates show considerable differences in these respects. In the case of 'FaMo', the phonology is problematic; but we do seem to be dealing with local variants of a widespread regional term (memem in A1 is shared by other prefixing languages to the west, apparently forming one bloc, while mu:mu in Y1 is a Yuu. form also borrowed into A2 as mu:mu, and another variant mimi occurs in B1 and B2 to the south). Hence recent local borrowing between A1 and Y1 is not attested; but it may well be that memem and mu:mu are divergent offshoots of an old, regionally diffused lexical item.

In the case of 'FaFa', the nature of the borrowing has been clarified by Shapiro 1977, who shows how the system of kin categories in the central Yuu. languages (presumably representing Proto-Yuulngu) has been transformed in Y1 (the most southerly Yuu. language). In Proto-Yuulngu, and now in central languages, we find a modified form of what anthropologists call a Kariera kinship system, in which 'FaFa' and 'MoMoBr' are combined into one term (Yuu. ma:ri, though 'FaFa' is optionally distinguished as ma:riʔmuŋu). The prefixing languages to the south, such as A1 and A2, have the Aranda system in which 'FaFa' (A1 mudi, A2 muri) is distinguished from 'MoMoBr' (A1 gokok, A2 ga:gu) and in which the agnatic descendants of these men are also terminologically distinguished. Y1 has adjusted to the Aranda system of its southern neighbors; and to distinguish 'FaFa' from 'MoMoBr' more rigorously, it has borrowed its present term for 'FaFa' (mu:di), not found in central Yuu. languages, from one of the A-group languages or proto-languages. Similarly, 'WiFa' (A1 roydoy, Y1 ruyday) is a probable borrowing from A1 into Y1, and most probably constitutes a new category in Y1 (where originally 'WiFa' was probably just an instance of the more general category 'FaMoBrSo'). As these examples show, borrowings in the area of kinship terminology often involve significant semantic restructurings (see Scheffler 1978 for a comprehensive treatment of Australian kinship semantics). However, in cases like A1 Y1
A CASE OF INTENSIVE LEXICAL DIFFUSION

Gaykay ‘MoBr’ we are almost certainly dealing with long-standing semantic categories in which one language has lost its old term and borrowed a new term from its neighbor.

Of the same 27 A1 terms, nine or 33% also occur in A2, higher than the 22% for A1–Y1. All the A1–A2 sharings are definite or possible retentions from Proto-A, and several show the effects of post-Proto-A sound changes and morphological restructuring. In this domain, at least, A1 has more sharings with its genetic relative A2 than with its diffusional partner Y1—though the A1–Y1 figure is still respectable, and the A1–A3 figure is not overwhelmingly impressive. I would suggest that the A1–Y1 (diffusional) sharing index would be higher were it not for diffusional reinforcement by other Yuu. languages on inherited vocabulary in Y1 (despite the fact that partial semantic re-alignment has taken place), and similar diffusional pressure on A1 from its non-Yuu. neighbors. All (or almost all) Y1 kin terms not shared with A1 are general Yuu. kin terms. Some A1 kin terms not found in Y1 can be demonstrated to be borrowings from languages to the south. A1 yanuri ‘FaFa’ is obviously based on B1 ya-muri (with 1sg. prefix ya-; cf. the other A1 variant mu-di); and A1 yanbajaja is likewise from B1 ya-bijaja.

Of twenty-two B1 kin terms, six fairly clear matches occur in contiguous A2 (27% shared). Some of these are regional terms, but may go back to Proto-AB in some cases; it is difficult to identify clear borrowings. Sharings are B1 and A2 baba ‘Fa’ (cf. Yuu. ba:pa); B1 bibi, A2-wibi ‘Mo’ (also bibi ‘MoBr’ in A2); B1 A2 mu:ri ‘FaFa’ (cf. A1 and Y1 forms discussed above); B1 mimi, A2 mu:mu ‘FaMo’ (indirectly related, cf. discussion above); B1 gaga, A2 ga:gu ‘MoMo(Br)’ (cf. A1 gokok, B2 gugu); and B1 wuruy, A2 gurumbaj (*guruy-baj) ‘WiMo and other avoidance relatives’ (A1 Y1 guruy).

From the above data it is fair to conclude that kin terms are somewhat more resistant to borrowing than other vocabulary, such as that for flora and fauna. Particularly in the context of more detailed philological analysis involving identification of archaic layers (e.g. in dyadic forms not considered here), kin terms can be useful evidence for genetically-based historical linguistics. However, even in this domain diffusion is fairly extensive. Precise statistics are less valuable here than elsewhere—because of the small number of items, possibilities of common retention instead of diffusion for some shared items, and possibilities of coincidental sharing with forms like baba ‘father’. However, approximately 20% sharing is found between A1 and Y1, and I believe this is entirely the result of diffusion (partly recent, partly old).

Finally, we ask whether the borrowings are concentrated in comparatively distant genealogical categories, such as grandparents, as opposed to immediate kin like parents, children, and siblings. While there may be some variation in rates of diffusion of core and peripheral kin terms, it is certainly incorrect to think that core relations are immune to diffusion. A1–Y1 diffusional sharings include ‘elder Br’ and the very important ‘MoBr’ category. We cannot tell whether diffusion is involved in B1 A2 baba and Yuu. (including Y1) ba:pa ‘Fa’ because of the possibility of coincidence; but we certainly cannot rule out diffusion. ‘Mo’ seems to have been relatively impervious to diffusion; but A3
\(\eta\text{nda} 'Mo'\) seems to be either a borrowing from some Yuu. language (\(\eta:\text{ndi} 'Mo'\)) or an inherited form (cf. \(\text{A}_1 \eta\text{na}\)) showing diffusional contamination (all inherited word-final vowels become \(a\) in \(\text{A}_3\), so final \(a\) in \(\eta\text{nda}\) does not decide the question).

6. OTHER NOUNS. In this final, residual section on nouns, I deal informally with semantic classes not specifically discussed above. I emphasize the extent of diffusion of important and high-frequency nouns between \(\text{A}_1\) and \(\text{Y}_1\), with lesser attention to other language pairs.

All these languages contain a set of important human age/sex category terms like 'man' and 'old woman'. \(\text{A}_1\)-\(\text{Y}_1\) sharings in this domain are simply astounding; most of the terms are shared, and all the sharings are clearly diffusional. Hence \(\text{A}_1 \text{Y}_1 \ '\text{gadaku} '\) boy (before circumcision)\'; \(\text{A}_1 \text{Y}_1 \ '\text{gurmul} '\) young man, bachelor (from circumcision to marriage)\'; \(\text{A}_1 \text{deremu}, \text{Y}_1 \ '\text{daramu} '\) man, male\'; \(\text{A}_1 \text{Y}_1 \ '\text{gul} '\) woman\'; \(\text{A}_1 \text{Y}_1 \ '\text{jawulp} '\) old man\'; \(\text{A}_1 \text{Y}_1 \ '\text{yaciwianggu} '\) old woman\'; and even the generic \(\text{A}_1 \ '\text{yu:lu} ', \ '\text{yu:lu} '\) person, human being'. All these terms are very common in both languages, and in most cases are the only terms for the glosses indicated. Among the major age/sex terms, only 'child' (same as terms for 'small') and 'girl' are distinct. In contrast to these seven \(\text{A}_1\)-\(\text{Y}_1\) sharings, \(\text{A}_1\) and \(\text{A}_2\) share only four (the \(\text{A}_2\) forms matching \(\text{A}_1\) forms shown above are \(\text{nun-gara} ' '\) boy, metathesized \(\text{wulmur} ' '\) young man', and \(\text{najiwianggu} ' '\) old woman'; cf. also \(\text{A}_1 \text{A}_2 \ '\text{yala} '\) 'girl').

Here is a miscellany of high-frequency terms for important natural objects in the region: \(\text{A}_1 \text{Y}_1 \ '\text{bulug} ' '\) ashes\'; \(\text{A}_1 \text{Y}_1 \ '\text{garkan} ' '\) charcoal\'; \(\text{A}_1 \text{Y}_1 \ '\text{gunug} ' '\) cloud\'; \(\text{A}_2 \text{Y}_1 \ '\text{wukun} ' '\) cloud\'; \(\text{A}_1 \text{Y}_1 \ '\text{bal} ', \ '\text{ba:l} ' '\) firewood\'; \(\text{A}_2 \text{Y}_1 \ '\text{gajiwianggu} ' '\) sand\'; \(\text{A}_1 \text{Y}_1 \ '\text{bijudu} ' ' '\) whirlwind\'). \(\text{Y}_1\) has borrowed \(\text{gujark} ' '\) water from \(\text{A}_1 \ '\text{jark} (\text{gu-jark} with noun-class prefix), though it is less common in \(\text{Y}_1\) than common Yuu. \text{gupu}.'

Cardinal-direction terms like 'north' are quite important in these languages; in most cases, 'up' (also meaning 'inland' and 'upriver') and 'down' (also 'on water'). For a similar case involving inclusion of an \(\text{A}_1\) noun-class prefix in a (synchronically unsegmentable) \(\text{Y}_1\) loanword, see fn. 3, above. \(\text{Y}_1\) itself lacks the noun-class system.

I do not deal here with the question of noun-class assignment of loanwords (e.g. \(\text{Y}_1\) nouns borrowed into \(\text{A}_1\)). Usually the noun class of a semantically similar noun is extended to the borrowed noun. If both languages have noun classes, and bilinguals are able to connect classes in one language with those in another, borrowings may simply retain the class assignment of the source language. However, there are many counter-examples (especially in \(\text{Bi}\)), and the situation is rather complicated. Most borrowings into \(\text{Y}_1\) from noun-class languages (like \(\text{A}_1\) and \(\text{A}_2\)) omit the noun-class prefix.
the coast’ and ‘downriver’) are also morphologically associated with these stems. All these are usually best analysed as nouns morphologically, though with some special features. Again we find astounding degrees of diffusional sharing between A₁ and Y₁. All four compass-point terms are diffusionaly related, though here (unlike most other A₁–Y₁ diffusional sets) some phonological differentiation has occurred: A₁ rawara, Y₁ rawaray ‘east’; A₁ yuri, Y₁ yururu ‘north’; A₁ bakay, Y₁ ba: gay ‘south’; A₁ Y₁ yanı ‘west’.

By comparison, A₂ has only one clear cognate (wagi- ‘south’) and one very dubious one (ramalı- ‘east’). However, A₂ has borrowed argali ‘west’ from B₁ to the south, and has even taken over the directional derivative argali-ni ‘westward’ with its B₁ suffix -ni. As for ‘up’ and ‘down’, Y₁ shares ‘up’ with A₂ (Y₁ garwar, A₂ arwar) and possibly with A₁ (garkala); it also shares ‘down’ with A₂ (Y₁ diripi, diripa-la, A₂ lirıbala). Examining only the A₁–Y₁ and A₂–Y₂ pairs, therefore, we see clear diffusion involving all six cardinal-direction terms, with A₁–Y₁ sharing all four compass-point terms.

Meillet mentioned numerals as highly resistant to diffusion (see quotation above), but they too have been borrowed in Arnhem Land. A₁ waygihn ‘one’ is diffusionaly shared with waygihn and other variants in several Yuulanguages—though Y₁ has a distinct stem widipiyay, so there is a slight interrup tion in the geographical distribution of waygihn. As for ‘two’, Y₁ bulal- may be diffusionaly related to A₂ wula- (*bula-), and more especially to reduplicated A₂ wulalwulal ‘four’; but *bulal(ı) is also found outside Arnhem Land, and I cannot now rule out a retentionist explanation. Numerals other than ‘one’ and ‘two’ are usually derived from them, and/or from words for ‘hand’ and ‘foot’.

A few other nominal A₁–Y₁ sharings are mul?mul ‘black’ (Y₁ also has unreduplicated mu:l), wurk ‘bushfire’, balpara ‘companion’, jambal ‘oven’, qa:lan ‘shell’, ji:ci2 ‘sore, wound’, and dunupa ‘straight’, all with identical sounds and meanings; note also A₁ dowo, Y₁ da:wu ‘word, speech’.

Meillet has been emphasizing A₁–Y₁ sharings, but a fair number of A₂–Y₁ sharings also exist. Many of the A₁–Y₁ pairs already mentioned involve words which also show up with A₂. Some found only in A₂ and Y₁ are these: A₂ landa, Y₁ danda ‘ant-mound’; A₂ Y₁ bungawa ‘boss, owner’; A₂ aṣa, Y₁ wa:ya ‘camp, country’; A₂ maragarj, Y₁ madakarić ‘dangerous, violent’; A₂ ma:bu, Y₁ mapu ‘egg’ (one of two synonyms in A₂); A₂ aban, Y₁ gaban ‘ground’ (more often julka in Y₁); and A₂ raṭag, Y₁ raṭa ‘tree, wood’ (more often ḏarpa in Y₁).

Over-all, it is clear that diffusion has been extensive in nouns of all types, particularly between A₁ and Y₁. I have found no nominal domain of any size in which diffusional sharings between these two languages are less than 20% (based on A₁ lexical lists); and some domains (trees/shrubs, human age/sex terms) show well over 50% sharing. In a small number of cases, related forms occur in non-Arnhem Land languages; but the (near-)exact phonological and semantic identity of A₁ and Y₁ items, with the fact that these lexical items are often not clearly reconstructible for proto-languages such as Proto-A and Proto-AB, suggests strongly that few or none of the nominal sharings result simply from independent retention from Proto-Australian.
However, Meillet and many others have noted that verbs are more difficult to borrow than nouns, and the question arises whether the A₁–Y₁ sharing indices are substantially lower for verbs than for nouns.

7. THEMATIC VERB CLASSES. My primary contention here and in §8 is that diffusion of verbs has been about as high as for nouns, where structural conditions were favorable—namely where either (a) an uninflected simplex or ‘root form’ of the verb occurs, or (b) the source language and borrowing language have closely congruent morphological frames for verb stems which permit direct conversion.

I use the term ‘thematic’ for those verb classes in which a root form is distinguished from inflected forms containing a thematizing suffix followed by inflectional endings. An example is A₂ -wulup-du- ‘to bathe’, which produces a root form wulup (optionally -wulup with pronominal prefix present), and inflected forms like past continuous -wulup-du-yi and future -wulup-du-ŋ.

In A₁ and Y₁, the largest and most productive verb class is thematic, with suffix -du- in A₁ and -du- (variants -yu-, -u- etc.) in Y₁. Although the inflectional suffixes which can follow the thematizer -du- are quite different in the two languages (e.g., Y₁ has past -du-na and future -du-ru), the identical suffix -du- results from morphemic diffusion (Heath 1978a). Because both languages permit the verb root to occur in isolation, it is possible for the roots to be borrowed from A₁ to Y₁, or vice versa, without the kind of interference from morphological structures which prevents diffusion of verb stems in many other language-contact zones. In addition, even the morphological frames for inflected forms are the same (Root-du-Suffix), and bilingual A₁–Y₁ speakers can easily make the association between the -du- thematic classes in the two languages; hence direct diffusion of verb roots is probably possible, even without the presence of root forms.

Other verb classes which might be considered here include a less productive thematic class with suffix -da- in A₂, and a type of ‘auxiliary’ construction in which a root is directly followed by a monosyllabic verb root, usually -bu- ‘to hit’, found in both A₁ and Y₁—possibly (but not necessarily) resulting from old diffusional interaction.

To get an idea of the extent of A₁–Y₁ sharings, I randomly chose the initial consonants b l ŋ, and made a list of A₁ roots of the -du-, -da-, and auxiliary (mostly -bu-) types. I ended up with 32 -du- roots, two -da- roots, and nine roots which took auxiliaries (six with -bu-). Of the 32 roots with -du-, at least 14 clear Y₁ matches are found, for a minimum 44% sharing index. Examples are A₁ -bal₂-du-, Y₁ bal₂-yu- ‘to hide (in grass)’; A₁ -da:-bar-du-, Y₁ da:-bar-yu- ‘to open mouth’ (compound with da: ‘mouth’); and A₁ -bilan₂- -du-, Y₁ bilan₂-gu- ‘to lick’. One probable match, not included in the clear matches, is A₁ -lul-du-, Y₁ lu:-yu- ‘to wade’, with slight phonological divergences (but lu:-yu- is exactly comparable to A₂ -lu:la- ‘to wade’, both from *-lu:-du- with regular changes). I also did not count A₁ bil?bil-du- ‘to sing with tapsticks’ and Y₁ noun bil-bil ‘tapstick’, because of the word-class difference. Thus 44% is a conservative figure for the sharing index here. Because the
-du- class contains several hundred verb roots in both A1 and Y1, total A1–Y1 sharings in this class are probably in the hundreds.

More modest figures are found for -da- and the auxiliary forms. No Y1 matches occur for the two A1 -da- verbs in the sample, and this seems true in general of -da- verbs. Of the nine A1 auxiliary forms in the sample, two exact matches (involving -bu-) are found: A1 Y1 lak-bu- ‘to rip flesh from shell’, and A1 Y1 lark-bu- ‘to cut up’. Non-exact matches involving distinct endings occur in two cases: A1 -liw-ga- ‘to hunt kangaroos with fires’, Y1 (-du- class) liw-u- ‘(fire) to blaze’, causative liw-(u-n-)mara- ‘to set (fire)’; and A1 -yer2-yu- ‘to be fond of’ (with -yu- ‘to lie down’ as auxiliary). Y1 yir2-yu- ‘to breathe’ (with -yu- allomorph of thematizer -du-). Because of semantic clustering of ‘breathe/heart/love’ in these languages, I am inclined to accept the last correlation as probable, but involving formal and semantic divergences.

We therefore have high rates of diffusional sharing for the productive -du- classes, and a reasonable amount (20% or more) for auxiliary compounds. Precise statistical figures should be used with caution; elicitation of verb forms is naturally less systematic than that of flora-fauna or kin terms, so the lexical inventories on which the statistics are based are somewhat less complete for the verbs. (It is possible that additional lexical materials for verbs would increase the sharing indices slightly.) Shared -du- verbs include many high-frequency stems; the glosses for the 14 sharings in the sample (i.e. with roots beginning in b l y) are ‘to hide (in grass)’, ‘to open mouth’, ‘to lick’, ‘to blame or reproach’, ‘to chant (in ritual)’, ‘to cross’, ‘to break open’, ‘to blow or puff’, ‘to make string’, ‘to sneak up’, ‘to fly’, ‘to cut up’, ‘to spill’, and ‘to go up’.

A2 has clear traces of the old *-du- thematizer (which has merged in A2 with *-da-); this *-du-, borrowed from Yuu. languages, was already present in Proto-A. However, A2 has tended to eliminate the simplex (root form), so the *-da- thematizer (representing *-du- as well as *-da- of Proto-A) is now largely unsegmentable as part of the verb root. A2 does have a simplex, but it is often suppletive (formally unrelated to the semantically associated stem); and even when etymologically cognate, the root form and inflected stem are no longer usually related by productive phonological regularities (yald- ‘to go past’ from *jolk-du-, with root form jalg from *jolk). Although *-da- has recently been re-activated as a way of incorporating creole English verbs into A2 (-warda- ‘to work’, cf. Y1 wark-u-; -bayanda- ‘to buy, pay’, cf. Y1 bayam2-bu-), it would appear that A2 has made considerably less use of thematizer *-da- in recent centuries to adopt borrowed verb stems from other languages.

A1 could introduce a Y1 root in the uninflected root form, and then incorporate the root by using the productive thematizer -du-; but the A2 thematizer *-da- was more frozen and less productively segmentable, so this procedure for incorporating borrowings was either unavailable or at least of limited significance. A2 stems containing reflexes of *-da- often correspond to -du- stems in A1 and/or Y1, but these are mostly inherited from Proto-A and show sound changes or other divergences vis-à-vis A1; the A1–A2 sharing index is lower than the A1–Y1 figure. For the 32 A1 -du- verbs in the sample, good matches in A2 occur for four items (13%): A1 -bilp-bil- -du- ‘to sing with tapsticks’, A2
-wilbila- 'to tap'; A1 -bolk-du-, A2 -wala- 'to appear, arrive'; A1 -bu₂-du-, A2 -bu:la- 'to blow'; and A1 -buju₂-du- 'to make string', A2 -wuyuda- 'to rub in hands'. A2 has a couple of other conceivable cognates involving sharper phonological and/or semantic divergence; but even counting these, the sharing index is far lower than the A1–Y1 figure of nearly 50%.

Languages of the B group do not have a precisely analogous thematic structure. Instead, they have a fairly small number of directly inflectable roots, several of which can also be used as auxiliaries. For languages like A1, I use the term ‘auxiliary’ for compounds of the type -bit-bu- ‘to climb’ (as in ŋ-bit-bu-ni ‘I climbed’, with 1sg. ŋa- and past continuous -ni), so that the root (-bit-) and auxiliary are welded together; however, in the B languages the ‘auxiliary’ construction involves a root preceding a complete inflected form, hence root war ‘to see’ and auxiliary -windi- in B1 war + ŋa-windi-ma ‘I saw him’ (‘+’ indicating root boundary, 1sg. on 3sg. ŋa-, past continuous -ma). Although the B root can occasionally occur by itself, this is much less common than the root forms of A1 and Y1. It may be possible to incorporate borrowed roots in the root form, and then incorporate them grammatically using productive auxiliaries; but this is more difficult than in the northerly languages.

In the sections on nouns, above, we saw evidence for moderate degrees of diffusion between A2 and B1, though less than for the A1–Y1 pair. However, the figures for A2–B1 verbs are low. We expect this, since the two languages have significantly incongruent verbal morphological structures, and since neither language makes extensive use of root forms (except for suppletive root forms in A2). Although hundreds of B1 roots of the type seen in war ‘to see’ are attested, virtually none directly match A2 (or A1) roots; and the very few which do seem to be archaic are possibly inherited from Proto-AB. We can cite B1 B2 lar ‘to cut up’, connected with A1 -lar²lar-du-, Y1 lar²lar-yu- ‘to cut up’ (no match in A2); and B1 B2 yal, with irregular reduplication yaljal ‘to go past’, connected to A1 -jolk-du-, A2 -yalda- (archaic root form jalg), Y1 julk-u-, Y2 julk-du-, all ‘to go past’. In both sets I suspect that the A and B forms are inherited from Proto-AB, and that the Yuu. forms were borrowed from A1 or one of its direct ancestors.

Thus B1 shows no clear signs of having borrowed recently from A2 or A1; but B1–B2 sharings are very common—indeed, rather moreso than would be expected in view of the fairly remote time depth for Proto-B which is suggested by the extent of morphological divergences. Quantification of B1–B2 sharings is imprecise—since my B1 data were from the last good speaker, who also spoke B2 very well, and may have unintentionally introduced a few B2 elements into the B1 material which he gave me. Even with some allowances for this, a B1–B2 sharing index of 50% or more can be attributed to the pre-white period; and the very fact that my consultant could spontaneously introduce B2 roots into his B1 speech shows that diffusion between the two was not difficult. Although neither language made extensive use of root forms, they shared the same basic system of auxiliaries, and morphological integration of borrowed roots was not significantly impeded.
8. Non-thematic verb classes. Here I deal with inflectional conjugations in which the verb root is directly followed by inflectional endings, has no root form, and in the prefixing languages (A and B groups) is immediately preceded by pronominal prefixes. In the B group this refers to the auxiliaries (e.g. -windi- in B1 war + ɣa-windi-ma ‘I saw him’) and the few verbs which are directly inflected (-na- ‘to see’ in B2 ɣa-na-ji ‘I saw him’); some roots occur in both these functions. In all languages, non-thematic verb classes are usually unproductive and closed, in the sense that the majority of attested stems seem archaic (some appear to go back to Proto-Australian). Consequently, both the quantity and patterning of sharings in this section will be considerably different from those seen in earlier sections.

We consider first A1 and A2, in which inflectional classes containing principally bisyllabic and longer roots are distinguished from those with monosyllabic roots. A1 Class III contains seven roots, and there are two A2 matches: A1 -naki-, A2 -nagi- ‘to burn (intr.)’; and A1 -waki-, A2 -awi- ‘to return’. From ten Class IV roots in A1, we get four A2 cognates: A1 -maka- ‘to call’, A2 -maga- ‘to tell’; A1 -benya-, A2 -wanya- ‘to tread on’ (also ‘to dance’ in A2); A1 -gopa- ‘to keep’, A2 -waba- ‘to wrap’; and A1 -nima-, A2 -nima- ‘to hold’. Four A1 Class VI roots show one A2 match: A1 -watu-, A2 -arua- ‘to abandon’. Phonological and inflectional class correspondences are regular, and all these roots go back to Proto-AB (the t/r correspondence in the last example is not fully regular, but a similar d/r correspondence is found in A1 -ru-du-ni, A2 -ya:-ri: ‘goes’; cf. B1 -ra-ra-ni).

In A1, Class IV stems (less than ten) do not exactly match A2 stems in the corresponding class. However, A1 Class IV is dominated by -ga- ‘to carry’ and compounds built on it; similarly, most of the A2 stems are frozen compounds with *-ga- (which does not survive by itself in A2).

In monosyllabic roots, mostly with highly irregular and obviously archaic paradigms, the cognate density between A1 and A2 increases dramatically; for 13 such A1 stems (excluding -ga- ‘to carry’, just mentioned) we find no fewer than 11 A2 cognates, and the paradigms in the two languages are clearly derived from the same Proto-A paradigms: A1 -yay- ‘to hear’ (compound -yan-yay- ‘to hear voice’) and A2 -yaya- (*-yan-yay-) ‘to hear’; A1 -na-, A2 -na- ‘to see’; A1 A2 -nu- ‘to eat’; A1 -wo-, A2 -i/-u- ‘to give’; A1 A2 -ma- ‘to get, pick up’; A1 -do-, A2 -la- ‘to chop’; A1 -ba-, A2 -wa- ‘to bite’; A1 -bu-, A2 -wu- ‘to hit, kill’; A1 -jaka-di-, A2 -la- ‘to stand’; A1 -yo-, A2 -ya- ‘to sleep’; A1 -ni-, A2 -na- ‘to burn’ (tr.) (Apparent homonyms like A1 -na- ‘to see’, -na- ‘to burn’ are differentiated by distinctive inflectional endings.) Of the remaining two, A1 -yu- ‘to put on’ is conceivably identical to A2 -yu- (various meanings, including ‘to produce, create’) with similar inflectional endings; A1 -ni- ‘to sit’ has no A2 match (A2 -bura- ‘to sit’), but does match B1 -na- ‘to sit’. All the A1–A2 matches involve retentions from Proto-A; these matches, plus point-for-point paradigmatic affinities (overlaid by sound changes and some morphological restructuring), are prime evidence for the close genetic relationship between the two languages.
Indeed, several of these items are so archaic that they can be reconstructed for Proto-Australian (which, for present purposes, I take as the immediate common ancestor of the Pama-Nyungan family, including the Yuu languages, and the A and B groups of prefixing languages). Consequently, some matches between A1 and Y1 are found involving the same set of A1 roots. However, these are almost exclusively common retentions, rather than recently diffused items; this is reflected in the significant phonological, morphological, and (to some extent) semantic divergences found in these A1–Y1 matches (standing in sharp contrast to the A1–Y1 matches in earlier sections—which show near-exact phonological and semantic identity, strongly suggesting recent diffusion). Hence A1 Y1 bu- ‘to hit, kill’ (with very different paradigms, though traces of Proto-Australian augmented form *bu-m occur in both in different parts of the paradigm); A1 -ga-, Y1 -ga: ‘to hear’; A1 -na-, Y1 -na: ‘to see’; A1 -ma-, Y1 ma:ra- ‘to get, pick up’; A1 -do-, Y1 du: ‘to chop’; A1 -jaka-di-, Y1 da:ra- ‘to stand’; A1 -ni-, Y1 ni:na- ‘to sit’; and possibly A1 -ba- ‘to burn’, Y1 batha- ‘to burn, bite’. Most of these are Proto-Yuulngu (as well as Proto-A) forms, found in other Yuu. languages as well as in Y1; and most have cognates in other Pama-Nyungan groups elsewhere in Australia (e.g. Warr. *na- ‘to see’, ja- ‘to stand’, *ni- ‘to sit’). For the majority of the A1–Y1 matches, then, consideration of all available evidence (geographical distribution, phonological and semantic divergences, and comparative morphology) points strongly to retentionist rather than diffusional explanations.

This list contains one clearly diffusional item: Y1 du: ‘to chop’ (competing with other roots in the same meaning, and showing only certain inflectional forms) is not a general Yuu. term, and is almost certainly a borrowing from A1 -do- ‘to chop’ (which goes back at least to Proto-A). It is also conceivable that Y1 and Proto-Yuulngu. -na: and Proto-A *-na- ‘to hear’ (A1 -na-, A2 -na: from *-yaNa-) are related diffusionally, with Proto-A borrowing from Proto-Yuulngu. Although cognates (e.g. *yami reflected in Cape York Peninsula languages—K. Hale, p.c.) occur in other Pama-Nyungan groups, I am not yet sure that Proto-A *-na- can be attributed to earlier proto-languages in its line of descent, such as Proto-AB; and unless such attribution is established, I will regard the Proto-A form as a possible borrowing. Nevertheless, the pattern of retention vs. diffusion of irregular monosyllabic verbs differs sharply from patterns seen in nouns and even in thematic verbs.

Comparisons between A and B languages are complicated by the fact (mentioned above) that the B languages have each reduced their inventory of Proto-B inflectible verb roots as they have evolved toward a verbal system with a few inflectible auxiliaries. Thus a Proto-B *-na- ‘to see’, matching Proto-A *-na-, is reconstructible on the basis of B2 (-mi)-na-; but this has been replaced in B1 by a new auxiliary construction with root war and auxiliary -windi-. However, B1 preserves some inflectable roots like -na- ‘to sit’ (matching Proto-A *-nV-, with variable vowel, in the same meaning) which are not preserved in B2. Allowance for such morphological changes must be considered in evaluating sharing indexes across the A–B boundary, or even among B languages. If B1 and B2 are combined, they have cognates for Proto-A *-na- ‘to see’,
*-dV- ‘to stand’, *-ga- ‘to carry’, *-wo- ‘to give’, *-yu- ‘to eat’, and perhaps one or two others. Proto-B paradigms may have been rather different from those of Proto-A for cognate stems, but some striking resemblances occur (cf. B1 -ju-ra with A1 -du-da, A2 -la-ra, reflecting *-Ju-da or the like with some laminal *J, a paradigmatic form of *-JV- ‘to stand’, Proto-A *-dV-). Other parallels were cited in §1, and cumulatively these correspondences (along with others, e.g. in pronominal prefixes) are sufficient to establish clearly that Proto-B and Proto-A were much more closely related genetically than either was to Proto-Yuulngu or Proto-Pama-Nyungan.

9. SUMMARY OF DATA. Table 1 presents a summary of the sharing indices for the most significant pairs of languages dealt with here: A1–Y1, A1–A2, and A2–B1. Recall that the A1–Y1 sharings are almost exclusively diffusional, except for monosyllabic verbs; but A1–A2 sharings are mostly retentions from Proto-A with some continuing diffusion, and A2–B1 sharings are mostly diffusional for nouns and retentions for verbs.

<table>
<thead>
<tr>
<th>Language Pair</th>
<th>Trees/ Shrubs</th>
<th>Insects</th>
<th>Flora-Fauna Over-All</th>
<th>Major Physical-Feature Nouns</th>
<th>Body-Part Nouns</th>
<th>Kin Terms</th>
<th>Major Human Age/sex Terms</th>
<th>Cardinal-Direction Nouns</th>
<th>Thematic Verbs (Sample)</th>
<th>Non-Thematic Monosyllabic Verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1–Y1</td>
<td>65% (69/106)</td>
<td>35% (14/40)</td>
<td>78% (7/9)</td>
<td>44% (14/32)</td>
<td>57% (8/14)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1–A2</td>
<td>44% (47/106)</td>
<td>35% (14/40)</td>
<td>53% (9/17)</td>
<td>28% (19/69)</td>
<td>44% (14/32)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2–B1</td>
<td>30% (31/103)</td>
<td>(poor data)</td>
<td>25–30% (est.)</td>
<td>26% (18/69)</td>
<td>22% (6/27)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Figures in A1–Y1 and A1–A2 columns are based on the percentage of items in A1 lists found in Y1 or A2 (exact semantic and formal identity not required). Figures in A2–B1 column are based on B1 list. Cardinal directions include compass points, ‘up’, and ‘down’. B1 uninflected verb roots in leftmost position in auxiliary constructions are counted here as ‘thematic verbs’. All figures include only relatively certain matches. Elimination of matches showing semantic divergence would reduce figures in bottom-most row for all columns; in other rows, it would largely leave A1–Y1 figures unchanged, but would sharply lower A1–A2 and A2–B1 figures in most categories. The verb *-ga- ‘to carry’ is counted in non-thematic monosyllabic verbs.

A few additional figures have been given in the table (e.g. for A2–B1 cardinal-direction terms), but most have been presented earlier. I will stress again that, to varying degrees, the statistics shown are to be considered approximate—because they were produced according to particular procedures (taking A1 and B1 lists as basic), because some of the percentages are based on small samples, because the data are not entirely complete (notably for B1), and because decisions had to be made concerning possible but doubtful matches. Since the doubtful matches were generally excluded from the figures shown, and the other sources of statistical error (or arbitrariness) do not seem to point simultaneously in the same direction, I consider the figures to be representative of
linguistic reality. Of course, the sharing indices do not themselves directly indicate diffusional as opposed to retentionist sharings; but in \( A_1 - Y_1 \) we can safely assume that the vast majority (and in some cases all) of the sharings are diffusional except for the monosyllabic verbs (most of which are from Proto-Australian), while the \( A_1 - A_2 \) and \( A_2 - B_1 \) figures involve various proportions of diffusion and retention.

It is apparent that \( A_1 - Y_1 \) diffusional sharings are very extensive, amounting to approximately 50% of total attested \( A_1 \) vocabulary, widely distributed across lexical categories; lows approach 20%, and highs approach 70%. However, diffusion between \( A_2 \) and \( B_1 \) has been considerably less extensive, though rather high by non-Australian standards; in unfavorable morphological contexts (thematic verbs), it is almost nil.

10. IMPLICATIONS FOR LEXICOSTATISTICS. I will use the term ‘lexicostatistics’ here for those approaches, including traditional lexicostatistics and glottochronology, in which lexical sharings between two languages are expressed by a single numerical figure (usually based on percentage of shared items relating to a 100- or 200-word list). I argue that, at least in the present case, it is essential to replace this with a more flexible approach to the statistical investigation of lexical sharing in which (a) specific morphological and semantic idiosyncrasies of the languages are considered; (b) sharings which are sufficiently close phonologically and semantically to be of diffusional origin are distinguished from those which cannot be (because of sound changes and other language-internal developments); and (c) different figures are computed for different types of vocabulary (from different semantic domains, with productive and unproductive inflectional classes distinguished) and subsequently contrasted to each other. Enlightened ‘lexicostatisticians’ who subscribe to some or all of these practices are not the targets of my criticism.

We can see the problems of traditional lexicostatistics by computing sharing-indices based on the Swadesh 100-word list (for the various lists and discussion thereof, see Hymes 1960). For this computation, I have prepared lists for each relevant language based on the 100 glosses in the list. For some of these glosses, I was unable to produce items from these languages—because the concepts in the list do not correspond to lexical concepts in these languages, because the concepts are expressed by stems whose basic or primary meaning is something else (‘seed’ was eliminated because some languages express it by the word for ‘eye’), or because of major gaps in the data (the excluded items are ‘all’, ‘bird’, ‘seed’, ‘bark’, ‘horn’, ‘claw’, ‘come’, and ‘round’). For a few other items, I have forms from most languages, but one or two lack attested forms (e.g., ‘new’ has no simple lexical expression in \( Y_1 \) or \( B_1 \)). For ‘breast’, I used words for ‘milk, breast’; for ‘walk’, I used ‘go’; for ‘kill’, I used ‘hit, kill’; for ‘say’, I used ‘do that, say that’; for ‘white’, I used ‘white, light-colored’; for ‘black’, I used ‘black, dark’; for ‘we’, I used the exclusive plural form; for ‘that’, I used the distant non-anaphoric form. In a few cases, a given language had two synonymous stems of equal frequency (or for which relative frequency had not been established); matches involving just one of the two synonyms
were counted as one-half match. In general, phonological criteria for acceptance of matches were unproblematic, since much of the comparative phonology is understood. Figures obtained in this fashion from the 100-word list are as follows (the first figure includes only relatively certain matches; the parenthesized figure includes doubtful matches as well):

\[
\begin{align*}
A_1-Y_1 &: 28 \ (35) & A_2-Y_1 &: 14 \ (17) \\
A_1-A_2 &: 21 \ (22) & A_1-B_1 &: 12 \ (14) \\
A_2-B_1 &: 7 \ (13) & Y_1-B_1 &: 4 \ (7)
\end{align*}
\]

Regardless of what decisions are made concerning doubtful matches, it is clear that the relatively high \(A_1-Y_1\) figure greatly exaggerates the real genetic relationship between the two languages—especially since some of the 100 items were not included, for reasons just given. Although the \(A_1-Y_1\) figure includes some probable independent retentions from Proto-Australian (chiefly monosyllabic verb roots), the figure is grossly inflated by diffusional sharings, many of which have been shown above. Although \(Y_1\) and \(A_2\) are in a considerably less intensive diffusional relationship than \(Y_1\) and \(A_1\), their sharing-index is still somewhat inflated by such diffused items as ‘stone’ (\(Y_1\ nu:ka?, A_2\ nuga\)). We find an \(A_1-A_2\) figure lower than that for \(A_1-Y_1\), which severely misrepresents the close \(A_1-A_2\) relationship which can be proved on the basis of comparative morphology and phonology. We can take some consolation from the fact that the \(A_1-A_2\) figure is correctly higher than the \(A_2-B_1\) and \(A_1-B_1\) figures. Note that the \(Y_1-B_1\) figure (involving non-contiguous languages with remote genetic relationship) is much lower than the other figures, all of which involve contiguous languages. Although \(A_1\) and \(B_1\) are equally remote from \(Y_1\), the \(A_1-Y_1\) and \(Y_1-B_1\) figures are sharply divergent.

One way to identify instances of extensive diffusion, and thus to avoid serious mistakes in genetic classification, is to use complex mathematical techniques to analyze relationships among many language pairs simultaneously, instead of merely doing pair-by-pair computations. I appreciate that this is possible, assuming that we are fortunate enough to have adequate data on all relevant languages in a region, and that we know the geographical configurations. However, it is probably better (and in many cases more feasible) to attempt the same thing by more sophisticated analyses of sharings within language pairs, by contrasting data from different semantic categories, and by distinguishing productive from unproductive (archaic) classes. Despite the extensive lexical sharings involving \(A_1\) and \(Y_1\), any competent historical linguist in possession of the facts could see immediately that the two are not closely related genetically, and hence that many of the sharings (even in the 100-word list) are diffusional. Even if we leave aside the overwhelming evidence from comparative morphology, we can see from the precise patterning of lexical sharings that the two languages are genetically remote. Sharings are very common in such easily diffusible lexical domains as flora-fauna vocabulary and thematic verb classes (since both languages have the same thematizing system). Sharings are less common in unproductive, closed verb classes in which root forms do not occur; and the apparent sharings which do occur here are mostly irregularly
related (as with A1 -ma-, Y1 ma:ra- ‘to get, pick up’). From these data we should at least suspect that A1 and Y1 have a possible remote genetic relationship, reflected in some sharings in the most archaic lexical stratum, and that they have in more recent times undergone massive diffusional interaction. We can achieve plausible hypotheses of this type (to be tested further through comparative morphology and phonology) provided we break the vocabulary down to begin with into different classes—based in part on our knowledge of the inflectional structures of the particular languages in question—and then contrast sharing-indices obtained separately from each class. Moreover, we do not limit ourselves to gross statistical figures, but also consider the extent of phonological and semantic differentiation reflected in the comparisons. Although our primary justification for rejecting a close genetic connection between A1 and Y1 is that A1 can be derived from Proto-A, that Y1 can be derived from Proto-Yuulngu, and that these two proto-languages differ sharply in morphology, we ought to be able to recognize the genetic gulf between A1 and Y1 by careful analysis of their lexical sharings, even if we had no data at all from other Yuulngu or prefixing languages.

In short, I emphasize strongly that even massive lexical diffusion of the sort seen here does not mean the collapse of traditional, genetically-based historical linguistics. It is possible to identify diffusional patterns through appropriate statistical techniques and to arrive at reasonable conclusions about genetic subgrouping. However, if we reduce sharing indices for each language pair to a single aggregate figure, as in traditional lexicostatistics, it will be exceptionally difficult to avoid making serious errors.

I observe, finally, that although the sharing indices above are rather lower, on the whole, than those in Table 1, this does not by itself mean that the items on the 100-word list are especially immune from borrowing. The lower figures result partly (though not entirely) from the fact that more stringent criteria for acceptance of matches were used; the words not only had to be historically related, but also had to have identical meanings and had to be the common words for the concept in both languages.

11. SOCIOLINGUISTICS. White settlers and missionaries became significant in this part of Arnhem Land in various stages from the turn of the century until 1952, when the A2 people were settled at an Anglican mission. Each EG (ethnolinguistic group, roughly a linguistically defined ‘tribe’ with no organized political structure) consisted of from fifty to a thousand persons, usually under three hundred. EG’s in the far northeast (in the Yuu. area) were larger than those for A1 (less than one hundred), A2 (perhaps two hundred), and B1 (perhaps two hundred).

Currently observable sociolinguistic processes involve bilingual diglossia in most settlements. B1 is recently extinct; A2 is the dominant language at Numbulwar Mission; A1 is becoming extinct, and is down to a few fluent speakers; Y1 is still thriving, being dominant at one settlement and represented at others. English is now important for all except the oldest Aboriginals, though least so for Y1. In addition to standard English, various forms of a widespread English-
based creole are spoken, particularly at one settlement in this area. Current sociolinguistic processes differ significantly from pre-white sociolinguistics; diffusion is occurring between surviving Aboriginal languages and standard and creole English, but its analysis and interpretation require separate study, and are of limited relevance to the understanding of pre-white diffusional processes responsible for the data in this paper. Fortunately, some basic data on traditional sociolinguistics can be reconstructed from ethnographic data, interviews, and the like. Although our information on these points is incomplete and generally resistant to quantification, it helps to explain the remarkable extent of diffusion in this region.

Each EG was defined by language, rather than by residential or political unity. Clans and subclans (often around forty persons) were of greater social and ritual significance; marginal clans were sometimes regarded as 'half' members of two distinct EG's; all forms of social relationship extended across, as well as within, EG boundaries. Demography was seasonal and fluid; fairly small bands, existing during the wet season and early dry season, coalesced later into larger congregations (roughly corresponding to an EG in some cases, but always with visitors from neighboring areas)—this occurred toward the end of the dry season, around large drying swamps which provided root foods. Major rituals and other large-scale social interaction occurred at this time. Permanent dwellings and villages were absent, but certain locations near swamps were habitually occupied by roughly the same people at the appropriate time each year.

Marriage was regulated by incest prohibitions within clans and larger social classes (moieties or semi-moieties), by a classificatory kinship system which identified preferred and possible spouses through abstract genealogical connections, by rightful claims based on concrete genealogical relationships resulting from marriages in earlier generations, and by political considerations favoring marriages which linked clans and bands neither too close nor too distant from each other. EG boundaries appear to have played no role whatever in marital politics, and the effect of the other factors was to insure a high rate of inter-EG marriages. An average EG with four clans and two hundred persons would have around 50% inter-EG marriages; hence half the girls born into the EG would be sent at puberty to join husbands in other EG's, and half the wives obtained by men in the EG would be imported from neighboring EG's. Each EG was contiguous to approximately four other EG's, and had some marital and social relationships with all of them; but typically, special relationships existed between particular pairs of clans, resulting in a predominance of marital exchanges linking one EG to one or two other particular EG's. My evidence on this point suggests especially prominent exchanges between A₁ and (much larger) Y₁, and between A₂ and B₁. Broadly, density of marital and ritual relationships correlates well with density of diffusion.

Substantial cultural uniformity in the area points to relatively stable EG boundaries over long periods of time—even though, at an earlier time, the Yuulngu and prefixing languages came together because of population movements. Systems of social classification (clans, moieties etc.), ritual, mythology,
and semantic structure show no major breaks corresponding to genetic lin-
guistic divisions, though there are many local idiosyncrasies.

Balancing the tendency for social relationships to cross EG boundaries was
a kind of cultural centripetalism and ethnocentrism. Efforts by anthropologists
(Peterson 1976a, and especially Tindale 1976) to connect linguistic and cultural
boundaries with external environmental barriers have had limited success, and
are virtually useless in this part of Arnhem Land. Centripetalism was shown
by the dread displayed toward unfamiliar territory, by attributions of sorcery
and pathological personalities to Aboriginal groups on the fringe of the social
horizon, by intense possessiveness regarding one’s own clan territories and
associated myths and rituals, and by studied disinterest in those of groups to
which one was not related by specific genealogical ties (Biernoff 1978). For an
individual, centripetalism did not crystallize around EG boundaries as such,
since an individual’s relationships were to a particular set of clans to whom
he had genealogical links (the patrilineal clans of his Mo, MoFa, FaMo, and
spouse, in addition to his own). However, from a collective perspective, the
establishment and maintenance of sharp language boundaries can be seen as
a mechanism to encourage narrow circumscription of the social horizon. Since
there was no wide-ranging regional lingua franca, even if every individual spoke
three languages, this only meant that one could communicate directly by a
shared language with perhaps a thousand persons, in a region as small as 5000
square miles (higher figures would apply in the Yuu. region).

It is particularly important to stress the relationship between a man and the
clan of his Mo (and to some extent MoMo). In addition to the personal kin
relationship to the Mo and her siblings, a man had an institutionalized rela-
tionship called juygayi (often translated ‘manager’) to all men in the Mo’s clan,
and to the clan’s sacred territories and rituals. Only the juygayi could paint
the clan’s totemic designs on the clan’s ritual dancers; almost every rite in
every ritual required the cooperation of clansmen and their juygayi; it was the
juygayi who prescribed punishment for ritual transgressions by the clansmen.
Of course, this official, institutionalized relationship was merely one aspect of
a more comprehensive social relationship, beginning early in a boy’s life when
his Mo’s clansmen (as well as his own) began to train him in their ritual and
mythological traditions. A man said, of his Mo’s clan country, ‘It gave birth
to me’ (A2 ŋungu-yaba-ń), and his spiritual relationship to it was as intense as
his relationship to his own clan territory.

For a woman, the Mo’s clan was less important. Instead, having been be-
stowed (perhaps years before her birth!) as future wife to a particular man,
even as a girl she might be sent to familiarize herself with the man and his
relatives. She would move permanently to join him at (or just before) puberty;
and sometimes he would insist on having her join him well before that, in order
to have her under his control and assure the fulfillment of the bestowal promise.
If she had a younger brother in her own clan, he might be sent to visit his
sister’s husband for several months as a child; and it would be the sister’s
husband who would hold the boy during his circumcision.
Such relationships, and others (such as between a man and his MoMoBr and his clanmates), were important in this connection because they often crossed EG boundaries. If a boy’s Mo was from another EG, bilingualism could begin within the nuclear family itself (one language spoken to Mo, the other to Fa), though the Mo probably used both her native language and that of her husband. In any event, a boy would learn some of his Mo’s language, and would use it off and on throughout his life when dealing with his Mo’s clanmates. He might well learn one or two other languages because of his relationships to his wife, his MoMoBr, and other persons (along with their clans and close kin).

Certain other features of southeastern Arnhem Land society can be mentioned here simply to indicate that they are not likely to have been significant factors in promoting lexical diffusion. First, the existence of taboos on the pronunciation of names of recently deceased persons, and of phonetically similar words, might be thought to provide a key to the unusual extent of lexical diffusion: words which cannot be uttered in an extended mourning period are temporarily replaced by alternative words (including borrowings from neighboring languages), and in some cases the original word is dropped permanently. Second, the existence of various kinds of special registers (song language, respect language used chiefly in the presence of real and classificatory WiMo and WiMoBr, and ritual language) might be thought to have a destabilizing effect on the lexicon over time, and to facilitate borrowings from neighboring languages.

I reject these as sufficient explanations for the peculiar extent of lexical diffusion in southeastern Arnhem Land. In the first place, both the death taboo and the special registers are very much less developed and less productive in this part of Australia than in most others. Death taboos which I observed typically lasted less than two years, even for the personal names of the deceased; and taboos on phonologically similar common words occurred, but were much briefer. After a respectable interval, such common words came back into regular use. Personal names are typically special totemic epithets taken from songs, and in most cases are not ordinary lexical items; even those which do happen to be common nouns are in the flora-fauna area. At most, explanations for lexical diffusion based on the death taboo could only account for a limited percentage of the sharings in this semantic domain. Much more significant death taboos occur in other parts of Australia, requiring languages to maintain elaborate stocks of synonyms for common nominal concepts, so that one noun may be temporarily taboo without creating problematic lexical gaps.

Similarly, special registers are highly limited in this area, in contrast to several other parts of Australia. We find no elaborate respect lexicon, and no special ritual language in the usual sense. Instead, we find highly frozen ritual songs, consisting almost entirely of unanalyzable and rather long epithets for various totems. Although these songs have been borrowed entire, across language boundaries, epithets in them are so unlike ordinary words in the local languages that they could hardly spread into the general lexicon. Even were
such spreading formally possible, it would be culturally impossible—since ritual song epithets are by definition secret, known only to, and jealously guarded by, initiated adult males.

In addition to the secret ritual songs, public songs exist. For the B₁ and B₂ groups, these are similar in formal structure to the secret songs, are only marginally analysable in terms of local grammars, and are highly unlikely to be conduits of lexical diffusion involving the general vocabulary. In the A₁, A₂, A₃, and Yuu. (Y₁ etc.) areas, a distinctive form of popular song is found, sometimes more or less analysable and interpretable in terms of the ordinary languages. However, a given EG need not sing in its own language. A₁ songs of this type, crucially, are in the A₂ language; so these songs cannot possibly explain the heavy borrowing of Y₁ words into A₁ (along with a more modest number of A₂ borrowings). Similarly, Y₁ songs are performed in a more northerly Yuu. language (of the Y₂ or Y₃ type). In addition, special songwords which are not also ordinary words tend to be unusually long epithets; they constitute a semi-specialized song vocabulary which is normally kept apart from the ordinary lexicon, though not so rigorously as in the case of secret ritual songwords.

It is true that songwords in one language are often borrowed from the common vocabulary of another language. Thus A₂ songs contain a fair number of common Y₁ or other Yuu. nouns, functioning only as specialized songwords in A₂. For example, one honey bee sp. is called birkudja in Y₁, and this shows up as songword birguda in popular A₂ songs; but the ordinary A₂ word nabi for this sp. is not shared with Y₁. Because totemic songs typically require a large number of synonymous epithets for a single species, it is hardly surprising that borrowings abound in popular songs; but there is virtually no evidence that songs are a primary channel for the diffusion of common words in one language into the common vocabulary of another. Even the type of diffusion is different in songs; e.g., while A₂ shares only one or two of its biological life-form terms like ‘fish’ or ‘bird’ with Y₁, in songs A₂ has borrowed several Y₁ life-form terms (and a few others) as classifying particles, so that many A₂ songwords for fish spp. begin with guya (borrowed from Y₁ and general Yuu. guya ‘fish’). Although the study of songwords and their diffusion is interesting in itself, my work in this area has convinced me that songs are not a primary vehicle of diffusion for ordinary vocabulary. Most songwords are epithets for flora-fauna terms, so such explanations cannot account for the broad range of lexical diffusion seen above; and even for flora-fauna, the nature of song language makes me very skeptical of suggestions that this is a central avenue for borrowings.

More generally, special registers (of which song language is a case) typically discourage rather than encourage diffusion of common vocabulary from one language into another. Since ordinary words in one language are often borrowed into the unstable special-register vocabulary of a neighboring language, and since there are strong internal pressures on the latter to maintain clearly distinct ordinary and special-register vocabularies, the typical pattern should be that
speakers of the borrowing language will keep the borrowed word in their own special registers, and will not permit it to intrude into their ordinary vocabulary.

I summarize the major relevant features of pre-white society in southeastern Arnhem Land as follows: stable, long-term maintenance of language boundaries, separating EG’s typically with fewer than 300 persons each; frequent intrafamilial bilingualism, with the Fa (and most other persons in the subsistence unit) speaking one primary language, and the Mo being a native speaker of a different language; EG’s each having a territorial base, with boundaries sometimes fuzzy, and with social and marital relations radiating outward in several directions (not always of equal importance); and particular bilingual or multilingual capabilities of individuals, varying even within small groups according to particular kinship relations and personal residential histories. Other important features are the virtual absence of asymmetrical functional or prestige differentiation of the languages, and the tendency for cultural centripetalism to operate on speech only in terms of language boundaries, while not blocking diffusion of individual words (or even bound grammatical morphemes).

12. WIDER COMPARISONS. It would be presumptuous to propose a comprehensive theory of lexical diffusion from a single case study, particularly one from such an unusual sociolinguistic situation. My goal here has been to describe a single interesting case in such a way that readers with expertise in other regions can identify similarities and differences. However, it is important that some effort be made to contextualize these data within the framework of a more comprehensive comparative theory of lexical diffusion.

Scotton & Okeju 1973, in a rare case study sufficiently detailed to permit comparison, report that isolated fringe groups of the Ateso-speaking population of Uganda and Kenya have adopted many lexical items from neighboring or hierarchically superimposed languages, including Swahili. Those authors emphasize that some ‘core’ vocabulary items are involved; however, from the perspective of Arnhem Land, the Ateso case is most notable for the limited extent of lexical diffusion. Almost all the ‘core’ borrowings involve independent words or phrases presenting no problems of morphological adaptation: (a) particles and similar forms like ‘but’ and ‘how much?’; (b) fixed phrases like ‘I don’t know’, borrowed whole; (c) words for times and calendar dates; and (d) fixed greetings. Virtually no high-frequency noun stems are borrowed; and verbal borrowings are very rare, involving new concepts rather than diffusional replacement of old lexical items. Some of this borrowing might more properly be called code-switching (see Pfaff 1979 for a review of terminology). By contrast, A₁–Y₁ sharings include many high-frequency nouns and verbs.

In Newman’s study of Bella Coola (Salishan) and its recent diffusional relationships to Wakashan and Athabaskan languages, he stressed the extent of diffusion from Wakashan into Bella Coola. However, he pointed out that diffusion in the reverse direction was sharply limited; he emphasized that, despite much intermarriage, diffusion between Bella Coola and Athabaskan was almost non-existent in either direction; and even for diffusion from Wakashan to Bella
Coola, he distinguished between ‘massive’ diffusion of ecological (mostly flora-fauna) terms (30% sharing between Bella Coola and Bella Bella) and core vocabulary: ‘Bella Coola borrowed almost nothing of its basic stock of words from Wakashan’ (207). Although Newman’s article, like that of Scotton & Okeju, presents data considered to exhibit striking degrees of lexical diffusion, the A1–Y1 data show far more over-all diffusion, more even distribution of borrowings across semantic and word-class categories, and extensive diffusion of bound grammatical morphemes and phonological patterns (Heath 1978a).

Haugen’s thorough linguistic study of American Norwegian (1953), abbreviated AmN, concludes that lexical diffusion from English into immigrant AmN, even in the second and third generations, never reached massive proportions if code-switching is distinguished from (assimilated) borrowings: fewer than 5% of Norwegian items on a basic thousand-word list were replaced by English lexical items. Haugen summarizes thus:

‘... individually and sporadically any word can be borrowed; the writer has even heard the conjunctions and and but used in a N context by some speakers. But in general the speakers of AmN were not borrowing beyond the number of words that one might expect; the core of their language was wholly Norwegian, and in no way the kind of non-descript pidgin that some have suggested.’ (97)

I have some personal experience with bilingual situations in other parts of the world (Basque and Romance; Moroccan Arabic and French; Mississippi Choctaw and English); but neither in this direct experience nor through the literature have I encountered a region characterized by stable multilingualism in which lexical and other linguistic diffusion has reached the extent seen in the A1–Y1 pair. Even in other pairs such as A2–B1 and A2–Y1, in which diffusion is less substantial than for A1–Y1, it is high by normal standards.

It might be argued that diffusion in the A2–B1 and A2–Y1 pairs, while high quantitatively, is nonetheless similar in type to that found in other cases—insofar as diffusion is high in domains like flora-fauna, but very low or non-existent in closed verb classes, and fairly low in 100-word list items. However, the A1–Y1 case differs more radically. It seems that when pressures for lexical diffusion build up past a certain point, the distinction between core (non-diffusable) and peripheral (diffusable) lexicon tends to break down, except where the way is blocked by insurmountable problems of morphological adaptation. Hence a considerable number of 100-word list items have been diffused between A1 and Y1, and only monosyllabic non-thematic verbs have avoided extensive diffusion (a fact which is partly meaningless, since several of these verbs were already cognate and still phonologically similar because of common inheritance from Proto-Australian, so that diffusion was unnecessary to achieve sharings). Indeed, A1 (or rather Proto-A) solved the problem of morphological obstacles for diffusion of verbs by borrowing thematizer *-du- from Y1 (or its ancestors) along with the root form system, whereupon roots of this highly productive thematic class could easily be borrowed back and forth. Although the ancestors of A1 and Y1 must have had only a handful of shared lexical items when they came into contact, diffusional pressures have been so intense and so long in operation that a kind of quantum jump of lexical diffusion has occurred,
whereby morphological barriers were demolished and even core vocabulary became highly susceptible to diffusion. It is likely that the particularly small size of the $A_1$ EG (less than a hundred persons), and the consequently very high rate of outmarriage of this group (more than 50%) was responsible for the exceptional rate of diffusion.

Although complete information is not readily available on Ateso, Bella Bella, and AmN, it seems that intrafamilial bilingualism, with the two parents having different native languages, has been much less extensive than in Arnhem Land. In addition, there is evidence that the other cases involved functional asymmetries, or self-conscious avoidance of linguistic contamination (see also Gumperz & Wilson 1971 on the latter point). Consequently, Arnhem Land shows not only a higher general rate of diffusion, but also essentially symmetrical diffusion. We cannot deal extensively with the question of directionality here, since we would need an extended philological discussion of each lexical set, as well as data from ten or so additional languages (most of which currently lack adequate dictionaries). However, preliminary study suggests that asymmetries in the direction of diffusion are small ($A_1$ has borrowed somewhat more from $Y_1$ than vice versa, but this probably results from the greater population of the latter EG, rather than from cultural prestige factors). Directionality was examined in cases of grammatical-morpheme diffusion in Heath 1978a, and the general conclusion there was that such diffusion was approximately symmetrical for most language pairs.

I want to distinguish sharply between the Arnhem Land case (and typologically similar cases), on the one hand, and two other kinds of contact situation, on the other. First, southeast Arnhem Land cannot be usefully compared to pidgin or creole zones of any kind. Arnhem Land (prior to white contact) was a stable multilingual zone, completely distinct sociolinguistically from plantations, ports, or even colonial societies. Despite intensive lexical and grammatical diffusion, Arnhem Land languages have shown no tendency whatever toward morphological simplification or regularization, nor toward morpheme-by-morpheme intertranslatability. Indeed, in several instances (Heath 1978a) diffusion of morphemes has increased morphological irregularity and/or complexity; and $A_1$, $A_2$, and $A_3$ in particular are famous among Australianists for their astonishingly intricate morphologies.

Second, the Arnhem Land case must be sharply distinguished from diffusion in dialect chains or across weak language boundaries. To the north, part of the Yuu group of languages can be described as a dialect chain (Morphy 1977); and much of the interior west central part of Australia is a vast dialect chain known as the Western Desert language. However, the languages I am dealing with here have sharp boundaries, with almost no significant internal dialectal variation. $A_1$ and $Y_1$, despite sharing 50% of over-all vocabulary (somewhat less if text frequency is considered), are mutually unintelligible at first, and have different morphological systems. ($A_1$ has noun-class prefixes for nouns, and intransitive or transitive pronominal prefixes and numerous derivational prefixes with verbs, but $Y_1$ has none of these; both languages have verbal inflectional suffixes and case suffixes, but they are distinct except for a couple
of diffused case suffixes.) A2 speakers who have not specifically learned neighboring languages (A1, B1, A3, Y1) find all of them entirely unintelligible. Lexical and grammatical diffusion in this region has therefore involved the crossing of sharp linguistic boundaries; grammatical diversity from one language to the next in Arnhem Land is greater than in any other part of Australia.

Hence the most useful direct comparisons will be with studies of intimate contact (including extensive intermarriage) among hunter/gatherer or other tribal societies, preferably with small populations, in certain other parts of Australia (Cape York Peninsula), New Guinea, Africa, and North and South America. Unfortunately, Cape York shows less initial linguistic diversity than Arnhem Land because genetic relationships are closer; and some languages have died out without being adequately recorded.5 Furthermore, in the other areas mentioned, tribal populations are normally far greater than the 200–300 per EG typical of Arnhem Land (let alone the less than 100 for A1); and we often find prestige and other cultural factors which reduce the extent of diffusion. We can hope for the appearance of case studies similar to this one from these other parts of the world, but it is at least possible that the A1–Y1 pair will stand as a unique, extreme case.

REFERENCES


5 Important work on Cape York is currently being undertaken by Peter Sutton and Bruce Rigsby (Sutton 1978, Sutton & Rigsby 1979), who belong to an interdisciplinary research team including social anthropologists. However, their material is not yet directly comparable to mine from Arnhem Land, since they have been concentrating on the development of appropriate theories of speech-community and social-network structures. At the present time, more properly historical linguistic studies on Cape York are insufficient to permit close comparisons with my own results, although we can expect this situation to change as more of their material is completed and published.
A CASE OF INTENSIVE LEXICAL DIFFUSION


—. 1980d. Nunggubuyu myths and ethnographic texts. Canberra: AIAS.


—. 1976b. (ed.) Tribes and boundaries in Australia. Canberra: AIAS.


[Received 16 January 1980; Revision received 8 July 1980; Accepted 18 September 1980.]