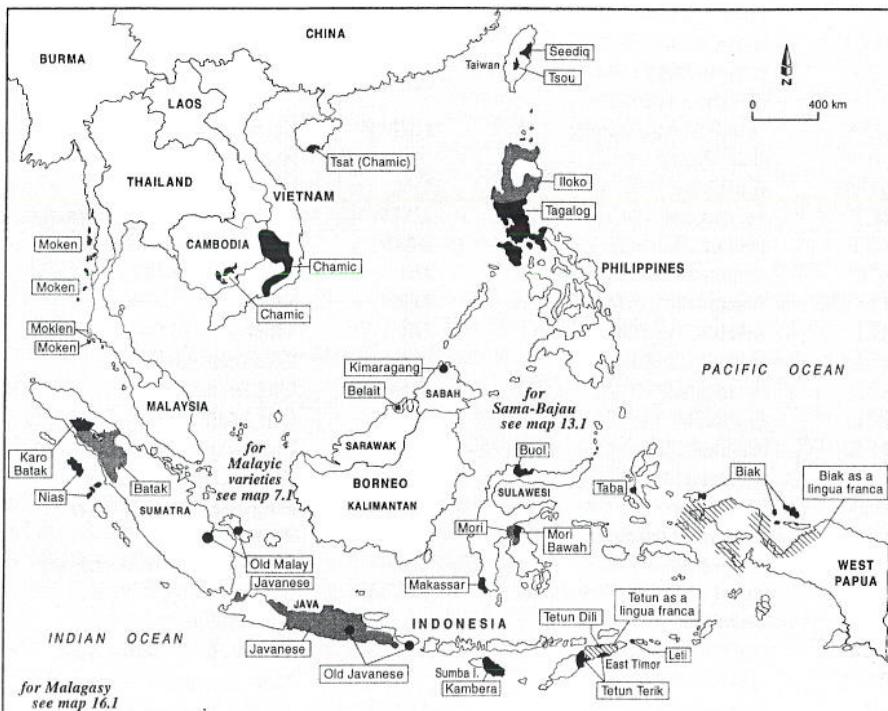


CHAPTER ONE

THE Austronesian Languages of Asia and Madagascar: A Historical Perspective

Alexander Adelaar



MAP 1 LANGUAGES DISCUSSED IN THE DESCRIPTIVE CHAPTERS OF THIS VOLUME

1 INTRODUCTION

The Austronesian language family has about 1,200 members, which together are spoken by some 270 million people. It is arguably the largest existing language family in terms of the number of its member languages (making up for 20% of the world's languages) and it is second in terms of its geographic spread (after the Indo-European language family). Austronesian languages range from Malagasy (in Madagascar and on the island of Mayotte) in the western part of the Indian Ocean to Rapanui or Easter Island in the southeastern part of the Pacific Ocean, and from the Formosan languages in Taiwan and Hawaiian in the northern reaches of the Pacific to Maori in New Zealand. The Philippines, Brunei, Polynesia, Micronesia, Melanesia and most of Indonesia and Malaysia are traditionally Austronesian-speaking, and there are pockets of Austronesian speakers on the Southeast Asian mainland in Thailand, Vietnam, Cambodia, Burma and Hainan (People's Republic of China). As a result of colonial administration and its aftermath there are also pockets of Austronesian speakers in Sri Lanka (Sri Lankan Malay), South Africa (Malay but now extinct), Netherlands (Moluccan Malay, Javanese and various Moluccan languages), Surinam (Javanese), Australia (Malay, various East Timorese languages) and Portugal (various East Timorese languages).

This chapter introduces the reader to various background aspects of the Austronesian languages of Asia and Madagascar which will not be individually discussed in the chapters that follow. It contains basic source literature on Austronesian linguistics (§2), a short explanation about writing systems (§3), basic information on Proto-Austronesian (PAn) and Proto-Malayo-Polynesian (PMP) grammar and how it might have evolved further into the present-day languages (§4), recent developments in the classification of the Austronesian languages in question (§5), Austronesian homeland and prehistory (§6), and language contact affecting the Austronesian languages west of the Pacific (§7).

Chamorro and Palauan (Belau) are spoken in Micronesia (in the Western Pacific area) and fall geographically outside the scope of this volume. However, they are included in the discussion on classification in this chapter because they are not Oceanic languages.

2 IMPORTANT REFERENCE WORKS CONCERNING ASTRONESIAN COMPARATIVE HISTORICAL LINGUISTICS

Modern Austronesian comparative historical linguistics begins with Dempwolff (1934, 1937 and 1938). This is a reconstruction of PMP phonology following the comparative method. It is based on a systematic comparison of eleven languages, and it includes a PMP lexicon. Although it is in many ways out of date, knowledge of this study is indispensable for a good understanding of the issues and further developments in the field that have emerged since, and it effectively incorporates the works of previous comparatists in the field of PAn phonology and lexicon.

Important expansions and refinements of Dempwolff's phonology are the works of Dyen (1947a, 1947b, 1951, 1953a, 1953b, 1962) and, later on, of Dahl (1976 and 1981). Dyen's merits lie particularly in the inclusion of Formosan languages in his comparisons (1963, 1965a), and in his study of the so-called Proto-Austronesian 'laryngeals' (which are in fact PAn *h, *ʔ (glottals), *q (a uvular stop), and *S; Dyen 1953a). His lexicostatistical classification of Austronesian languages became the subject of much controversy but is nowadays outdated (Dyen 1965b). Dahl (1976) is a balanced criticism of Dyen's work; it focuses on phonological issues but it also discusses grammar and homeland theories and it gives a short historical overview of Austronesian studies. Dahl (1981) is a further elaboration on Proto-Austronesian phonology. Important refinements to PAn and PMP phonology are Blust (1990, 1993a, 1994a). Blust also made major contributions to the reconstruction of PAn lexicon; see his publications in *Oceanic Linguistics* (Blust 1970, 1980a, 1983–4, 1986, 1989). His unfinished *Comparative Austronesian Dictionary* is already partly available on request from the author (Blust to appear). Ross (1995) gives a clear summary detailing all major current issues and developments regarding Proto-Austronesian phonology, morphosyntax and classification.

In spite of comparative morphological studies of MP languages by various previous scholars (e.g. Humboldt 1836–9, Brandstetter 1893), Dempwolff did not venture into a grammatical comparison of MP languages, nor did other scholars for a long time after the publication of his reconstruction. It was not until 35 years later that such a reconstruction (for PAn) was undertaken, and Wolff's reconstruction (1973) was something of a *tour de force* considering the lack of previous systematic attempts based on a representative sample of Austronesian languages. He established a 'Philippine-type' morphology with verbs marked for aspect, mode and semantic role of subject. Several other studies of importance followed, most notably Starosta, Pawley and Reid (1982) and Ross (Ross 1995). The former claimed that the Austronesian verbal voice markers were originally nominalizing affixes which only developed into verbal voice markers after the break-up of PAn.

A recent book edited by Wouk and Ross (2002) contains several chapters dealing with PAn and PMP morphosyntax as well as with the changes that happened after the split-up of PMP. One major change appears to have been the development of an 'Indonesian-type' morphosyntactic model alongside the Philippine-type model inherited from PAn. Here again, the chapter by Ross provides the most comprehensive overview of changes that have taken place since PAn (a topic which will be dealt with in detail in Zobel to appear). Blust (in press) discusses the PAn affixes *ka- and *ka- -an, and various causative prefixes including *paka- (generic causative of stative verbs) which Zeitoun and Huang (2000) claim to be a concatenation of (causative) *pa- + (stative) *ka-. Blust has also written a monograph on Austronesian root theory (1988) and articles on reduplicated monosyllables (Blust 1976a), vocatives (Blust 1979) and Ca-reduplication (1998a).

Recent achievements in the fields of Austronesian language classification will be treated in §5.

In the area of culture history, there are a number of leading publications by Bellwood and by Blust. Bellwood (1997) is a comprehensive multidisciplinary study of the Austronesian past combining research data mainly from archaeology, linguistics and biological anthropology (for further discussion, see §6). Blust's publications cover Austronesian social organization (1976a, 1980b), PMP phratry dualism (1980c), the linguistic value of the Wallace line (1982), the settlement of mainland Southeast Asia (1994b), and the Austronesian homeland and subsequent spread of Austronesian speakers (1984–5, 1999). Other publications of interest are Mahdi (1988, 1994a and 1994b) and various papers in Pawley and Ross (1994).

A number of introductions to the anthropology of Austronesian societies have recently appeared or are about to appear: Fox (1993, 1997), Bellwood, Fox and Tryon (1995), Fox and Sather (1996), Vischer (in press) and Reuter (forthcoming).

Some atlases and wordlists have proven to be of great importance to the study of Austronesian linguistics. Wurm and Wilson's (1975) English-Austronesian finder-list is a very rudimentary tool for finding PAn, PMP and lower order etyma that had been reconstructed before 1975. The list has now become largely outdated. Wurm and Hattori's linguistic atlas of Southeast Asia and Oceania has detailed maps with explanatory notes on the back. Being the first linguistic atlas of such a wide scope, it contains various errors and drawbacks, as pointed out in Steinhauer (1986), but this does not prevent it from being an extremely useful tool in the study of Austronesian linguistics (an updated version is to appear). It is now supplemented by a linguistic atlas including a much larger part of the world and dealing specifically with language contact (Wurm, Mühlhäusler and Tryon 1996).

Tryon *et al.* (1995) is basically an 'annotated dictionary of synonyms for some 1200 lexical items in 80 different Austronesian languages' (Tryon 1995:1). It has for each language an introduction of varying length and depth providing a map, facts and figures, important bibliographical references and phonological and morpho-syntactical information. It also contains a listing of Austronesian languages which is excerpted from Grimes' *Ethnologue* and as such gives basic data on alternative names, dialects, classification and location, and general population figures for the area in which each language is spoken. No attempt has been made here to reproduce these data in detail as the same information is contained in the *Ethnologue*, where it is continually updated (Grimes 2000). Tryon's introductory chapter treats the language family in general and includes an overview of the history of Austronesian studies (a topic not covered here in any major way) (Tryon 1995). It also provides basic information on language distribution, ranking of languages according to numbers of speakers, classification, phonology and grammatical typology, settlement and migrations. And finally, it contains Ross's overview of important trends and issues in Austronesian comparative historical linguistics already mentioned above (Ross 1995).

3 WRITING

Except for Vietnamese, the major Southeast Asian languages have (or had) scripts that were derived from the Grantha- or Pallava-script from South India, which is indirectly related to the North Indian Devanagari script.

In Malaysia, Indonesia and the Philippines, these scripts have almost entirely been replaced by the Roman script. The oldest Austronesian written records are Cham stone

inscriptions which date from the 4th century AD and were found in the former Champa empire (nowadays part of Vietnam, see Thurgood, this volume). The earliest Malay inscriptions date from the seventh century AD (see Mahdi, this volume); those in Javanese and Balinese date from the ninth century AD (see Ogleblin, this volume).

Pallava-derived scripts read from left to right, and are neither alphabets nor syllabaries. Each letter (or *aksara*) denotes a consonant followed by a default vowel *a* (some scripts use a default vowel other than *a*); if in a word a consonant is followed by a vowel other than this default vowel, this is indicated by small vowel signs on top, under or after the consonant (in some scripts vowel signs can even occur before – or before and after – the consonant). In some scripts, a dummy letter is used in order to write words beginning with a vowel. If a consonant occurs word-finally, this is indicated by a special following sign (sometimes called the ‘vowel-killer’). Finally, clusters of consonants are indicated by writing the consonants under each other, as in Javanese, or by using the vowel-killer, as in the (North Sumatran) Batak script and the (South Sumatran) Rencong script, or they are not indicated at all. The last causes considerable phonological uncertainty and occurs in the Philippine and Buginese-Makassar scripts, which lack a vowel-killer. In Buginese texts, for instance, consonant gemination, final final velar nasals and final glottal stops are not indicated; furthermore, homorganic nasal + consonant clusters can be indicated but this is often not done (see also section 2.5 in Jukes, this volume). The Javanese and Batak scripts are very precise in contrast (although the Batak script does not indicate, (phonemic) stress in Toba Batak).

The Javanese script is also used for writing Madurese, Sundanese and (in a variant form) Balinese. Versions of the Buginese-Makassar script were (still are?) also in use in Southeast Sulawesi (for Wawonii), on Sumbawa (for Bimanese and Sumbawanese) and on Flores (Endenese) (Noorduyn 1993:535).

The different shapes of Pallava-derived scripts have probably developed through the use of various writing materials: for instance, the use of bamboo led to cuneiform letters in the Rencong script, whereas the use of palm-leaf allowed round-shaped letters in the Javanese script. Bataks and people in the Philippines used writing mainly (but not exclusively) for recording magico-religious and medicinal knowledge. The Javanese, Makassarese and Buginese used their scripts for more general purposes.

Pallava-derived scripts are hardly used any more nowadays, although there are still people practicing them in Java, Bali, the Batak regions, South Sulawesi, Mindoro and Palawan (Postma 1971:1). Among the Mangyan in Mindoro there have been initiatives to revive the local variant, which is now equipped with an (unauthentic) *pamudpod* (vowel-killer) to distinguish final consonants and consonant clusters (*A Primer* 1986: Introduction).

In Muslim areas (Malaysia, Southern Philippines, Indonesia), Arabic script came into use in the thirteenth century at the earliest. In East Madagascar it has been in use at least since the fourteenth century. It is often assumed that the Arabic script as used for Malay is originally an Indo-Persian adaptation, but this is not clear from a cursory inspection and remains to be investigated.

Originally designed for a Semitic language structure, the Arabic script (in its various adaptations) causes some inherent orthographic problems for Austronesian languages, where vowels are phonemically more salient and (usually) more variegated than in Arabic.

Some adaptations of the Arabic script are in reality adaptations of Jawi, the Malay version of the Arabic script, or (possibly) of Pégón, the Javanese version. For instance, the Tausug adaptation (used in the Sulu archipelago) follows Malay closely in the way

(such as *c*, *n* and *y*) and in the way it interprets Arabic sounds that do not occur in Tausug or Malay (e.g. the letter for Arabic *f*, which is pronounced [p]). Furthermore, some Tausug technical terms for the vowel diacritics used in Arabic script are derived from Malay, not from Arabic, e.g. *baris* is a general term denoting diacritic vowel signs (< Malay *baris* ‘1. line; 2. vowel sign’); *dapan* denotes the vowel sign for *u* (< Malay *dapan*, ‘in front’, cf. Arabic *dammah*) (cf. Cameron 1917). There are also various indications that the Sorabe (or *Volay 'Onjatsy*) script reached Madagascar from Indonesia, and not from East Africa or the Arabian peninsula (Adelaar 1995a:332–339). For instance, the word *soratra* ‘writing’ is a Malay loanword (Adelaar 1989, 1995a). Metaphors like *reni soratri* ‘mother of writing’ for main letters and *zana 'tsoratri* ‘child of writing’ (Dahl 1983:71 fn.7) for diacritics are also found in the terminology for main letters and diacritics in Batak (cf. respectively *ina ni surat* and *anak ni surat*) and Buginese (cf. respectively *ina surə?* and *ana? surə?*). The use of the same metaphors here indicates that, even if the Malagasy did not have a Pallava-derived script, they seem to be an exponent of the writing traditions of the Malays and other Austronesian peoples who throughout the centuries have adapted Pallava-, Arabic- and European writing systems in a continuous tradition of literacy, sometimes carrying over writing conventions from one system into the next.

4 PAN AND PMP LINGUISTIC HISTORY: SOME BASIC INFORMATION

Blust (1999:34) proposes a PAN consonant system which is shown in Table 1.1 in main outline:

TABLE 1.1: PAN PHONOLOGY

	Labial	Dental	Palatal	Palato-velar	Velar	Glottal
Voiceless stops	p	t	c		k	?
			C			
Voiced stops	b	d	z	j	g	
Nasals	m	n	ñ		N	ŋ
Voiceless fricatives		S	s			h
Lateral		l				
Trills		r			R	
Semivowels	w		y			

PAN has moreover the following vowels: *a, *ə, *i, *u. It has the following diphthongs: -iw, -uy, -ay, and -aw.

Earlier interpretations of the PAN phoneme system often distinguish three *d's (Dahl 1981, Ross 1995). However, their evidence is based on Puyuma and Paiwan reflexes only. They occur in loanwords (usually borrowed from Puyuma into Paiwan) and should be discounted (Blust 1999:47–51).

The evidence for PAN *c, *g and *r is weak (Wolff 1988; cf. also Mahdi 1988 regarding *r). Ross (1995:58–59) accepts *c for PMP only, but argues that there is sufficient evidence from Formosan languages for PAN *g and *r, even if they must have had a low functional load.

The nasal *N is sometimes represented as a voiceless lateral *L.

The phonology of PMP is somewhat simpler than that of PAN (except for the acquisi-

TABLE 1.2: PAn VERBAL MORPHOLOGY

	Actor	Patient	Location	Circumstantial
INDICATIVE				
Neutral	<um>V	V-an	V-an	Si-V
Perfective	<umin>V	<in>V	<in>V-an	Si-<in>V
Durative	<um>-R-V	R-V-an	R-V-an	Si-R-V
NON-INDICATIVE				
Atemporal	V	V-u	V-i	áñ-i + V (V-áni)
Projective	<um>V-a	V-aw	V-ay	áñ-ay + V (V-áñay)

underwent the following changes: PAn *C and *t > PMP *t; PAn *S > PMP *h (a glottal spirant which kept distinct from PAn *h); PAn *N and *n > PMP *n.

The grammars of PAn and of PMP are thought to have had a Philippine-type structure with verbs marked for actor-, patient-, local- and circumstantial voice. Their voice markers could be combined with various other affixes indicating tense, mode or aspect. The PAn verbal paradigm proposed by Ross (2002) distinguishes the voices mentioned above as well as neutral-, perfective- and durative aspect and atemporal and projective mode (Table 1.2). Some further comments are in order for an accurate interpretation of this paradigm:

- 1 Although he is very uncertain about the reconstruction of PAn stress, Ross reconstructs oxytone roots (with stress on last syllable) and paroxytone roots (with stress on penultimate syllable). In both cases stress moved one syllable to the right if the root was suffixed.
- 2 There were four formal verb classes: (a) verbs which took the actor infix <um> (as indicated in Table 1.2); (b) verbs without <um> affix in actor voice; (c) verbs consisting of a derived stem beginning with (possibly causative) *pa-; and (d) verbs consisting of a derived stem beginning with (possibly stative) *ka-. Both *pa- and *ka- appeared as *ma- in actor voice after inflection of <um>.
- 3 Locative voice was also used as a Benefactive voice.
- 4 Ross is hesitant about the status of the circumstantial voice in PAn. *Si- (as well as another circumstantial prefix *Sa-) did occur in nominalizations at the break-up of PAn but may have developed into verbal pivot markers only later on, and in individual post-PAn branches. Also, it is probable that the non-indicative circumstantial suffixes developed from independent verbal modifiers and were not yet cliticized to the main verb in PAn.

Typologically, PMP must have been rather similar to PAn. The various differences between PAn and PMP grammar as discussed by Ross (2002) seem to have had little structural impact. In the pronominal system (among some other changes) the PAn polite and neutral series of free pronouns merged in one single free set. Furthermore, PMP seems to have had a somewhat more complex set of noun phrase markers.

However, what is more important for further developments in Malayo-Polynesian languages (and what does not appear from Table 1.3 below) is that, in contrast to PAn, PMP had acquired a variety of derivational prefixes including *paN- [+distributive] and *paR- [+durative], which became *māN- and *māR- respectively in active voice. (However, according to Zobel, fossilized forms of *paR- (and possibly also *paN-) are also attested in Formosan languages, Zobel to appear). These prefixes are widely reflected in WMP languages and (in fossilized form) in Oceanic languages.

TABLE 1.3: PMP VERBAL MORPHOLOGY

	Actor	Patient	Location	Circumstantial
<i>Indicative</i>				
Neutral	<um>V	V-an	V-an	i-V
Perfective	<umin>V	<in>V	<in>V-an	i-<in>V
Imperfective	<um>-R-V	R-V-an	R-V-an	i-R-V
<i>Non-indicative</i>				
Atemporal	V	V-a	V-i	V-án
Projective	V-a	(V-aw) [†]	V-ay	

† Ross (2002a) is not certain about the PMP status of the patient projective ending *-aw.

The verbal structure of PAn and PMP has been largely retained in the languages of Taiwan, the Philippines, Sabah, North Sulawesi and Madagascar, which are often referred to as Philippine-type languages (but see HIMMELMANN, TYPOLOGICAL CHARACTERISTICS for a somewhat different definition).

In Indonesian-type languages, i.e. many Austronesian languages spoken in Malaysia and western Indonesia, the PAn voice system is reduced to a simple opposition between actor and undergoer voice. However, different types of undergoers are often distinguished with the help of applicative suffixes, and these undergoers may also appear in subject position if the verb is in undergoer voice. Compare the various derivations of the Indonesian verbal root *tulis* 'to write', which combines with the prefixes *məN-* (actor voice) or *di-* (undergoer voice) and with the suffixes *-i* (locative-oriented) or *-kan* (more general applicative):

- (1) a. *ia mənulis /məN-tulis/ nama-nya*
3s AV-write name-3.GEN
'She wrote her name.' (actor-voice)
- b. *nama-nya di-tulis-nya*
name-3.GEN UV-write-3.GEN
'She wrote her name.' (undergoer-voice)
- c. *aku mənulisi /məN-tulis-i/ amplop itu*
1s AV-write-LOC envelope DIST
'I scribbled on the envelope.' (actor-voice)
- d. *amplop itu ku-tulis-i*
envelope DIST 1s.ACT-write-LOC
'I scribbled on the envelope.' ('The envelope was scribbled on by me.')
(undergoer-voice)
- e. *Ia mənuliskan /məN-tulis-kan/ saya bon*
3s AV-write-APPL 1s receipt
'He wrote me a receipt.' (actor-voice)
- f. *saya di-tulis-kan bon*
1s UV-write-APPL receipt
'They wrote me a receipt.' (undergoer-voice)

However, note that in Indonesian many transitive verbs do not allow all six combinations, and that the derivations with *-i* and *-kan* sometimes have idiomatic meanings (see also EWING, COLLOQUIAL INDONESIAN).

Some languages (e.g. Batak languages, Old Javanese, Balinese, Sasak, several Sulawesi languages) in one way or another represent intermediate stages between the Philippine model and the Indonesian one. Hence it is not quite clear at this point whether it will be possible to establish a clear-cut boundary between the two types, and it is not obvious that their distinction would be a basis for subgrouping arguments (see below).

The historical transition towards the Indonesian model is explained by Starosta, Pawley and Reid (1982) and by Wolff (1996, 2002) through the fact that in Philippine-type (and some other Austronesian) languages, auxiliaries, negators and other preverbal markers can attract person-marking pronouns, while the lexical verb follows in the atemporal form. This is demonstrated with the following two sentences from Cebuano (Philippines), *Gi-tawg-an nako siya* (PST-call-LV 1s.GEN 3s.NOM) 'I called him' and *Wa nako siya tawg-i* (NEG 1s.GEN 3s.NOM call-LV) 'I didn't call him' (Zobel 2000:409). The first sentence is declarative, and is headed by the verb 'to call' which is also marked for tense. The second sentence is a negation; it has the negator as its head, which has the Actor pronoun cliticized to it. The (lexical) verb follows and is not marked for tense. At some point in time, these preverbal modifiers may have been lost, leaving the actor pronoun 'stranded' before the atemporal verb to which it eventually became cliticized or prefixed (Starosta, Pawley and Reid 1982). Or alternatively, the position of the actor pronoun before the atemporal verb may have become generalized to cases in which there was no preverbal modifier (Wolff 1996). There had now arisen an opposition between an actor-oriented neutral verb form and three non-actor-oriented forms of the atemporal verb, and in a last stage, the pattern of three non-actor-oriented forms became generalized over the actor-oriented forms (Wolff 1996). Compare the following schematic development based on Wolff (1996), where (1st person proclitic) *ku=* stands for any proclitic pronoun.

(2)	<i>ku=V</i>	vs.	<i>maN-V</i>	→	<i>ku=V</i>	vs.	<i>maN-V</i>
	<i>ku=V-i</i>	---			<i>ku=V-i</i>	vs.	<i>maN-V-i</i>
	<i>ku=V-an</i>	---			<i>ku=V-an</i>	vs.	<i>maN-V-an</i>

Languages allowing this proclitic actor marking often do not have a full pronominal set to do so: Toba Batak has only first person proclitics (*hu-* '1s' and *ta* '1pi'), Indonesian has only 1s and 2s proclitics (*ku-* and *kau-* respectively), etc.

Although there are many languages showing some or all these Indonesian-type features, it is unlikely that they are inherited from a common (post-PAn) protolanguage. It is more likely that the relevant changes have happened independently or via contact in the various languages or language groups (Himmelmann 2002, Ross 2002, Wolff 2002).

5 THE INTERNAL CLASSIFICATION OF AUSTRONESIAN LANGUAGES

For a full picture of the main issues and developments in Austronesian classifications the reader is referred to Malcolm Ross (1995). His overview basically serves as a vantage point for the present section, which discusses various publications that have appeared since.

The classification of Austronesian languages proposed by Blust (1977 and later publications) is not without its problems (see below). Nevertheless, it is the one that

is referred to most frequently in recent literature, including Ross (1995). It will therefore also serve as a point of reference in this section. Compare the following tree diagram:

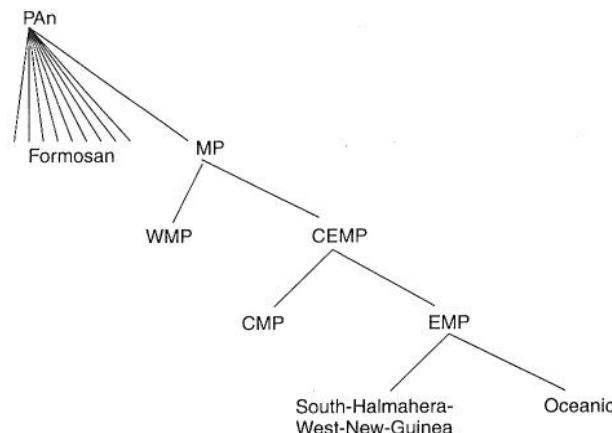


FIGURE 1.1 A TENTATIVE TREE DIAGRAM OF AUSTRONESIAN LANGUAGES (cf. BLUST 1980a AND 1999)

In this tree diagram, the Formosan languages (the Austronesian languages of Taiwan) belong to nine primary branches of the Austronesian family, whereas all non-Formosan languages belong to a tenth primary branch, Malayo-Polynesian (Blust 1999).

Furthermore, according to Blust (1980a), Malayo-Polynesian split into WMP and Central-East-Malayo-Polynesian (CEMP), whereas the latter in turn split into Central-Malayo-Polynesian (CMP) and East-Malayo-Polynesian (EMP).

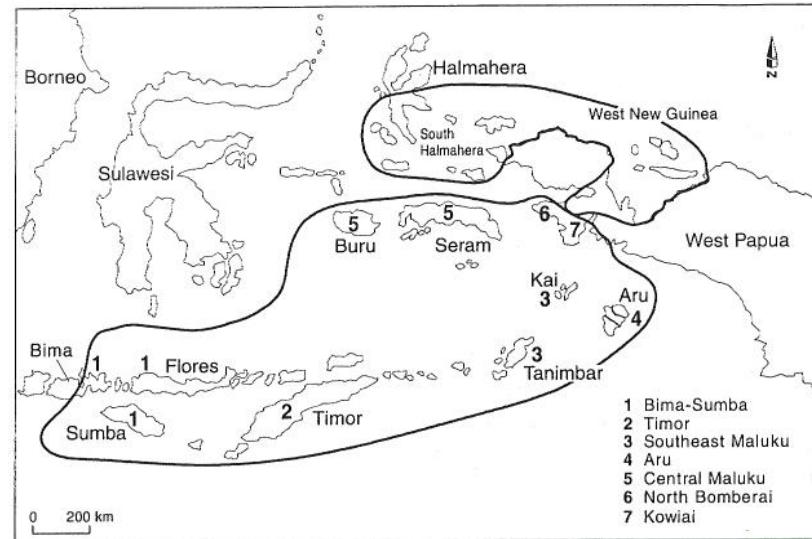
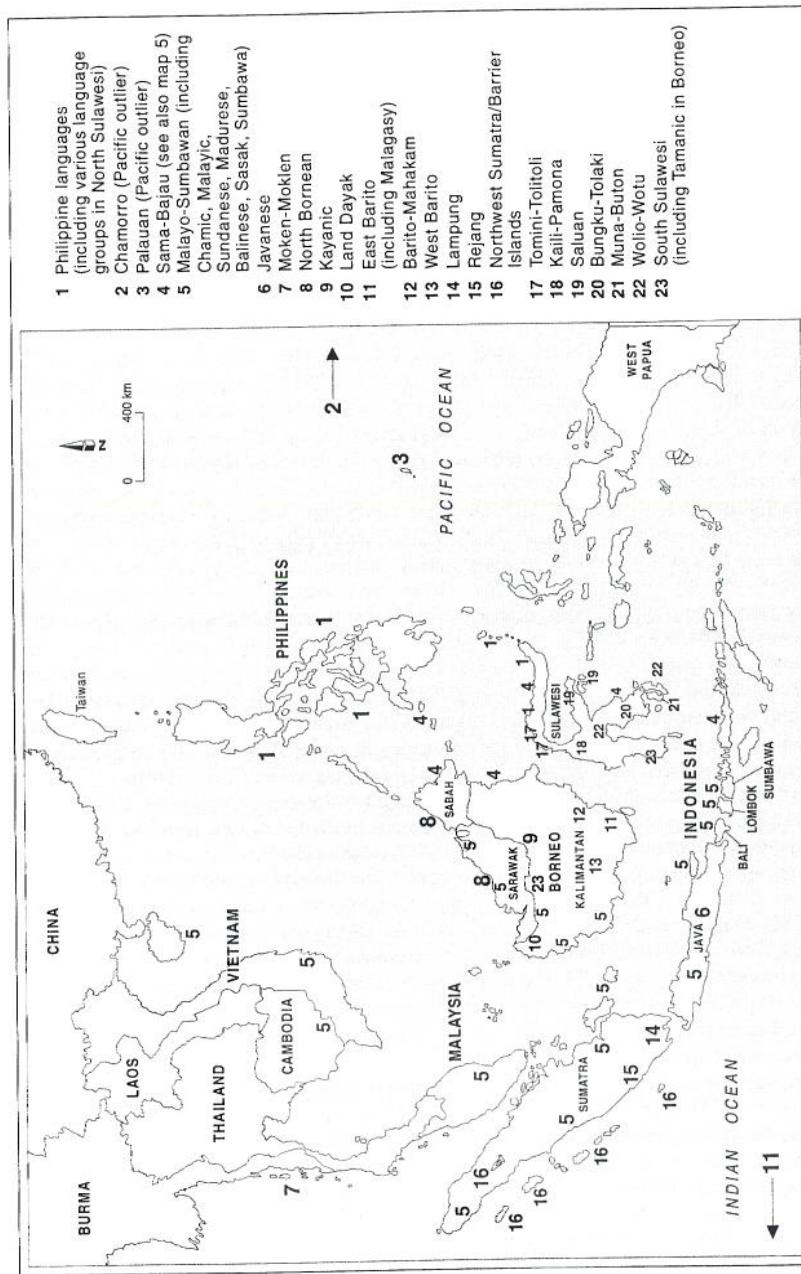
Languages belonging to WMP are languages of the Philippines and West Indonesia as well as Chamorro, Palauan, Malagasy and Chamic; West Indonesia includes Bali, Lombok, West Sumbawa and Sulawesi, with the Banggai islands to the east and the Tukang Besi – and Muna-Buton islands to the southeast (see Map 1.1).

Languages belonging to CMP are languages of the Lesser Sunda Islands from East Sumbawa (with Bimanese) onwards to the east, and languages of the central and southern Moluccas (including the Aru Islands and the Sula archipelago, but not Obi, Misool or parts north; see Map 1.2).

The South-Halmahera-West-New-Guinea group consists of languages of Halmahera, Cenderawasih Bay as far as the Mamberamo River, and of the Raja Ampat Islands (Waigeo, Salawati, Batanta, Misool) together with their satellites (see Map 1.2).

Finally, Oceanic consists of Micronesian languages, Polynesian languages and the various language subgroups located in Melanesia (Blust 1980a:11–12).

The remainder of this section discusses the structure of some of the higher- and lower-order branches of the above family tree insofar as they have been the topic of recent investigation.



5.1 Formosan languages

There are about fifteen living Formosan languages. There are furthermore some extinct Formosan languages that are on record. As they represent several primary branches in the classification of Austronesian languages, they are of great importance for a comparative-historical study of the language family.

Most Austronesian linguists and archaeologists would agree that the Formosan languages take a special classificatory position within the Austronesian language family and that the original ‘homeland’ of Austronesian languages must be sought in Taiwan and, prior to Taiwan, in coastal South China (see section 6).

However, there is much more controversy about the internal classification of Formosan languages and about their exact position vis-à-vis the Malayo-Polynesian languages. The nineties have seen some important debates evolving around the following issues:

- (1) How many Formosan subgroups are there?
- (2) Do the Formosan subgroups together form one single primary branch of Austronesian, or does each of them constitute a primary branch directly continuing Proto-Austronesian?
- (3) Do the Formosan languages form a primary branch (or several primary branches) with the Philippine languages, or are they a category on their own within the Austronesian language family?
- (4) Are the Malayo-Polynesian languages a totally distinct primary branch within the Austronesian language family, or are they a subgroup within one of the Formosan subgroups?

(1) As far as the number of Formosan subgroups is concerned, until recently the Formosan languages were often classified into Atayalic (Atayal and closely related languages), Tsouic (Tsou and closely related languages) and Paiwanic (Paiwan and closely related languages), following Ferrell 1969. However, this classification has met increasing criticism. More particularly, of the three subgroups mentioned above, Paiwanic is ill-defined and far too heterogeneous. There is a case for the inclusion of Paiwan, Puyuma, Bunun and Amis, but the inclusion of other alleged members is much less evident (Ross 1995:69).

In an attempt to avoid some of the randomness that is often inherent in lexical and typological subgrouping evidence, Blust (1999) makes a classification of Formosan languages based on phonological evidence only. He reaches a division into the following nine primary branches: 1. Atayalic; 2. East Formosan, with a Northern branch (Basai-Trobiawan and Kavalan), a Central branch (Amis), and a Southwest branch (Siraya); 3. Puyuma; 4. Paiwan; 5. Rukai; 6. Tsouic; 7. Bunun; 8. Western Plains, consisting on the one hand of Central Western Plains with Taokas-Babuza and Papora-Hoanya, and on the other hand of Thao; and 9. Northwest Formosan, with Saisiyat and Kulon-Pazeh. Whether each of these subgroups can be maintained requires further investigation on the basis of more than phonological evidence alone. However, the obvious merit of Blust's classification is that it is more systematic and provides a better working hypothesis for further research than Ferrell (1969) and other classifications involving Formosan languages (see below).

(2) Dyen (1963) and Tsuchida (1976) suggested that the Formosan languages together form a primary branch of Austronesian. Blust's classification of Formosan languages into nine primary branches clearly precludes the existence of such a unitary branch. The idea of one primary branch reappears in the nineties in an expanded form (including Philippine languages, see below).

(3) There exist various scenarios involving the subgrouping of Formosan and Philippine languages. Already in 1982 Reid defined Malayo-Polynesian languages on the basis of whether or not they exhibit nasal infixation. This led to a tree diagram which is very different from Blust's, with only the Central Philippine languages belonging to Malayo-Polynesian, and Northern and Southern Philippine languages being placed at various levels higher up in the Austronesian tree. Dyen and Tsuchida (1991) conclude that Formosan and Philippine languages do form a single primary branch on account of the large amount of vocabulary that these languages share. Wolff (1995) reaches the same conclusion on the basis of shared vocabulary as well as of some striking typological similarities in verbal morphosyntax.

Others (Blust 1995, 1999, Ross 1995) reject a Philippine-Formosan branch. Blust is able to show that part of the lexical evidence amassed by Dyen, Tsuchida and Wolff is not exclusively shared between Formosan and Philippine languages and is therefore not critical for a subgrouping (Blust 1995:606–616), nor is there lexical evidence for a Formosan primary branch (Blust 1995:594–606). Another, more general, issue raised by Blust is the lack of directionality inherent in lexical evidence: if Formosan and Philippine languages share a word X and the remaining Malayo-Polynesian languages share a word Y, it cannot always be demonstrated that X is an innovation and Y is a retention, rather than the other way around. Phonological evidence, on the other hand, often provides that directionality: if one language shows *s* and a related language shows *h* in cognate vocabulary, it is phonologically highly probable that *h* is an innovation and not a retention. As to Wolff's argument of shared morphosyntactic typology, Blust points to the fact that other Malayo-Polynesian languages also share this typology (e.g. Malagasy) or at least

show traces of it (e.g. Old Javanese, Chamorro or Toba Batak). It must therefore be a retention from Proto-Austronesian.

Blust's arguments are convincing. Nevertheless, the bulk of lexical and morphosyntactic similarities between Formosan and Philippine languages remains impressive and cannot be dismissed altogether as retentions. Some similarities may be due to contact. Contact did take place, not only between Formosan languages and Philippine languages (cf. Wolff 1995:573–574) but also between Formosan languages and other Austronesian languages (cf. §7), and among Formosan languages themselves, which remains an under-studied phenomenon. The likelihood of contact cautions against the use of evidence from only a few Formosan sources in conjunction with evidence from (especially northern Philippine) Malayo-Polynesian languages in cases where the possibility of borrowing cannot be ruled out.

(4) It is possible that certain Formosan languages subgroup with PMP. Amis had already been singled out by Reid (1982) as being closer to Malayo-Polynesian than are other Formosan languages. He did so on account of two shared phonological mergers: PAn *C and *t > t, and PAn *N and *n > n.

As a general theoretical principle, Ross makes a distinction between a subgroup, which refers to a group of languages that have diverged because of a separation, and a linkage, which refers to a group of languages that have arisen by dialect differentiation (Ross 1995:45–47). 'A linkage is formed when a chain of diverse dialects persists for long enough for innovations to diffuse across parts of the chain, in overlapping or linking patterns, without spreading across the entire dialect chain' (Pawley 1999:130). Ross believes 'that Proto-Austronesian had already diversified into a linkage of dialects before the ancestors of Malayo-Polynesian speakers left Taiwan'. It is therefore important to ask from which part of the linkage this pre-PMP must have broken off. Ross thinks that the merger of PAn *C and *t shared by Amis, Bunun and PMP may be a lead: if Amis and/or Bunun also share other important innovations with PMP, they are probably the source of pre-PMP (Ross 1995:69).

Blust (1999:54), however, opposes attempts to subgroup Amis with PMP because phonological evidence clearly demonstrates that Amis belongs to the East Formosan branch (together with Siraya, Basai-Trobiawan and Kavalan). This branch is not only defined by the merger *C, *t > t, but also by the more significant change from *j to n, which never occurred in Bunun, nor anywhere else in the Austronesian language family.

This does not entirely exclude the possibility of a close kin relationship between Amis and PMP, but if there is such a relationship, it would involve the entire East Formosan branch, not just Amis. PMP and East Formosan would remain separate subgroups within a primary Austronesian branch, as they are both defined by a number of mutually exclusive phonological changes. The only evidence for such a branch would be the merger of PAn *C and *t to *t. On the other hand, if the phonological basis for an East Formosan–PMP branch is not immediately evident, there is some lexical evidence to support it. Blust himself (Blust 1999:46–47) points out that 'members of [the East Formosan] group have been in continuous contact with the sea since the break-up of PAn'. He gives evidence for the existence of maritime terms that are shared by MP and East Formosan languages but are not found in other Formosan languages.

In view of a possible close link between Malayo-Polynesian and East Formosan languages it is enlightening to have a further look at Siraya, a language which has not yet been fully exploited for its subgrouping evidence. As a member of the East Formosan branch, Siraya shares with Amis the changes *C, *t > t and *j > n. The merger of PAn *N and *n > n is reflected in some Siraya dialects, but not in others, which have l,

e.g. *malituk* vs. *vanitok* (written ‘vannitock’) ‘silver, money’; *sulat* vs. *sunat* (‘sonnat’) ‘document’ (Tsuchida and Yamada 1991:5). The last instance incidentally shows the directionality and relatively recent occurrence of this innovation: *sunat* must be borrowed from Tagalog (cf. *sulat* ‘writing’ which in turn derives from Malay *surat*, cf. Adelaar 1995a:332). It shows that the change (from a lateral to a nasal) was still in process at the time of borrowing, which must have been after the Malays had begun to extend their trade network to the Philippines.

Siraya shares a number of lexical features with Malayo-Polynesian that are not found in Amis or other Formosan languages, such as (PMP) **Ratus* ‘hundred’ and (Siraya) *ka-xatux-ay* id.; **laba* ‘spider’ and *rawa* id.; **kuliC* ‘skin’ and *k-m-urit* ‘to peel’; **tuRut* ‘follow’ and *t-muxot* id.; **qaRus* ‘stream’ and *axu-aj* ‘river’. It furthermore shares some irregular phonological changes that are not attested in other Formosan languages, and that are considered diagnostic for PMP by Blust (1995). These changes are syncope of **u* in PAn **payudaN* ‘pandanus’ > *pandal*, and reduction of the medial cluster in **biRbiR* ‘lip’ > (gospel dialect) *bibix* (but cf. *vixbix* in the Utrecht Manuscript dialect, Van der Vlis 1842). It appears, then, that various members of the East Formosan branch provide evidence for a closer link with Malayo-Polynesian. However, they do so each in their own way, and their evidence does not always seem to corroborate.

5.2 Malayo-Polynesian

Ross (1995:69) states that ‘the unity of the Malayo-Polynesian languages is probably not open to serious question, and PMP is readily reconstructible’. However, if the existence of a Malayo-Polynesian subgroup has met relative acceptance among Austronesianists, its internal classification is much less established. Blust’s subdivision is weak in several branches, and he himself accedes to the fact that West-Malayo-Polynesian is not a genetically attested subgroup (see below). However, as Ross points out, CEMP is also problematic, and so are its sub-branches CMP and EMP (Ross 1995:81, 82, 85).

5.3 West Malayo-Polynesian

In contrast to Malayo-Polynesian, WMP does not have a clear linguistic foundation. In the literature this category is often referred to as if it were an established subgroup, but the linguistic criteria for it have not been formulated and tested properly, and the genetic affiliations of its putative members remain to be investigated. This is acknowledged once again by Blust (1999:68), who says that this category has been used as a convenient ‘catch-all’ category for all MP languages other than the CEMP ones. ‘No phonemic innovations characterize the WMP languages as a group, and a hypothetical PWMP is thus phonologically indistinguishable from PMP.’

Much more comparative research than has been done so far will be required to sort out the exact genetic relations between these languages. Many of them evidently do cluster in microgroups, but their exact genetic affiliations are very difficult to trace at the higher nodes, and scholars differ as to their classification. In some cases, existing lower-order subgroups and individual languages may be linked up together to form a major branch within WMP, but others, like Chamorro and Palauan (two WMP outliers in Micronesia), and also Rejang (Sumatra), are ‘nuclear’ members which do not seem to form a further subgroup at all with any particular language within WMP (see further below).

More or less in an attempt to create order in the chaos of subgrouping propositions regarding the languages that are allegedly included in the WMP ‘fake’ subgroup, Ross (1995:74ff.) distinguishes 24 groups that have been identified as WMP by previous scholars. His list does ‘not entail a commitment to their genetic unity’ but it is meant to highlight ‘current issues’ in the classification of languages labeled as WMP (Ross 1995: 74). Ross’ 24 subgroups are only a working hypothesis to be used as a base for future adjustment into a more genetic classification. They are:

- 1 Batanic (south of Taiwan);
- 2 Northern Philippines;
- 3 Meso-Philippines;
- 4 Southern Philippines;
- 5 South Mindanao;
- 6 Chamorro and Palauan (outliers in Micronesia);
- 7 Sangiric (northern Sulawesi);
- 8 Minahasan (northern Sulawesi);
- 9 Gorontalo-Mongondic (northern Sulawesi);
- 10 Sama-Bajau (‘Sea Gypsies’ living in Philippines, Sabah, Kalimantan, Sulawesi and East Indonesia);
- 11 Malayo-Chamic;
- 12 Moken and Moklen (Burma, Thailand);
- 13 Northwest Borneo;
- 14 Land Dayak (Borneo);
- 15 East Barito (including Malagasy; Borneo);
- 16 Barito-Mahakam (Borneo);
- 17 West Barito (Borneo);
- 18 Lampung (Sumatra);
- 19 Northwest Sumatra/Barrier Islands;
- 20 Java-Bali-Sasak (Lombok Island);
- 21 Central Sulawesi;
- 22 South Sulawesi;
- 23 Muna-Buton (Southeast Sulawesi); and
- 24 Tamanic (central Borneo).

Several publications have appeared since Ross (1995) was submitted to the publisher (which in fact happened several years before it was actually published). It is therefore necessary to adjust Ross’ inventory of subgroups and to reduce it to some extent, as I will do in the remainder of this section. I will discuss the various new publications and present evidence for a rearrangement of his inventory in the following way:

- 1 Philippine languages (including the Sangiric, Minahasan and Gorontalo-Mongondic languages of North Sulawesi)
- 2 Chamorro (Pacific outlier)
- 3 Palauan (Pacific outlier)
- 4 Sama-Bajau (Philippines, Sabah, eastern Indonesia)
- 5 Malayo-Sumbawan (Malayic, Acehnese and Chamic, Balinese, Sasak, Sumbawanese, Madurese, Sundanese)
- 6 Javanese
- 7 Moken-Moklen (Mergui archipelago, Myanmar, insular and coastal West Thailand)
- 8 North Bornean

- 9 Kayanic (central and eastern Borneo)
- 10 Land Dayak (western Borneo)
- 11 East Barito (southeastern Borneo, Madagascar)
- 12 Barito-Mahakam (East Borneo)
- 13 West Barito (central and southern Borneo)
- 14 Lampung (South Sumatra)
- 15 Rejang (Southwest Sumatra)
- 16 Northwest Sumatra/Barrier Islands
- 17 Tomini-Tolitoli (northwestern Sulawesi)
- 18 Kaili-Pamona (western Sulawesi)
- 19 Saluan (eastern Sulawesi)
- 20 Bungku-Tolaki (southeastern Sulawesi)
- 21 Muna-Buton (southeastern Sulawesi)
- 22 Wolio-Wotu (southeastern Sulawesi)
- 23 South Sulawesi.

5.4 Philippines

Some caution is in order regarding the term ‘Philippine languages’. Although Philippine languages seem to be closely related to each other (see below), in the literature the term often refers to a typological category, usually including the languages of the Batan Islands (South Taiwan) and of North Sulawesi (Sangiric, Minahasan and Gorontalo-Mongondic) but excluding the Sama-Bajau languages (spoken in several spots along the coasts of the southern Philippines, Sabah, eastern Borneo and eastern Indonesia, see Akamine, this volume). This configuration of languages also underlies Zorc’s Philippine subgroup (Zorc 1986), although his subgrouping claim is based on lexical evidence. Some authors also include the languages of North Borneo in this category (Ross 1995:73).

Blust (1991) agrees with Zorc’s Proto-Philippine subgroup but he makes a different internal classification. He identifies South Luzon and the Visayan islands as areas with a low linguistic diversity, which, he asserts, is due to language leveling. He distinguishes a ‘Greater Central Philippine subgroup’, the members of which all have merged PAn *R and *g to g and furthermore share some lexical and semantic innovations. These members are South Mangyan, Tagalog, Bikol, the Visayan languages, Palawanic (but not Kalamian), the Mindanao languages (but not South Mindanao) and the Gorontalo-Mongondic languages. The Greater Central Philippine subgroup in turn belongs to the Philippine subgroup together with all other Philippine languages and with the Sangiric and Minahasan languages (but excluding the Sama-Bajau languages). Around 500 BC speakers of Greater Central Philippine began to expand outward from a center somewhere in northern Mindanao or the southern Visayas. This expansion affected many other Philippine languages as well as the languages of Sabah but it did not cause their disappearance. Nevertheless, the influx of lexical borrowings caused by the Greater Central Philippine expansion makes it difficult to assess the genetic affiliations of these other Philippine and Sabahan languages (Blust 1991, 1998b:42).

In accordance with Blust’s integration of all Philippine languages and various northern Sulawesi groups into an encompassing Philippine subgroup, Ross’ first five subgroups together with his seventh, eighth and ninth ones can be reduced to one, the (greater) Philippine subgroup (1).

5.5 Chamorro, Palauan, Sama Bajau

Various speculations exist as to the classification of Chamorro and Palauan. On the basis of their status as ‘WMP outliers’, or rather as non-Oceanic languages within Micronesia, they are usually mentioned together, which may unintentionally create the impression that they are closely related. The fact that Chamorro retains the PAn voice markers has sometimes created the impression that it is a Philippine language, or even a Formosan one. Using phonological and morphosyntactic arguments, Zobel shows that both languages are nuclear members within (West) Malayo-Polynesian. Their phonological histories do not link them up with any other Austronesian language or subgroup in particular, and they are both somewhat idiosyncratic representations of the Indonesian-type morphosyntactic model (Zobel 2002). However, see also Reid (2002) who argues that Chamorro is a Philippine language.

Chamorro (2) and Palauan (3) are classified as isolates. The position of Sama-Bajau as a nuclear group remains unchanged (4).

5.6 Malayic languages, Chamic languages and Bali-Sasak-Sumbawa

5.6.1 *The Chamic subgroup and its affinity with Malayic*

Thurgood (1999) studies the internal relationships and linguistic developments of Chamic languages. In the sixth century AD Cham became the lingua franca of the Champa kingdom in South Vietnam. It was spoken in a largely undifferentiated dialect chain extending along the coast of Vietnam and possibly even connecting with the east coast of Malaysia (see below). Vietnamese territorial expansion towards the South forced Cham speakers to spread over a wider area and the dialect chain developed into distinct Chamic languages including Tsat (Hainan), Rade, Jarai, Haroi, Chru, Roglai, Phan Rang Cham (or eastern Cham) (highlands of southern Vietnam) and Western Cham (Cambodia).

One form of Chamic, later to become Acehnese on Sumatra’s north coast, must have split off at an earlier date but already after a dialect chain had begun to form. Remarkably, Thurgood considers Acehnese as a Chamic language directly subgrouping with Roglai. Earlier scholars tended to define Acehnese as an early branch-off within an Aceh-Chamic subgroup, giving it a status co-ordinate with the Chamic subgroup and assuming a break-away from Acehnese before the Chamic languages had begun to diverge (Thurgood 1999).

Scholars had been speculating on a close relationship between Cham, Acehnese and Malay since Morrison (1975). Blust (1994b) argues for it on the basis of lexical evidence, also speculating that at one point in history (before Thai invasions) there might have been a more or less continuous dialect chain of Chamic and Malayic (see below) along the east coast of Southeast Asia. Although Thurgood (1999) does not directly aim at demonstrating a direct link between Chamic and Malayic, his reconstruction of Proto-Chamic confirms it beyond reasonable doubt.

5.6.2 *Subgrouping attempts involving Malayic languages*

Malayic is a genetic subgroup which includes Malay and all dialects and languages that are directly related to it, such as Kelantan-, Jakarta-, Kerinci-, Banjar- and Brunei-Malay, and furthermore Iban, Minangkabau and Kendayan. The homeland of Malayic is

believed to be in West Borneo, although Malay (at least in its literary and standard forms) most likely originated in South Sumatra (Blust 1984–85, Adelaar 1985, 1992 and to appear). The internal classification of Malayic remains a much disputed topic. Adelaar (1985) divides the Malayic languages into one primary branch including Iban (West Borneo), and another including all other Malayic varieties. This is contested by Nothofer (1988), who proposes a wider Malayic subgroup also including Salako, Embaloh and Rejang. I present evidence against their inclusion (Adelaar 1993, 1994a and below), and I also refrain from an internal classification of the Malayic varieties on the basis of our present – insufficient – knowledge of them (Adelaar 1992).

In recent years, several other internal Malayic classifications have been made by Nothofer (1995, 1997a, 1997b), Collins (1994, 1996a) and Ross (in press).

Collins (1994) believes that the Malay varieties of Borneo form a single subgroup. His ‘Proto-Malay’ (the exact scope of which remains somewhat unclear, Adelaar 2004:14) existed as one or several closely related Austronesian languages some 2000 years ago. It split into a Sumatra/Malay peninsula branch and a Borneo branch; the Borneo branch developed into a northern branch (including Sambas and Brunei Malay) and a southeastern branch (with Banjar-, Kutai- and Berau Malay). A close historical link between Brunei Malay and Banjar Malay among others appears from their verb morphology, which includes a combination of the locative *-i* suffix with a following beneficiary-oriented suffix *-akan* (Collins 1996a:83–84). Bacan Malay (spoken on Bacan Island, North Moluccas) is an offshoot from Brunei Malay and also allows a concatenation of these suffixes (Collins 1996a).

Nothofer (1996, 1997a) proposes several branches which each correspond to a migration route from a Malayic homeland in Northwest Borneo. One route was in a northeastern direction and then further southwards and westwards along the Bornean coast, generating, among others, Brunei Malay and Banjar Malay. Another route was to the Tambelan and Riau archipelagos and the Malay peninsula. A third route was to the Bangka and Belitung Islands, South Sumatra and Java’s north coast. Among the dialects established along the last route, Nothofer (1995) distinguishes some closely related varieties in the ‘southwest corner’ of the South China Sea, including Jakarta-, Seraway- and Bangka Malay. Nothofer (1997b) argues for a close relationship between these ‘southwest corner’ dialects and languages of Northwest Borneo (including Iban, Sarawak Malay and Kendayan), Tioman Malay (off the Malay peninsula’s east coast) and Palembang Malay. He now classifies the Malayic varieties in West Borneo into West Borneo Malayic and Southwest Borneo Malayic.

As it stands, the data suggest a special genetic position for some of the Northwest Bornean varieties (primarily Kendayan and possibly Iban). Nevertheless, this remains to be demonstrated on the basis of linguistic data of a more critical nature than adduced so far. The evidence for some of the above classifications is too limited and in some cases even contradictory. In Adelaar (2004) I therefore maintain my earlier position that the Malayic subgroup is not capable of internal classification unless more linguistic data become available.

Ross’ subgrouping argument is primarily based on morphosyntax (Ross in press). He divides the Malayic subgroup into Western Malayic Dayak (Kendayan) and ‘nuclear Malayic’ (all other Malayic varieties). Critical for this division is the application of the prefix *di-* as well as some other innovations. In nuclear Malayic varieties, *di-* became a verbal prefix occurring in passive constructions with a third person agent. In Kendayan, however, *di-* has not fully developed into a verbal prefix: the actor noun phrase is allowed to occur between it and the following verb, and it is prefixed directly to the latter only if

the actor is not expressed. Compare *di-* in the following sample sentence (from the Salako subdialect): the first *di-* is directly followed by the actor and functions as an actor preposition, whereas the second *di-* functions as an undergoer voice marker which is prefixed to the verb in a clause without explicit actor:

- | | | | | | | |
|-----|------------------|------------------|-------------|----------------|--------------|-------------|
| (3) | <i>Taykitn-e</i> | <i>dah</i> | <i>tabà</i> | <i>di</i> | <i>darah</i> | <i>kayo</i> |
| | sword-3poss | already | thick | ACT | blood | enemy |
| | <i>mujkus</i> | /N-bujkus/, and? | <i>bisà</i> | <i>di-buaj</i> | | |
| | CPL-wrap | NEG | can | uv-throw.away | | |

‘His sword had become thick, covered with the enemy’s blood which could not be wiped off.’

Furthermore, nuclear Malay lost the PAn projective suffix *-a* (retained in Kendayan as *-a?*). Finally, according to Ross, nuclear Malayic uses the inflected verb form for atemporal aspect whereas Kendayan retained the original PMP atemporal verb marking, which is expressed by a bare stem (Ross in press).

Ross does not consider Old Malay (cf. Mahdi, this volume) to be a member of the Malayic subgroup, excluding it on the ground that it has not yet acquired the passive marker *di-* and that it still reflects the PMP transitive verbal prefix **maR-* as *mar-*. Contemporaneous Malayic varieties have *di-*, and they reflect **maR-* as *bər-* (or a related form with initial *b*).

As already mentioned, Kendayan may ultimately turn out to represent a separate primary branch. It is very different from other forms of Malayic and its aspectual distinction represents a substantial deviation from the consistent lack of aspect marking on verbs in other Malayic varieties. However, although Ross considers this distinction a PMP retention, it is far from identical with the way temporality and neutral aspect are marked in PMP (including most Malayic varieties), and it might just as well be an innovation. Furthermore, the other arguments adduced by Ross hinge on the assumption that Old Malay is not Malayic, but there are strong arguments favoring its inclusion in the Malayic subgroup. On the one hand, Old Malay shares its phonological history with other Malayic varieties, and on the other hand, both the absence of *di-* and the presence of a reflex of the projective marker **-a* are also observed in (nuclear) Malayic varieties. The prefix *di-* does occur in inscriptions of a somewhat later date, and it is also the only prefix not to have undergone the all-pervasive phenomenon of antepenultimate neutralization of vowels to schwa in most forms of Malayic, which shows that it is a late development. There is therefore little doubt that Old Malay and modern Malay are forms of the same language, in spite of some considerable differences between them. Once this is established, it follows that some critical features used to set off Kendayan against nuclear Malayic are not exclusive to it (Adelaar in press a).

5.6.3 A Malayo-Sumbawan subgroup

Mbete (1990) made a reconstruction of Proto-Bali-Sasak-Sumbawa, which gives ample evidence of the common phonological history of these languages and the lexical innovations they share. Ross (1995) refers to a Java-Bali-Sasak[-Sumbawa] subgroup, while admitting that it is a weak hypothesis.

Apparently Esser (unseen 1930 publication) included Bali-Sasak-Sumbawa as a separate branch in a Javanese-Madurese subgroup (cf. Mbete 1990:10), but he does not do so on his 1938 map which shows the Bali-Sasak-Sumbawa without affiliations to other

Austronesian languages in particular. Blust (1984–85) speculates on a subgroup including Javanese, Balinese, Sasak and Sumbawa, which together with Malayo-Chamic and the various Barito groups would have a common homeland in Southwest Borneo. The basis for this subgroup remains unclear. Its inclusion of Javanese is moreover at odds with an (equally problematic) Malayo-Javanic hypothesis, which in Nothofer's 1985 amendment posits Javanese as a primary branch coordinate with another branch including all other Malayo-Javanic members, including Malay, Madurese and Sundanese.

Similarities between Javanese and the Bali-Sasak-Sumbawa languages appear most clearly between Javanese and high register Balinese and Sasak. Sociolinguistically, however, there is no ground to use the high register vocabularies of these languages for the purpose of investigating their genetic affiliations because high register vocabularies are much more susceptible to borrowing than low register ones (cf. Clynes 1989, Nothofer 1975).

Adelaar (in press b) proposes a 'Malayo-Sumbawan' subgroup, which excludes Javanese and has a Madurese branch, a Sundanese branch, and one branch comprising the Malayic-, Chamic- and Bali-Sasak-Sumbawa languages. By using low Balinese lexicon only, he finds that Balinese, Sasak and Sumbawa share a significant part of their phonological history with Malayic and Chamic. Although some of the phonological developments are not forceful in themselves, or even unique to Balinese-Sasak-Sumbawa and Malayic, their configuration is striking. It includes PMP *w- > Ø; PMP *q > *h; PMP *R, *r > *r; PMP *z > *j; PMP *j, *d > *d. In contrast, Madurese, Sundanese and particularly Javanese are phonologically more divergent from Malayic as well as from Chamic or Bali-Sasak-Sumbawa. Furthermore, Balinese, Sasak and Sumbawa have some basic vocabulary in common with Malayic as well as formal similarities in vocabulary such as the metathesis in *qudip 'to live' which is reflected in Malay *hidup* and in Balinese and Sasak *idup* (but not in Chamic *hudip). These common features should be seen against the histories of Lombok and especially Bali, where linguistic and cultural influence from Java has been overwhelming since the twelfth century, whereas Malay influence has been relatively weak.

This new subgrouping hypothesis requires a replacement of Ross' Malayo-Chamic and Java-Bali-Sasak groups with (5) Malayo-Sumbawan (including Chamic etc., as defined above) and (6) Javanese.

5.7 Moken and Moklen

Larish classifies the Moken-Moklen languages as a primary ('Moklenic') branch in a macrogroup together with Malay, Acehnese and Chamic (Larish 1999 and this volume).

However, Thurgood emphasizes that the phonological history of these languages sets them off rather sharply from Malayic and Chamic (including Acehnese). He points out that they do not share the change from PAn *q to h, but reflect k instead. Furthermore, both Chamic and Moken-Moklen have diphthongized PAn word-final high vowels, but they have done so each in very different ways (Thurgood 1999:58–59). One could add some other obvious differences in phonological history, such as PAn *w-, which became Moken and Moklen b- whereas *w- was lost in Malayic and, after a following *a became colored to u, in Chamic; PAn *R, which became Moken and Moklen l but in Malayic and Chamic became r.

Moken-Moklen languages here will keep their status as a separate subgroup (7); more intensive research is required in the future to determine their position vis-à-vis Malayic and Chamic.

5.8 North Bornean

The Sabahan languages (Banggi, Dusunic, Murutic, Paitanic, Ida'an, Tidong) show various striking similarities with Philippine languages. They have a Philippine-type morphosyntax. They furthermore share some vocabulary, such as a reflex of *siam for 'nine', and some phonological developments exhibiting, for instance, g for PAn *R. Blust (1998b) points out that the morphosyntactic patterns of Sabahan are also shared by non-Philippine languages such as the languages of Formosa, Malagasy, Chamorro and Toba Batak. They are very likely a retention from Proto-Austronesian and therefore not diagnostic for subgrouping. The g reflex for *R, although widely attested among Sabahan languages, is not attested in all of them and is not criterial for the Sabahan subgroup as a whole. Blust makes further investigations into the genetic affiliations of Sabah languages and finds that they have most of their critical phonological developments in common with the languages of North Sarawak (Kelabitic, Kenyah, Berawan-Lower Baram and Bintulu). These developments are: 1. Loss of *h; 2. Schwa epenthesis; 3. Consonant fortition; 4. Loss of initial *q; 5. reduction of consonant clusters; and 6. antepenultimate neutralization. Especially significant for a classification are schwa epenthesis and consonant fortition as well as the conditions under which these changes took place. Sabahan languages seem to share with Philippine languages the merger of an original palatal and dental series. However, on closer inspection it appears that the parts of this merger are either widely attested among Austronesian languages in general (i.e. the merger of *c and *s), or they are not clearly attested in Philippine languages (i.e. the merger of *z and *d and of *ñ and *n). Sabahan languages do share lexical innovations with both North Sarawakan languages and Philippine languages, but here again it appears that the number of shared innovations is higher with North Sarawak languages than with Philippine ones. Sabahan languages must have been affected by lexical diffusion coming from both other groups, as Blust (1998b) argues. All these facts taken together clearly show that Sabahan languages have more linguistic history in common with the languages of North Sarawak than with those of the Philippines, and Blust proposes a North Borneo subgroup. He points out that if Sabahan languages share lexicon with Philippine languages this is almost always with members of the Greater Central Philippines group. The reflex g for *R shown in many Sabahan languages is also a typical feature of Greater Central Philippines languages. Blust attributes these factors to contact-induced change as a result of the Greater Central Philippines expansion (see above). This expansion also had far-reaching contact effects in the peripheral parts of the Philippines and in North Sulawesi.

It is noteworthy that Blust's North Sarawak branch of North Borneo languages does not include the Kayanic languages (Kayan, Modang, Murik). In Hudson's (1978) classification of Bornean languages, Kayanic and Kenyahic languages belong together in the same lower-order subgroup (compare also Rousseau 1988:5), but this is rejected by Blust (Blust 2002:29–30). One of the developments that clearly set off Kayanic languages from Kenyahic ones is the fact that they often exhibit a final glottal stop where the latter have Ø, and vice versa. Blust explains this as the result of an ordered set of phonological changes in Kayanic: 1. vowels were lengthened before a final glottal stop (< PAn glottal stop); 2. all words acquired a final glottal stop; 3. short vowels before a final glottal stop were lowered; and 4. final glottal stops were lost after long vowels. The genetic affiliations of Kayanic obviously need further study but for the time being it seems safe to give it a separate subgroup status. Ross' (1995) Northwest Borneo group is replaced by North Borneo (8) and Kayanic (9).

5.9 Other Bornean subgroups

Ross (1995) lists four other Bornean subgroups which are maintained here, to wit Land Dayak in western Borneo (10), East Barito in Southeast Borneo (including Malagasy in Madagascar) (11), Barito-Mahakam in eastern Borneo (12) and West Barito in central and southern Borneo (13).

With regard to East Barito, Dahl (1991) claimed that the Lom language on Bangka Island belongs to this group and is originally a form of Maanyan. However, it is clearly Malayic according to Nothofer (1994) and Adelaar (1995a).

5.10 Lampung, Northwest Sumatra/Barrier Islands, Rejang

Ross (1995) lists Nothofer's (1985) Northwest Sumatra/Barrier Islands group which includes the Batak languages and Gayo in Northwest Sumatra, and Nias, Mentawai, Sichule, Simalur and Enggano on the islands of the same name off Sumatra's west coast (14). He lists Lampung in South Sumatra as an isolate (15).

Rejang, another South Sumatran language (spoken in Bengkulu Province in southwest Sumatra) has a rather unique phonological history and should also be treated as an isolate (16) in the list of subgroups (cf. Blust 1984).

5.11 Sulawesi subgroups

Ross (1995) lists six independent Sulawesi subgroups: Sangiric (northern Sulawesi); Minahasan (northern Sulawesi); Gorontalo-Mongondic (northern Sulawesi); Central Sulawesi; South Sulawesi; Muna-Buton (Southeast Sulawesi). But other authors usually distinguish around ten subgroups. For instance, Sneddon (1993) lists the following ones: 1. Sangiric; 2. Minahasan; 3. Gorontalo-Mongondic; 4. Tomini (northern part of Central Sulawesi); 5. Bungku-Tolaki (eastern part of Central Sulawesi); 6. Kaili-Pamona (western part of Central Sulawesi); 7. Saluan (eastern tip of Central Sulawesi); 8. Banggai (id.); 9. Muna-Buton (Southeast Sulawesi); and 10. South Sulawesi. A slightly different constellation of nine subgroups is listed in Noorduyn's (1991) bibliography.

As mentioned above, Blust (1991) considers the Sangiric, Minahasan and Gorontalo-Mongondic languages as belonging to the Philippine group. The South Sulawesi languages will be discussed in the next subsection.

Van den Berg (1996) tries to combine the Kaili-Pamona, Bungku-Tolaki and Muna-Buton languages in a 'Celebic' group on the basis of various phonological changes: the loss of final consonants and of consonant clusters, the monophthongization of PMP final diphthongs and the shift of PMP to a back rounded vowel. However, the loss of final consonants appears to be due to an areal feature in most of the Sulawesi languages, and it postdates the emergence of the various Sulawesi subgroups (cf. Sneddon 1993). Sneddon concedes that this process may have begun in a stage ancestral to Kaili-Pamona, Bungku-Tolaki and Muna-Buton, but Mead (1996) demonstrates that this is not the case, which weakens Van den Berg's subgrouping claim.

On the basis of some sound developments in the Wolio-Wotu languages (consisting of Wolio, Wotu, Kamaru, Laiyolo and Kalao), Donohue (in press) gives them a separate subgrouping status from the Muna-Buton group. In Donohue's view, Wolio-Wotu may be closer to the Kaili-Pamona group, while Mead adduces evidence for macrogrouping Muna-Buton with Bungku-Tolaki languages (Mead 1998 and 2002).

In a recent paper presented at the 9th International Conference of Austronesian Linguistics in Canberra (2002), Van den Berg (in press) accepts the exclusion of Wolio-Wotu from the Muna-Buton group. He gives some phonological and grammatical arguments in support of the reduced Muna-Buton group, which in his analysis includes the Tukang Besi languages, albeit in a separate sub-branch.

In another paper presented at the same conference, Mead (in press a) has a closer look at the Saluan group, which is defined by eleven sound changes that its members have undergone since PMP. He includes Banggai in the group and argues that together with Balantak it forms a separate branch vis-à-vis the other Saluan languages, which together form another (western) branch. Mead (in press b) furthermore combines Bungku-Tolaki, Muna-Buton and Tukang Besi in a Southeastern Celebic group. This group shares the first eight sound changes that are diagnostic for Saluan-Banggai, and, according to Mead, forms Eastern Celebic with the latter. He speculates that the Tomini-Tolitoli, Kaili-Pamona and Wolio-Wotu groups may belong together in the western branch of an encompassing Celebic macrogroup. In so doing he in some way takes up Van den Berg's idea of a Celebic group (Van den Berg 1996), although he arrives at a different configuration of Sulawesi micro-groups.

Mead's inclusion of Banggai in a Saluan-Banggai group is convincing. Some of his eleven sound changes are of a general nature, but others are more diagnostic, such as (PMP) **awa* > Saluan **oa*. On the other hand, his evidence for a direct link between Banggai and Balantak is weak because it essentially depends on one change, the lowering of final **u* to /*o/* (the merger of **r* and **R* being of rather low diagnostic value). Except for the South Sulawesi, Gorontalo-Mongondic, Sangiric and Minahasan languages, languages of Sulawesi reflect PMP **j* as either /*y/* or *o*. Mead reasonably assumes that in the case where **j* became *ø*, it must first have changed to **y* before it finally disappeared. Such a course of change strengthens the case for his Celebic macrogroup considerably. Other critical sound changes for it are PMP **ə* > **o*, PMP **d* > **r*, PMP *-aw, *-aw > *-o and PMP *-ay, *-ay > *-e (note that *-aw and *-ay are not distinguished from *-aw and *-ay respectively in Blust (1999)).

The various subgrouping hypotheses above are difficult to assess. There seems to be agreement on Donohue's claim that the Wolio-Wotu languages should be separated from the original Muna-Buton group. Progress has been made on the establishment of a possible Celebic subgroup, but it may still be too early to accept Mead's definition of it without additional evidence from various linguistic levels, especially regarding the inclusion of Tomini-Tolitoli, Kaili-Pamona and Wolio-Wotu. In the mean time, Ross' (1995) Central Sulawesi and Muna-Buton groups are too general. To remain on the cautious side, they are replaced by (17) Tomini-Tolitoli, (18) Kaili-Pamona, (19) Saluan, (20) Bungku-Tolaki, (21) Muna-Buton, and (22) Wolio-Wotu. While most of these are reasonably well established subgroups, the genetic unity of Tomini-Tolitoli has in fact never been proven, and it is not obvious that Tolitoli forms a low-level subgroup with the Tomini languages (see Himmelmann 2001:20).

5.12 A Tamanic-South Sulawesi subgroup

The Tamanic languages (Embaloh, Taman, Kalis) are spoken at the upper reaches of the Kapuas River in the northeastern part of West Kalimantan, Borneo. Ross (1995) lists them as an independent subgroup within WMP. However, there is substantial phonological, lexical, semantic and morphosyntactic evidence linking these languages directly to

the South Sulawesi subgroup, and more particularly to Buginese (Adelaar 1994a). South Sulawesi languages are spoken in the southwestern peninsula of Sulawesi and comprise Makassar, Buginese, Seko and the northern subgroup (Sa'dan Toraja, Mandar, Massenrempulu, Mamuju, and Pitu Uluunna Salu) (Sneddon 1993:21).

Typical phonological innovations shared by Tamanic and Buginese are $*j > s$, loss of $*p$ in various cognate forms, lenition or loss of $*b$, merger of $*z$ and $*d$ into d - and $-r$ -, and monophthongization and heightening of final diphthongs. Exclusively shared lexical innovations include (Embaloh and Buginese) *sao* 'house', *si-ala* 'get married', *lino* '(the physical) world', Embaloh *lilu?* and Buginese *lilu* 'forget', Embaloh *iyar* 'nose' and Buginese *iya?*, Embaloh *sera* 'one' and Buginese *a-sera* [(ten) minus one=] 'nine' etc. An important grammatical innovation shared with South Sulawesi languages is the presence of four series of pronouns (free, absolute, ergative and possessive). These series by and large reflect the Proto-South-Sulawesi pronominal system as reconstructed by Mills (1975). Similar pronoun systems are also found elsewhere in Sulawesi and the Philippines but in central Borneo there seems to be no system with the same formal and structural similarities.

Many of these innovations are idiosyncratic and regular enough to rule out chance resemblance as an explanation. Language contact is also an unlikely explanation considering that the Tamanic communities live deep in the almost impenetrable heart of Borneo. Furthermore, Tamanic languages as a rule maintained PAn root-final consonants whereas South Sulawesi languages (especially Buginese) have often lost them. Although some Buginese did migrate to Borneo, they have done so in fairly recent times (seventeenth century AD and later) and they are not to be confused with the Tamanic communities under discussion. Adelaar (1994a) concludes that Tamanic languages belong to the South Sulawesi language group and that they subgroup more directly with Buginese. Speakers of Proto-Tamanic must have left South Sulawesi in ancient times, long enough ago to wander (probably in stages) to the remote area where they have ended up today and to lose collective memory of the event. The inclusion of Tamanic in the South Sulawesi group brings the number of WMP subgroups to (23).

5.13 CEMP and CMP

Blust (1993b) endeavors to give a stronger foundation to his CEMP, CMP and EMP hypotheses put forward in Blust (1978). (This 1993b article was already referred to by Ross (1995) as 'Blust 1990a', an unpublished manuscript.) Blust collects a set of phonological, lexical, morphosyntactic and semantic innovations that are unique to CEMP languages. He also presents such evidence for the establishment of CMP, and he tries to make a clearer boundary between CMP and South-Halmahera-West-New-Guinea languages in the West Papua (West New Guinea) area by showing that the Northern Bomberai languages (Sekar, Onin and Uruangnirin, spoken on the Bomberai peninsula on the south coast of West Papua) are not South-Halmahera-West-New-Guinea but CMP. There is some strong evidence that they directly subgroup with Yamdena (spoken on the Tanimbar Islands, Southeast Moluccas) and must be the result of a migration from the Tanimbar Islands to West Papua. At a higher level, Yamdena and Northern Bomberai subgroup with Kei and Fordata (Southeast Moluccas), forming what could be defined as the 'Southeast Maluku' subgroup (not to be confused with 'Southeast Maluku' as defined by Mills (1991)).

Blust admits that the innovations that are critical for CMP do not always feature in each member of the proposed CMP group. For instance, the truncation of PMP

word-final glides, (cf. PMP $*-ay$, $*-aw > -a$, and PAn $*-uy > -u$) is reflected in many, but not all, alleged CMP languages. Their overlapping distribution shows that CMP languages must have developed from a linkage, and not from a (basically uniform) protolanguage. But, as implied in Ross (1995:82), the partly overlapping distribution of various innovations does weaken the argument for CMP.

According to Blust, CEMP is easier to demonstrate than CMP. Nevertheless, the morphosyntactic evidence for CEMP is also problematic. Blust adduces two innovations: (1) the use of proclitic subject markers on the verb; and (2) a morphologically marked distinction between alienable and inalienable possession. He points out that the proclitic subject markers in Moluccan languages have a different form than those in West Melanesian languages. It is therefore possible that the subject markers in these two areas are not commonly inherited but the result of a convergent development, which potentially weakens the evidence for CEMP. Moreover, proclitic subject markers are also found in many Sulawesi languages, as Blust admits, and in Barrier Island languages (see JUKES, MAKASSAR; MEAD, MORIBAWAH; BROWN, NIAS). These are WMP languages, and although, here also, the sets are formally different than the Moluccan and West Melanesian ones, their occurrence shows that the phenomenon of proclitic subject marking is not unique to CEMP.

Blust considers the distinction between alienable and inalienable possession as stronger evidence. Nevertheless, this distinction is not unique to CEMP either. It has also been observed in Puyuma (Tsuchida 1995), where, however, it is expressed in a formally different way. Historically, it can also be demonstrated to have happened in various Bidayuh (or Land Dayak) languages in West Borneo. These languages have a petrified suffix *-n* or *-tn* (both $< *-n$), which can be explained as a historical inalienable noun suffix attached to terms for body-parts and kinship-relations ending in a vowel. Compare for instance the following evidence from Sungkung, which is spoken in six villages on the Sungkung mountain ridge in the Sambas and Sanggau regencies of West Kalimantan Province (Indonesia) along the Sarawak border (Adelaar personal fieldnotes):

PAn	<i>*maCa</i> 'eye'	> Sungkung	<i>batətn</i>
	<i>*qaCay</i> 'heart'		<i>atiin</i>
	<i>*qapəjuS</i> 'bile'		<i>əmpudutn</i>
	<i>*bulu</i> 'body hair'		<i>burutn</i>
	<i>*paqa</i> 'thigh'		<i>pa?ətn</i>
	<i>*bana</i> 'husband'		<i>banən</i>
	<i>*sawa</i> 'wife'		<i>sawətn</i>
	<i>*(ən)pu</i> 'grandchild'		<i>sikutn</i>
	<i>*su-aji</i> 'younger sibling'		<i>siditn</i>

Note that Sungkung final nasals become preloaded ($*-n > -tn$) unless the preceding consonant is also a nasal. This inalienable suffix does not appear in body-part and kinship terms already ending in a consonant, which includes non-historical glottal stops as in *sinə?* 'mother' ($< *si + *ina$) and *sama?* 'father' ($*si + *ama$).

The suffix is also found in other Bidayuh languages such as Biatah (cf. Nais 1988) and Singhi (Reijffert 1956), which are both spoken in Sarawak. However, not all Land Dayak languages have the inalienable suffix, and those that have do not always show the same distribution as found in Sungkung. But then again, lack of uniformity in exhibiting an alienable/inalienable distinction – as well as in the extent to which it is exhibited – is also typical for CEMP languages. The innovation of a grammaticalized distinction between alienable and inalienable possession would be strong subgrouping evidence for CEMP if it were a single construction type involving the same basic marker(s) throughout the area where the

alleged CEMP languages are spoken. What one finds instead, however, is quite a number of significant differences in possession constructions, as appears from a brief comparison of the relevant sections in STEINHAUER, BIAK; KLAMER, KAMERA; WILLIAMS-VAN KLINKEN AND VAN ENGELHOVEN, LETI AND TETUN; BOWDEN, TABA.

Ross (1995) also doubts the evidence for EMP: although the two putative branches within this subgroup, the South-Halmahera-West-New-Guinea group and Oceanic, are each phonologically well-attested, there is essentially no phonological evidence to support the claim that together they form an exclusive higher-level subgroup (Ross 1995:84–85).

Ross does not give a further division of the South-Halmahera-West-New-Guinea group. On the basis of the same practical criteria he used for establishing 24 WMP subgroups, he divides CMP into seven subgroups, as follows:

- 1 Bima-Sumba (eastern Sumbawa, Sumba, Flores);
- 2 Timor (Timor and Waimaha groups);
- 3 Southeast Maluku (Tanimbar and Kai Island groups);
- 4 Aru;
- 5 Central Maluku (Seram, Buru and their offshore islands);
- 6 North Bomberai (south coast of MacCluer Gulf, West Papua); and
- 7 Kowiai (south coast of Bird's Neck, West Papua).

Some further subgrouping has been done with regard to the Timor group. Van Engelenhoven puts Tetun and Waimaha together with the 'Luangic-Kisaric' languages in a South-East-Timor sub-branch of the Timor group. The Luangic-Kisaric languages are further divided into Kisaric (Roma and Kisar) and Luangic (Wetan, Leti and Nuclear Luangic, which in turn consists of Luang, Lakor and Moa) (Van Engelenhoven 1995:17). This classification also seems to be borne out by lexicostatistical percentages presented in Taber (1993).

In a rather unusual approach, Hull (1998) considers the Timor languages to be heavily creolized, linking them genetically with the 'Celebic' languages (primarily the Muna-Buton and Bungku-Mori languages) and relegating CMP features to a substratum.

To summarize, the classification of Austronesian languages remains uncertain. A systematic application of phonological criteria to Formosan languages has shown that their previous division into three primary branches does not hold and that there are probably many more primary branches. The Malayo-Polynesian subgroup seems well-defined, but its internal structure remains unclear. The South-Halmahera-West-New-Guinea and Oceanic subgroups are reasonably well-founded, and some progress has been made in lower-level subgrouping. However, a major branch such as WMP misses a linguistic foundation. Other major branches, such as CEMP, CMP and EMP, also remain contested and need further investigation.

6 THE Austronesian Homeland and Austronesian Migrations

The homeland issue is much less controversial among scholars than is the issue of classification. The study of Austronesian prehistory is moreover remarkable for the successful multidisciplinary approach that has been taken by various linguists, archaeologists and anthropologists over the last two decades.

It is generally accepted among these scholars that Taiwan is the area from which Austronesian speakers dispersed, if it is not the Austronesian homeland itself. However, before the Neolithic Austronesians settled in Taiwan more than 6000 years ago they must

have come from coastal South China (from present-day Fujian or Guangdong, Bellwood 1997:214).

Bellwood (1997) provides an updated and comprehensive multidisciplinary account of Austronesian prehistory based on combined research results from archaeology, linguistics, biological and social anthropology, and the paleoenvironmental sciences. It forms the basis of the present section.

Crucial for an appreciation of this prehistory is the development of agriculture in East Asia. Bellwood reminds us that agriculture originated autonomously in only a very few centers in the world. The areas around the Yellow River and Yangzi River seem to represent such a center (or possibly two independent centers) in East Asia. Millet was first cultivated along the Yellow River, and rice in the Yangzi River basin. These rivers run through much of central China, but 8000 years ago, when systematic agriculture began in this region, it was populated by a variety of ethnic groups. Some of these may have been among the ancestors of modern Chinese, but others must have been more Southeast Asian in cultural affiliation (Bellwood 1997:205).

From here agriculture must have spread to other parts of East Asia, including Southeast Asia. It must have spread to large parts of the present-day Austronesian world via Taiwan, where it was most prominent in the western part of the island. Neolithic sites representing the so-called Ta-pen-k'eng (Da-ben-keng) culture usually have radiocarbon dates going back to some 5000 years ago, although some sites are as old as 6300 years. The Ta-pen-k'eng culture is clearly linked to mainland Chinese culture complexes, and it is characterized by cord-marked and incised pottery and polished stone tools. It diversified later on in various other cultures in Taiwan, some of which are identified by Bellwood as Austronesian. These cultures include evidence for red-slipped and stamped pottery, and for 'stone net weights, hoes and possibly bones of domesticated pigs and dogs' (Bellwood 1997:219). Strangely enough, for a long time Ta-pen-k'eng and related sites did not provide evidence for cereal cultivation prior to 4500 years ago in Taiwan. However, Tsang (2001:11) recently found charred rice grains in a newly discovered Ta-pen-k'eng site, the Nan-kuan-li site near Tainan in south-western Taiwan.

The cultures that succeeded Ta-pen-k'eng in Taiwan have clear relatives in the Philippines and Indonesia (Bellwood 1997:ch. 7). Various Neolithic sites dating as far back as 2500 BC are found in the Philippines (cf. Dimolit, an open site in northern Luzon). Sites are found in Indonesia in the Talaud Islands north of Sulawesi (possibly 2500 BC but uncertain), in west Central Sulawesi (undated), on Kayoa Island to the west of Halmahera in the Moluccas (1200–300 BC), in Sabah (1500–500 BC), and on Timor (2500–2000 BC). These sites leave a fairly clear trail of red-slipped ceramics including decorated and (often) globular vessels, and furthermore of neolithic stone flake tools and bones of pigs and other animals.

After circa 1400 BC this chain of historically related archaeological sites continues eastward through Melanesia and (past the Solomons) into previously uninhabited territory. A clear series of sites ranges over a distance of 6500 km from the Admiralty Islands north of New Guinea to Samoa, representing the so-called Lapita culture which existed between 1400 and 800 BC. These thoroughly studied sites are well-dated and again contain red-slipped pottery as well as the bones of domesticated animals (including pigs, dogs and fowl).

The trail of sites becomes less clear towards western Indonesia. Java and Sumatra appear to be rather poor in Neolithic sites, which may be because 'Neolithic sites along former northern coastlines are now likely to be buried under many meters of alluvium and beneath the water table...and hence unavailable for archaeological research'

(Bellwood 1997:231). The Neolithic sites in peninsular Malaysia are unrelated to the Austronesian ones found in Sabah, Sulawesi and East Indonesia: they are linked to the immigration of Austro-Asiatic speakers. The origins of the latter, incidentally, also seem to be in present-day South China (Bellwood 1997:ch. 8).

The early Austronesians combined agriculture with maritime subsistence, animal husbandry (involving pigs, but also dogs and possibly other animals), and a technology that included canoes, wooden houses and pottery (possibly also weaving technology). Although generally agriculturalists, Austronesian peoples often continued to combine their agricultural skills with other economic resources such as hunting and fishing. Comparative linguistic evidence suggests that the Austronesians planted rice, sugarcane and gourds (Bellwood 1997:20). However, the rice that they must have brought along with them when they moved southward did not spread into Oceania and had difficulty at first adapting to large parts of insular Southeast Asia because equatorial climates were not favorable to its cultivation.

The peoples that lived in insular Southeast Asia before the arrival of Austronesians were evidently of an Australo-Melanesian or Negrito appearance (Bellwood 1997:ch. 7, cf. also Pawley 1999:105). Some of them had edge-ground stone axes and shell adzes but they did not use pottery, and they were not agriculturists. Some may have combined some form of plant management in their foraging economies (planting tubers or fruit trees) but this did not lead to the development of full-fledged agricultural societies. As a result, they were not equipped to resist the spread of Austronesian-speaking agriculturists, who were sedentary (and technologically more advanced). This led to their gradual assimilation, although some pre-Austronesian groups still survive in the (nowadays Austronesian-speaking) Negritos in the Philippines, the Orang Asli in Malaysia (some of whom are Mon-Khmer-speaking, and some of whom are hunter-gatherers), and Melanesian peoples in eastern Indonesia and Melanesia. One notable exception to this trend is New Guinea, which had an agricultural complex developed independently perhaps 10,000 years ago. Typically, the Austronesian influence only affected some coastal areas of the island and never penetrated it as thoroughly as it did elsewhere in Melanesia.

The Austronesian dispersal was slow given the fact that, after the Austronesians left Taiwan, it took them circa 4000 years to reach New Zealand (1200 AD) and Madagascar (possibly 700 AD). However, Bellwood points out that within this time span there was very much activity between 2000 BC and 1000 BC (roughly between the dates obtained from the sites in the northern Philippines and the period of colonization of Melanesia, western Micronesia and western Polynesia (Bellwood 1997:311)). It appears that the oldest dates for red-slipped and plain pottery anywhere south of Taiwan are not much older than 2000 BC, with the possible exception of Luzon. As Austronesians were already in Samoa (West Polynesia) by 1000 BC, Austronesian migrations must have taken place within a relatively short time within the second millennium BC (Bellwood 1997:232). Pawley suggests that there is a correlation between the long period of rest before the Austronesians moved from Taiwan to the Philippines (in the second millennium BC) and the development of an Austronesian linkage. A similar process must have occurred later on, when the (Polynesian) Austronesians paused for a thousand years in western Polynesia before they began to occupy eastern Polynesia, which happened between 300 and 1200 AD (Pawley 1999:112) or between 700 and 1200 AD (Bellwood 2001:12).

Blust speculates that the pause of at least a millennium between the settlement of Taiwan and that of the northern Philippines may be due to an initial lack of suitable boat technology. The first Austronesian settlers in Taiwan had bamboo sailing rafts, which did not enable them to move on to make the more difficult passage to the northern

Philippines. This passage would have been greatly facilitated by the invention of outriggers, which was apparently a late one. Blust is not sure whether outriggers were used prior to the break-up of PMP but he points out that many shipping terms can be reconstructed for PMP, but not for PAn. Outrigger canoes permitted a rapid settlement of Southeast Asia and Melanesia, where islands are 'relatively close-set' (Blust 1999:77). Blust relates the second long pause in Austronesian migrations, between the settlement of western Polynesia (and Fiji) and the settlement of East Polynesia, to the invention of the double canoe, which he believes occurred in the Fiji-Tonga-Samoa triangle.

There are presently no Austronesians living on the southern Chinese mainland. Blust points out that this area has been subject to a sinicization process throughout the last 2500 years of Chinese history. The gradual sinicization of Austronesian peoples in Taiwan is but a late rehearsal of this process. It has brought about an almost complete acculturation of earlier ethnic groups, and loss of their original languages. Therefore, Blust argues, even if there are no traces of an earlier Austronesian-speaking population on the mainland, it is not difficult to imagine that the Austronesian homeland was originally larger and included some mainland coastal areas, the P'eng-hu (Peng-hu, Pescadores) Islands and Taiwan. It is after all unlikely that the Austronesian migration to Taiwan from the outset would have left no trace in the earlier habitat on the mainland (Blust 1999:69).

7 LANGUAGE CONTACT

7.0 Introduction

This section gives only a very general and incomplete outline of language contact in Madagascar and the Asian part of the Austronesian world. It should be read in conjunction with Chapter 3 (this volume), and for further information and bibliographical references the reader is referred to Dutton and Tryon (1994) and to Wurm, Mühlhäusler and Tryon (1996; especially its sections on (insular) Southeast Asia, the Philippines and Taiwan).

In order to make the complicated issue of language contact more accessible to the reader, this section is further divided into three parts, dealing with influence from sources outside the Austronesian world (basically influence from languages from India, China, the Middle East and Europe, §7.1), language contact in areas bordering on the Austronesian world (contact between Austronesian and non-Austronesian languages in Austronesian border areas, §7.2), and contact among Austronesian languages (§7.3).

7.1 Influence from sources outside the Austronesian world (in chronological order)

The Malay peninsula and Indonesia are centrally located at a crossroads connecting three trade networks of global importance: the Indian Ocean, the China Sea, and the Spice Islands in East Indonesia. As a result, they became recipients of strong cultural and linguistic influences from India, China, the Middle East and Europe. Malay kingdoms along the Strait of Malacca and Javanese kingdoms became the main political forces in early insular Southeast Asian history, and it is largely through their languages (particularly Malay) that the linguistic influence from other civilizations found its way into other Malayo-Polynesian languages, including Malagasy and Philippine languages.

The Indian subcontinent was the first of these civilizations to have an impact. The first written evidence of this impact is some Sanskrit inscriptions from the fourth–fifth

century AD. Linguistically, Indian influence is shown by the great number of Sanskrit loanwords in the languages under investigation, and by loanwords from contemporary languages in India (cf. Gonda 1973). Of these, the most important language was Tamil (which was also the vehicular language for the spread of part of the Sanskrit loanwords), although other languages from northern and southern India as well as Sinhalese also left their trace (Gonda 1973:80, Asmah Haji Omar 1975:303–335, Van Ronkel 1902). Influence from these other Indian languages remains a seriously under-studied area. A recent annotated list of Sanskrit loanwords in Malay is De Casparis (1997) which is reviewed thoroughly by Mahdi (2000).

Indian languages had a direct impact on Cham, Malay, Javanese, Balinese and (especially Tamil influence) Karo Batak (cf. Tideman 1936). Their influence affected other Austronesian languages indirectly, predominantly via Javanese (especially in the case of Sundanese, Madurese, (early) Balinese, Sasak and modern Indonesian) and Malay (in most other cases, including Philippine languages, cf. Wolff (1976), and Malagasy, cf. Adelaar (1989, 1994b, 1995a, 1995b)).

With the establishment of Persian rule in Northern India (thirteenth–seventeenth century AD), this area also became the center of the dispersal of Persian and early Arabic lexical influence on Indonesian and Philippine languages. It was only from the seventeenth century onwards that this role was taken over by Hadramaut and other trade and religious centers on the Arabian peninsula and (nowadays) also by Islamic centers elsewhere in the Middle East. The latter are currently having an enhanced effect on Malay/Indonesian and other languages of Muslim communities. Here again we see that particularly Malay (but in some cases also Javanese) became the main vehicles for the spread of Arabic and Persian lexicon in other parts of Southeast Asia.

Chinese influence in Southeast Asia is at least as old as the influence from Arabic and Persian, if not older. Chinese has exercised a lexical influence on various Malay dialects (especially Jakarta Malay, Bazaar Malay and Baba Malay, which is originally the language of a mixed Chinese-Malay community in Singapore, Malacca and Penang). There are also various speculations about the possible role of Chinese in the genesis of trade Malay (cf. Adelaar and Prentice 1996).

Chinese influence is strongest in Malaysia and Singapore, where the Chinese are a very large proportion of the population as a result of mass migrations especially from the late eighteenth century onwards. Various dialects assume the role of contact language: Hokkien (in most places), Hakka (Sabah), Cantonese (in Kuala Lumpur, Seremban, Kuantan). Nowadays Mandarin is gaining ground at the cost of other Chinese varieties (especially in Singapore).

Other parts of Southeast Asia have been affected in different ways. In Indonesia, Chinese migrations date back to the thirteenth century but they increased in the 1860s. This resulted in two communities, an acculturated mixed Chinese-Indonesian (*Peranakan*) one that speaks a local Indonesian language, and a more authentic Chinese (*Totok*) one speaking usually Hokkien but also Hakka (Bangka Island, West Java, West Borneo), Teochew (West Borneo, East Sumatra and Riau) or even Cantonese (Bangka, Borneo). Bangka also developed a creolized Hakka dialect with many Malay loanwords (Skinner 1963:104). Chinese influence in Indonesia is limited in extent and scope in comparison with Indian, Arabic and European influence. Most of it is from the Hokkien (Fujian) dialect and belongs to the domains of food, building and places, shipping, numerals and measures, and concepts pertaining to Chinese subculture (Kong 1987:459). For Chinese influence in the Philippines, see Chan Yap (1980).

Borrowing from European languages began with the arrival of the Portuguese in Southeast Asia. They were soon followed by the Spaniards who, while initially also

making efforts to get strongholds in Brunei and East Indonesia (especially Halmahera), assumed control over the Philippines and some islands in Micronesia (including Chamorro). The Portuguese and Spanish had a great impact on the linguistic geography of the areas under their respective administrations. They left many loanwords, which is probably due to the fact that they were the first Europeans to arrive in the area and therefore the first to introduce typically European objects and concepts (including Christianity). But they also left creole versions of their languages in the form of Portuguese-based creoles in Malacca (Baxter 1988), in Tugu near Jakarta (Schuchardt 1890) and in Bidau in East Timor (Baxter 1990), and in the form of Chavacano, which is a basically a cover term for various Spanish-based creoles (Lipski 1996:276–280). Portuguese remained the official language of East Timor until the Indonesian invasion in 1975 and has been re-established as such (along with Tetun) since its independence (Hajek 2000; STEINHAUER, LANGUAGE POLICY). It is becoming again a major source of borrowing into Tetun (Williams-Van Klinken, Hajek and Nordlinger 2002). The Dutch imprint on the languages of Indonesia and Malaysia was slight by comparison, considering that in most areas Dutch rule lasted much longer than Portuguese rule. The Dutch language nevertheless left many loanwords in Indonesian and in vernacular Indonesian languages, and also some in Malaysian Malay. After Indonesian independence there was a concerted effort by the Indonesian Language Center to replace Dutch loanwords with neologisms based on inherited Malay, Sanskrit, Arabic and Old Javanese lexicon. As a result, many Dutch loanwords disappeared from the Indonesian lexical inventory, although some of them persist in non-standard forms of Malay or in vernacular languages. Some Dutch creoles (such as Malay-based *Peco*' and Javanese-based *Javindo*) have become extinct or are on the verge of disappearance.

French influence played a major role in areas where the French were a colonial power, that is, in Madagascar and in Vietnam and Cambodia, where Chamic languages are spoken. It continues to play this role in Madagascar where it is in competition with Malagasy, the national language (cf. STEINHAUER, LANGUAGE POLICY). Malagasy has many French loanwords (and some English ones, see below). On the other hand, it seems that Malagasy has remained somewhat more resistant to European lexical influence than Malay and Tagalog. Finally, many French loanwords ended up in Malay and other Indonesian languages via Dutch in the colonial period.

Borrowing from English probably started in the seventeenth century (when the British had representations in Aceh and Bengkulu, in Sumatra) but it became a major lexical source in late eighteenth-century Malaysia and early twentieth-century Philippines. Malagasy also shows clear traces of borrowing from English, which stem from the nineteenth century when the British tried to get a foothold on the island and competed with the French to establish a colonial administration. In both Malaysia and the Philippines English influence continues in spite of the establishment of a native national language (see STEINHAUER, LANGUAGE POLICY). In this modern age of globalization, English has in general become the major source for lexical borrowing in the national languages of Southeast Asia (including Indonesian).

7.2 Language contact in areas bordering on the Austronesian world (in order from west to east)

Other manifestations of language contact are more local in their scope.

Malagasy has been in contact with Bantu languages in several stages of its post-migratory (African) history. Dahl (1988) identifies this influence as being from Comoran Swahili (Ngazije and Ndzuani).

Various Austronesian languages on the Southeast Asian mainland as well as Sumatra exhibit influence from Austro-Asiatic languages. This is clearly manifested in the phonology and lexicon of Chamic languages, including Acehnese (Thurgood 1999:47–59). Acehnese, in turn, affected the phonologies of Gayo (Eades in press) and Alas (a northern Batak variety). Austro-Asiatic influence also appears in Malayic: Malay words like *səmət* ‘ant’, *kətam* ‘crab’, *əlay* ‘kite’, *mərak* ‘peacock’, *pərət* ‘stomach’ are not Austronesian and are shared with Austro-Asiatic languages (Shorto 1975, Thurgood 1999:360, cf. also Collins (1985) detailing features of a Malay dialect spoken by Orang Asli in Pahang). Like the Chamic languages, Malayic varieties such as Kerinci and Minangkabau as well as the WMP isolate language Rejang have undergone radical phonological changes in their last syllables. Some of these manifest areal features have also affected non-Austronesian languages on the mainland. In turn, Orang Asli languages exhibit influence from Malay.

Thai has been affecting the Malay dialects of southern Thailand and of the northern states of West Malaysia for a long time, although its influence is mainly lexical and hardly phonological (Tadmor 1995:14). Thai influence is also manifest in Moken and Moklen (see LARISH, MOKEN AND MOKLEN), and Burmese has influenced Moken (Naw Say Bay 1995). There has also been considerable lexical borrowing from Malay into Thai (Suthiwat 1992). Thurgood (1994) points to Austronesian influence on Tai-Kadai. Lexical borrowing probably took place in prehistorical times when the speakers of both language groups were still living in mainland China.

In the northernmost regions of the area under investigation, the Formosan languages have been in contact with local (Southern Min or Hokkien) Chinese for at least four centuries. For instance, the Siraya translation of the Gospel of St. Matthew (Gravius 1661) already contains several local Chinese loanwords. The Formosans have also been in contact with Japanese (during the occupation by the Japanese between 1895 and 1945) and with Mandarin, which have both served as linguae francae and have both left traces in Formosan languages (Li 1996:741). Tai-Kadai languages and Southern Min Chinese also influenced Tsat, a Chamic language spoken in Hainan (China) (Thurgood 1999:22).

Eastern Indonesia and East Timor form an interface area between Austronesian and various Papuan linguistic spheres. If basic vocabulary is used as an indicator, the languages in this area are overwhelmingly Austronesian, except in large parts of New Guinea and in North Halmahera, in a few limited areas in East Timor and on the islands of Alor, Pantar and Kisar in Indonesia. However, as far as grammatical typology is concerned, Austronesian languages in eastern Indonesia and East Timor have features that are quite different from those of the more conservative Austronesian languages of Taiwan, the Philippines and western Indonesia. These features differ to the extent that in a number of cases the classificatory labels have become almost insignificant, as pointed out by Capell (1982), Foley (1986, 1998), Hull (1998) and others who have worked in the area. CMP and SHWNG languages have in all likelihood left their traces on neighboring non-Austronesian languages. What is certain, however, is the (lexical) influence from Indonesian and regional (Amboinese, Ternatan) Malay as well as Tetun on neighboring non-Austronesian languages.

Austronesian languages exerted some influence on Australia. In North Australia (between Bathurst and Melville Islands to the west and Limmen Bight to the east), contact between Makassarese (collecting sea urchins) and Aboriginals caused the emergence of a Makassar-based pidgin language, ‘Macassan’ which is still testified to in a number of lexical items remembered by local Aboriginals or borrowed into their languages. However, some of the alleged Macassan loanwords seem to be of Malay rather than Makassar

provenance (Urry and Walsh 1981:94–95). Makassar as well as Malay loanwords were also found in Yolngu-Matha (Northeast Arnhem Land; Walker and Zorc (1981)), and in languages of West Arnhem Land (Evans 1992, 1997).

7.3 Contact among Austronesian languages (contact languages in order from West to East)

Traditionally, the language area under investigation is by and large characterized by the presence of Austronesian regional contact languages that are used throughout a geographically delimited area (an archipelago, island, a river and its tributaries, a coast area, etc.). Since decolonization there has been a trend for these contact languages to be replaced by newly established national languages such as Indonesian, official Malaysian Malay, Pilipino, Thai, Vietnamese, etc.

Among these contact languages there are quite a few regional forms of Malay. The latter also occur in eastern Indonesia, where they include Ambon Malay (South Moluccas and in a subvariant in West Papua), Ternate Malay (North Moluccas) and Manado Malay (North Sulawesi). In the past, Malay was used as a contact language in the Philippines and, most likely, Madagascar. In Indonesia, one Malay contact language seems to have resisted the takeover by the national language rather well. Although Betawi, the original Malay dialect of Jakarta, is nowadays spoken by only a small minority, it has influenced official Indonesian to the extent that Jakartan Indonesian has emerged, a variant of Indonesian which is much more vibrant and popular than its formal counterpart. Another Malay contact language, Banjar Malay has had an expansion along the coast of East Kalimantan.

Malay has served various sociolinguistic purposes, including that of interethnic lingua franca, trade language, language for religious instruction, literary language, educational language, etc. An overview of Malay contact languages is Adelaar and Prentice (1996:676–688). See ADELAAR, MALAYIC for an insight into some of the linguistic variety of Malayic, Collins (1998) for a recent overview of the history of Malay, and Collins (1990, 1995a, 1995b, 1996b) for an annotated bibliography of Malay varieties.

Javanese is the Austronesian language with by far the highest number of mother-tongue speakers (see OGLOBLIN, JAVANESE and FLOREY, ENDANGERED LANGUAGES). It is also the most important regional language in Indonesia and has had an overwhelming impact on Indonesian lexicon. It has had an important role in the linguistic history of island Southeast Asia. In the fourteenth century AD, Java extended its political influence over many other areas in Indonesia and at one point even included parts of the Philippines and some of the coast of West Papua. However, in contrast to Malay, the influence of Javanese did not extend beyond the spread of loanwords, except for the languages in its vicinity. Of the latter, Sundanese, Madurese, Balinese and Sasak emulated the Javanese use of polite speech registers. The vocabulary of these registers is moreover largely borrowed from Javanese. Palembang Malay court jargon is largely based on Javanese lexicon. Banjar Malay adopted many Javanese loanwords (including much basic vocabulary) but it did not develop polite language registers. Javanese loanwords are occasionally also found in languages further away from Java (including languages as far away as Siraya (Taiwan) and Malagasy, cf. Adelaar 1994b).

In addition to Makassar influence on Australian languages (see above), lexical influence from South Sulawesi languages is found in various languages of Sulawesi, the Lesser Sunda Islands (Lombok, Sumbawa, and Flores), and Madagascar. South Sulawesi loanwords in Malagasy are most easily explained as having entered Malagasy and other

members of the Southeast Barito language group before the early Malagasy left Borneo (Adelaar 1995a:350–351).

For the roles of Indonesian, Tagalog, Tetun, Malagasy and other national languages, see the sketches of these languages and Chapter 3 (this volume). For various other Austronesian contact languages in Asia, see the *Atlas of languages of intercultural communication* by Wurm, Mühlhäusler and Tryon (1996).

ACKNOWLEDGMENTS

I would like to express my gratitude to Peter Bellwood, John Hajek, Nikolaus Himmelmann, David Mead, Malcolm Ross and Elizabeth Zeitoun for their valuable comments on earlier versions of this chapter. They are in no way responsible for any errors in the present version.

REFERENCES

- Adelaar, K.A. (1985) *Proto-Malayic: a reconstruction of its phonology and part of its morphology and lexicon*, PhD thesis, Leiden University.
- (1989) ‘Malay influence on Malagasy: linguistic and culture-historical inferences’, *Oceanic Linguistics*, 28/1:1–46.
- (1992) *Proto-Malayic: a reconstruction of its phonology and part of its morphology and lexicon*, Canberra: Pacific Linguistics (revised publication of Adelaar 1985).
- (1993) ‘The internal classification of the Malayic subgroup’, *Bulletin of the School of Oriental and African Studies*, 56/3:566–581.
- (1994a) ‘The classification of the Tamanic languages (West Kalimantan)’, in T.E. Dutton and D. Tryon (eds), 1–41.
- (1994b) ‘Malay and Javanese loanwords in Malagasy, Tagalog and Siraya (Formosa)’, *Bijdragen tot de Taal-, Land- en Volkenkunde*, 150/1:49–64.
- (1995a) ‘Asian roots of the Malagasy: a linguistic perspective’, *Bijdragen tot de Taal-, Land- en Volkenkunde*, 151/3:325–357.
- (1995b) ‘Bentuk pinjaman bahasa Melayu dan Jawa di Malagasi’, in Ismail Hussain, A. Aziz Deraman and Abd. Rahman Al Ahmadi (eds), *Tamadun Melayu. Jilid Pertama*, Kuala Lumpur: Dewan Bahasa dan Pustaka; 21–40.
- (in press a) ‘On the classifiability of Malayic’, in D. Gil (ed.), *Malay and Indonesian Linguistics*, London: Routledge-Curzon.
- (in press b) ‘Malayo-Sumbawan’, in P. Austin (ed.), *Working Papers in Bali, Lombok and Sumbawa*, London: School of Oriental and African Studies.
- (2004) ‘Where does Malay come from? Twenty years of discussions about homeland, migrations and classifications’, *Bijdragen tot de Taal-, Land- en Volkenkunde*, 160/1:1–30.
- Adelaar, K.A. and Prentice, D.J. (1996) ‘Malay: its history, role and spread’, in S.A. Wurm, P. Mühlhäusler and D. Tryon (eds), 673–693.
- A primer to Mangyan script. *Surat Mangyan* (1986), Panaytayan, Mansalay-4411, OR. Mindoro, Philippines.
- Asmah Haji Omar, (1975) *Essays on Malaysian Linguistics*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Baxter, A.N. (1988) *A Grammar of Kristang (Malacca Creole Portuguese)*, Canberra: Pacific Linguistics.
- (1990) ‘Notes on the Creole Portuguese of Bidau, East Timor’, *Journal of Pidgin and Creole Languages* (Amsterdam/Philadelphia: Benjamins), 5:1/1–38.
- Bellwood, P. (1997) *Prehistory of the Indo-Malaysian Archipelago* (revised edition) Honolulu: University of Hawai'i Press.
- (2001) Polynesian prehistory and the rest of mankind, in C.M. Stevenson, G. Lee and F.J. Morin (eds), *Pacific 2000. Proceedings of the Fifth International Conference on Easter Island and the Pacific*, The Easter Island Foundation, 11–25, Los Osos (CA): Bearsville Press.
- Bellwood, P., Fox, J.J. and Tryon, D. (eds) (1995) *The Austronesians: history and comparative perspectives*, Canberra: Dept. of Anthropology, Australian National University, RSPAS.
- Blust, R.A. (1970) ‘Proto-Austronesian addenda’, *Oceanic Linguistics* 9:104–162.
- (1976a) Dempwolff’s reduplicated monosyllables, *Oceanic Linguistics*, 15:107–130.
- (1976b) Austronesian culture history: some linguistic inferences and their relation to the archaeological world, *World Archaeology*, 8/1:19–43.
- (1977) ‘The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report’, *Working Papers in Linguistics of the University of Hawai'i*, 9/2:1–15.
- (1978) ‘Eastern Malayo-Polynesian: a subgrouping argument’, in S.A. Wurm and L. Carrington (eds), *Second International Conference on Austronesian Linguistics: proceedings, Fascicle I, Western Austronesian*, 181–234, Canberra: Pacific Linguistics.
- (1979) ‘Proto-Western-Malayo-Polynesian vocatives’, *Bijdragen tot de Taal-, Land- en Volkenkunde*, 135:205–251.
- (1980a) ‘Austronesian etymologies I’, *Oceanic Linguistics*, 19:1–181.
- (1980b) ‘Early Austronesian social organization: the evidence of language’, *Current Anthropology*, 21/2. 205–47.
- (1980c) ‘Notes on Proto-Malayo-Polynesian phrathy dualism’, *Bijdragen tot de Taal-, Land- en Volkenkunde*, 136:215–247.
- (1981) ‘The reconstruction of Proto-Malayo-Javanic: an appreciation’, *Bijdragen tot de Taal-, Land- en Volkenkunde*, 137:456–469.
- (1982) ‘The linguistic value of the Wallace Line’, *Bijdragen tot de Taal-, Land- en Volkenkunde*, 138:231–250.
- (1983–4) ‘Austronesian etymologies II’, *Oceanic Linguistics*, 22–23:29 149.
- (1984) ‘On the history of Rejang vowels and diphthongs’, *Bijdragen tot de Taal-, Land- en Volkenkunde*, 140/4:422–450.
- (1984–5) ‘The Austronesian homeland: a linguistic perspective’, *Asian Perspectives* (Honolulu), 26/1:45–67.
- (1986) ‘Austronesian etymologies III’, *Oceanic Linguistics*, 25:1–123.
- (1988) *Austronesian root theory. An essay on the limits of morphology*, Studies in Language Companion Series 19, Amsterdam/Philadelphia: John Benjamins.
- (1989) ‘Austronesian etymologies IV’, *Oceanic Linguistics*, 28:111–180.
- (1990) ‘Patterns of sound change in the Austronesian languages’, in Ph. Baldi (ed.), *Linguistic change and reconstruction methodology*, 231–267, Berlin: Mouton de Gruyter.
- (1991) ‘The Great Central Philippines hypothesis’, *Oceanic Linguistics*, 30:73–129.
- (1993a) ‘*S metathesis and the Formosan/Malayo-Polynesian language boundary’, in Øyvind Dahl (ed.), *Language – a doorway between human cultures: tributes to Dr. Otto Christian Dahl on his ninetieth birthday*, 178–183, Oslo: Novus.
- (1993b) ‘Central and Central-Eastern Malayo-Polynesian’, *Oceanic Linguistics*, 32/2:241–293.
- (1994a) ‘Obstruent epenthesis and the unity of phonological features’, *Lingua*, 93:111–139.
- (1994b) ‘The Austronesian settlement of mainland Southeast Asia’, in K.L. Adams and T.J. Hudak (eds), *Papers from the Second Annual Meeting of the Southeast Asian Linguistics Society*, Tempe (Arizona): Program for Southeast Asian Studies, Arizona State University, 25–83.

- Blust, R.A. (1995) 'The position of the Formosan languages: methods and theory in Austronesian comparative linguistics', in Li *et al.* (eds), 585–650.
- (1998a) 'Ca-reduplication and Proto-Austronesian grammar', *Oceanic Linguistics*, 37/1:29–64.
- (1998b) 'The position of the languages of Sabah', in Ma.L.S. Bautista (ed.), *Pagtanaw Essays on language in honor of Teodoro A. Llamzon*, 29–52, Manila: The Linguistic Society of the Philippines.
- (1999) 'Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics', in E. Zeitoun and P.J.-K. Li (eds), 31–94.
- (2002) 'Formalism or phoneyism? The history of Kayan final glottal stop', in K.A. Adelaar and R. Blust (eds), *Between worlds: Linguistic papers in memory of David John Prentice*, 29–37, Canberra: Pacific Linguistics.
- (in press) 'Three notes on early Austronesian morphology', to appear in *Language and Linguistics*, (Taipei: Academia Sinica).
- (to appear) *Austronesian Comparative Dictionary* (unpublished data on computer file).
- Brandstetter, R. (1893) *Die Beziehungen des Malagasy zum Malaischen*, Luzern.
- Cameron, C.R. (1917) *Sulu writing*, Zamboanga (Philippines): Sulu Press.
- Capell, A. (1982) 'Bezirkssprachen im UAN-Gebiet', in Reiner Carle *et al.* (eds), *Gava' Studies in Austronesian languages and cultures dedicated to Hans Kähler*, Veröffentlichungen des Seminars für Indonesische und Südseesprachen der Universität Hamburg Band 17, Berlin: Reimer Verlag.
- Chan Yap, G. (1980) *Hokkien Chinese borrowings in Tagalog*, Canberra: Pacific Linguistics.
- Clynes, A. (1989) *Speech styles in Balinese and Javanese*, unpublished MA thesis from the Australian National University.
- Collins, J.T. (1985) 'Dialek Melayu di kampung Landai, Pahang: menuju penelitian tatabahasa Melayu di kalangan Orang Asli', *Dewan Bahasa* 29/7:476–493.
- (1990) *Bibliografi dialek Melayu di pulau Borneo*, Siri Monograf Bibliografi Sejarah Bahasa Melayu, Kuala Lumpur: Dewan Bahasa dan Pustaka.
- (1994) 'Sumbangan dialek Brunei dalam pengkajian sejarah Bahasa Melayu', in *Tinggal Landas ke Abad 21*, Bandar Seri Begawan: Dewan Bahasa dan Pustaka; 62–75.
- (1995a) *Bibliografi dialek Melayu di pulau Sumatera*. Siri Monograf Bibliografi Sejarah Bahasa Melayu, Kuala Lumpur: Dewan Bahasa dan Pustaka.
- (1995b) *Bibliografi dialek Melayu di pulau Jawa, Bali dan Sri Lanka*, Siri Monograf Bibliografi Sejarah Bahasa Melayu, Kuala Lumpur: Dewan Bahasa dan Pustaka.
- (1996a) 'Borneo and Maluku: the evidence from the language of Bacan', in P.W. Martin, C. Ozog and G. Poedjosoedarmo (eds), 73–88.
- (1996b) *Bibliografi dialek Melayu di Indonesia Timur*. Siri Monograf Bibliografi Sejarah Bahasa Melayu, Kuala Lumpur: Dewan Bahasa dan Pustaka.
- (1998) *Malay, World Language of the Ages. A sketch of its history*, Kuala Lumpur: Dewan Bahasa dan Pustaka.
- (2002) 'The study of Sarawak Malay in context', in K.A. Adelaar and R. Blust (eds), *Between worlds: Linguistic papers in memory of David John Prentice*, 65–76, Canberra: Pacific Linguistics.
- Dahl, O.C. (1976) *Proto-Austronesian*, Scandinavian Institute of Asian Studies Monograph series No. 15 (2nd edition), Lund: Studentlitteratur – Curzon press.
- (1981) *Early Phonetic and Phonemic Changes in Austronesian*, Instituttet for sammenlignende kulturforskning, Oslo: Universitetsforlaget.
- (1983) *Sorabe révélant l'évolution du dialecte antemoro*, Antananarivo: Trano Printy Loterana.
- (1988) 'Bantu substratum in Malagasy' (special issue: *Linguistique de Madagascar et des Comores*), *Études Océan Indien*, 9:91–132.
- (1991) *Migration from Kalimantan to Madagascar*, The Institute for Comparative Research in Human Culture, Oslo: Norwegian University Press.
- De Casparis, J.G. (1997) *Sanskrit loanwords in Indonesian. An annotated check-list of words in Indonesian and traditional Malay*, Jakarta: Badan Penyelenggara Seri Nusa (Universitas Katolik Atmajaya).
- Dempwolff, O. (1934, 1937, 1938) *Vergleichende Lautlehre des Austronesischen Wortschatzes* (3 vols), Beihefte Zeitschrift für Eingeborensprachen 15, 17, 19, Berlin: Dietrich Reimer Verlag.
- Donohue, M. (in press) 'The pretenders to the Muna-Buton group' (to appear in *Pacific Linguistics*, Canberra).
- Dutton, T.E. and Tryon, D.T. (eds) (1994) *Language Contact and Change in the Austronesian World*, Trends in Linguistics. Studies and Monographs. Berlin: Mouton – de Gruyter.
- Dyen, I. (1947a) 'The Malayo-Polynesian word for "two"', *Language*, 23:50–55.
- (1947b) 'The Tagalog reflexes of Malayo-Polynesian D', *Language*, 23:227–238.
- (1951) 'PMP *Z', *Language*, 27:534–540.
- (1953a) *The Proto-Malayo-Polynesian laryngeals*, Baltimore: Linguistic Society of America.
- (1953b) 'Dempwolff's R', *Language*, 29:359–366.
- (1962) 'Some new Malayopolynesian initial phonemes', *Journal of the American Oriental Society* 82:214–215.
- (1963) 'The position of the Malayo-Polynesian languages of Formosa', *Asian Perspectives*, 7:261–271.
- (1965a) 'Formosan evidence for some Proto-Austronesian phonemes', *Lingua*, 14:285–305.
- (1965b) *A Lexicostatistical Classification of the Austronesian Languages*, International Journal of American Linguistics, Memoir 19, Baltimore: Waverly Press.
- Dyen, I. and Tsuchida, S. (1991) Proto-Philippine as the nearest relative of Proto-Formosan, in Harlow, R. (ed.), *VICAL 2: Western Austronesian languages*, Auckland: Linguistic Society of NZ; 85–101.
- Eades, D. (in press) *A grammar of Gayo, a language of Aceh, Indonesia*, Canberra: Pacific Linguistics.
- Esser, S.J. (1938) 'Languages', in *Atlas van Tropisch Nederland*, Batavia: Koninklijk Nederlandsch Aardrijkskundig Genootschap.
- Evans, N. (1992) 'Macassan loanwords in Top End languages', *Australian Journal of Linguistics*, 12:45–91.
- (1997) 'Macassan loans and linguistic stratification in western Arnhem Land', in P. McConville and N. Evans (eds), *Archaeology and linguistics. Aboriginal languages in global perspective*, 237–260, Melbourne: Oxford University Press.
- Ferrell, R. (1969) *Taiwan Aboriginal Groups*:(...). Taipei: Academia Sinica, Institute of Ethnology.
- Foley, W.A. (1986) *The Papuan Languages of New Guinea* (Cambridge Language Surveys), Cambridge: Cambridge University Press.
- (1998) 'Towards understanding Papuan languages', in J. Miedema, C. Odé and R.A.C. Dam (eds), *Perspectives on the Bird's Head in Irian Jaya, Indonesia. Proceedings of the Conference, Leiden, 13–17 October 1997*, 503–518, Amsterdam: Rodopi.
- Fox, J.J. (ed.) (1993) *Inside Austronesian houses: perspectives on domestic designs for living*, Canberra: RSPAS, Dept. of Anthropology, Australian National University.
- (ed.) (1997) *The poetic power of place: comparative perspectives on Austronesian ideas of locality*, Canberra: RSPAS, Dept. of Anthropology, Australian National University.

- Fox, J.J. and Sather, C. (eds) (1996) *Origins, ancestry and alliance*, Canberra: RSPAS, Dept. of Anthropology, Australian National University.
- Gonda, J. (1973) *Sanskrit in Indonesia* (2nd edn), New Delhi: International Academy of Indian Culture.
- Gravius, D. (1661) *Het heylige Euangelium Matthei en Johannis. Ofte Hagnau ka d'lig matiktik ka na sasoulat ti Mattheus ti Johannes appa*, Amsterdam: Michiel Hartogh.
- Grimes, B.F. (2000) *Ethnologue: Languages of the world* (14th edition), Texas: Summer Institute of Linguistics.
- Hajek, J. (2000) 'Language planning and the sociolinguistic environment in East Timor: colonial practice and changing language ecologies', *Current Issues in Language Planning*, 1:400–414.
- Himmelmann, N.P. (2001) *Sourcebook on Tomini-Tolitoli languages. General information and wordlists*, Canberra: Pacific Linguistics.
- (2002) 'Voice in western Austronesian: an update', in F. Wouk and M. Ross (eds), 7–16.
- Hudson, A.B. (1978) 'Linguistic relations among Bornean peoples with special reference to Sarawak: an interim report', in *Sarawak. Linguistics and development problems*, 1–45, Studies in Third World Societies (Williamsburg, VA) No. 3.
- Hull, G. (1998) 'The basic lexical affinities of Timor's Austronesian languages. A preliminary investigation', *Studies in languages and cultures of East Timor Volume 1* (Academy of East Timor Studies, University of Western Sydney):97–202.
- Humboldt, Wilhelm von (1836–1839) *Über die Kawi-Sprache auf der Insel Java* (3 vols), Berlin.
- Kong Yuan Zhi (1987) 'A study of Chinese loanwords (from South Fujian dialects) in the Malay and Indonesian languages', *Bijdragen tot de Taal-, Land- en Volkenkunde*, 143/4:452–467.
- Larish, M.D. (1999) *The position of Moken and Moklen in the Austronesian language family*, PhD thesis, University of Hawai'i.
- Li, P. Jen-kuei (1996) 'The lingue franche in Taiwan', in S.A. Wurm, P. Mühlhäusler and D.T. Tryon (eds), *Atlas of Languages of Intercultural Communication in the Pacific, Asia, and the Americas vol. II.1*, Trends in Linguistics Documentation 13, 741–743, Berlin: Mouton de Gruyter.
- Li, P. Jen-kuei, Tsang, Cheng-hwa, Huang, Ying-kuei, Ho, Dah-an and Tseng, Chiu-yu (eds) (1995), *Austronesian Studies Relating to Taiwan*, Symposium Series of the Institute of History and Philology, Academia Sinica Number 3, Taipei: Academia Sinica.
- Lipski, J.M. [with Mühlhäusler, P.; Duthin, F.] (1996) Spanish in the Pacific, in S.A. Wurm, P. Mühlhäusler and D. Tryon (eds), 271–298.
- Lynch, J. (ed.) (in press) *Issues in Austronesian phonology*, Canberra: Pacific Linguistics.
- Mahdi, W. (1988) *Morphonologische Besonderheiten und historische Phonologie des Malagasy*, Veröffentlichungen des Seminars für Indonesische und Südseesprachen der Universität Hamburg Band 20, Berlin-Hamburg: Dietrich Reimer Verlag.
- (1994a) 'Some Austronesian maverick protoforms with culture-historical implications I', *Oceanic Linguistics*, 33/1:167–229.
- (1994b) 'Some Austronesian maverick protoforms with culture-historical implications II', *Oceanic Linguistics*, 33/2:431–490.
- (2000) review of J.G. De Casparis (1997), *Bijdragen tot de Taal-, Land- en Volkenkunde*, 156/4:844–852.
- Marrison, G.E. (1975) 'The early Cham language and its relation to Malay', *Journal of the Malayan Branch of the Royal Asiatic Society*, 48(2):52–59.
- Martin, P.W., Poedjosoedarmo, G. and Ozog, C. (eds) (1996) *Language use and language change in Brunei Darussalam*, Athens: Ohio University Centre for International Studies.
- Mbete, A.M. (1990) *Rekonstruksi protobahasa Bali-Sasak-Sumbawa*, PhD thesis, Universitas Indonesia (Jakarta).
- Mead, D. (1996) 'The evidence for final consonants in Proto-Bungku-Tolaki', *Oceanic Linguistics*, 35:180–194.
- (1998) *Proto-Bungku-Tolaki: reconstruction of its phonology and aspects of its morphosyntax*, PhD dissertation, Rice University, Houston, Texas.
- (1999) *The Bungku-Tolaki languages of Southeastern Sulawesi*, Indonesia, Canberra: Pacific Linguistics, Australian National University.
- (2002) 'Proto-Celebic focus revisited', in F. Wouk and M. Ross (eds), 143–177.
- (in press a) 'The Saluan-Banggai microgroup of eastern Sulawesi', in John Lynch (ed.) (in press).
- (in press b) 'Evidence for a Celebic supergroup', in John Lynch (ed.) (in press).
- Mills, R.F. (1975) *Proto-South-Sulawesi and Proto-Austronesian phonology*, PhD dissertation, University of Michigan, Ann Arbor.
- (1991) 'Tanimbar-Kei: an Eastern Indonesian subgroup', in R.A. Blust (ed.), *Currents in Pacific Linguistics. Papers in Austronesian languages and ethnolinguistics in honour of George W. Grace*, 241–263, Canberra: Pacific Linguistics.
- Nais, W. (1988) *Bidayuh-English dictionary*, Kuching: Persatuan Kesusastraan Sarawak.
- Naw Say Bay (1995) 'The phonology of the Dung dialect of Moken', in Bradley, D. (ed.), *Studies in Burmese languages*, 193–205, Papers in Southeast Asian Linguistics No.13, Canberra: Pacific Linguistics.
- Noorduyn, J. (1991) *A critical survey of studies on the languages of Sulawesi*. KITLV Bibliographical Series. Leiden: Royal Institute of Linguistics and Anthropology.
- (1993) 'Variation in the Bugis/Makasarese script', *BKI* 149/3:533–570.
- Nothofer, B. (1975) *The reconstruction of Proto-Malayo-Javanic* Verhandelingen van Koninklijk Instituut voor Taal-, Land-en Volkenkunde 73, The Hague: Martinus Nijhoff.
- (1985) 'The subgrouping of the languages of the Javo-Sumatran hesion: a reconsideration', *Bijdragen tot de Taal-, Land- en Volkenkunde*, 141/3–4:228–302.
- (1988) 'A discussion of two Austronesian subgroups: Proto-Malay and Proto-Malayic', in Mohd. Thani Ahmad and Zaini Mohamed Zain (eds), *Rekonstruksi dan cabang-cabang Bahasa Melayu Induk*, 34–58, Siri monograf sejarah bahasa Melayu, Kuala Lumpur: Dewan Bahasa dan Pustaka.
- (1994) '‘Bahasa Lom’ – Dari mana asalnya?' in L. Sihombing *et al.* (eds), *Bahasawan Cendekia: seuntai karangan untuk Anton M. Moeliono*, 193–208, Jakarta: P.T. Intermasa.
- (1995) 'The history of Jakarta Malay', *Oceanic Linguistics*, 34/1:86–97.
- (1996) 'Migrasi orang Melayu Purba: Kajian awal', *Sari*, 14:33–52.
- (1997a) *Dialek Melayu Bangka*. Bangi: Penerbit Universiti Kebangsaan Malaysia.
- (1997b) 'Klasifikasi varian Melayik di antara Sungai Semandung dan Sungai Pawan/Sungai Keriau (Gerai, Tanjung Beringin, Randau Jeka)', Paper: Seminar Internasional Bahasa dan Budaya di Dunia Melayu (Asia Tenggara). Mataram, 21–23 Juli 1997.
- Pawley, A.K. (1999) 'Chasing rainbows: implications for the rapid dispersal of Austronesian languages for subgrouping and reconstruction', in E. Zeitoun and P.J.-K. Li (eds), 95–138.
- Pawley, A.K. and Ross, M.D. (1994) *Austronesian terminologies: continuity and change*, Canberra: Pacific Linguistics.
- Postma, A. (1971) 'Contemporary Mangyan scripts', *Philippine Journal of Linguistics*, 2/1:1–12.
- Reid, L. (1982) 'The demise of Proto-Philippines', in A. Halim, L. Carrington and S.A. Wurm (eds), *Papers from TICAL vol.2*, 201–216, Canberra: Pacific Linguistics.
- (2002) 'Morphosyntactic evidence for the position of Chamorro in the Austronesian family', in R.S. Bauer (ed.), *Collected papers on Southeast Asian and Pacific languages*, Canberra: Pacific Linguistics.
- Reijffert, A. (1956) *Vocabulary of English and Sarawak Land Dyak (Singhi Tribe)*, Kuching: Sarawak Government Printing Office.

- Reuter, T.A. (forthcoming) *Sharing the earth, carving up the land*, Canberra: RSPAS, Dept. of Anthropology, Australian National University.
- Ross, M.D. (1995) 'Some current issues in Austronesian linguistics', in D.T. Tryon *et al.* (eds), 45–120.
- (2002) 'The history and transitivity of western Austronesian voice and voice marking', in F. Wouk and M. Ross (eds), 17–62.
- (in press) 'Notes on the prehistory and subgrouping of Malayan', in J. Bowden and N.P. Himmelmann (eds) *Papers in Austronesian linguistics*, Canberra: Pacific Linguistics.
- Rousseau, J. (1988) *Central Borneo: a bibliography*, Sarawak Museum Journal Special Monograph No. 5, Kuching (Sarawak): The Sarawak Museum.
- Schuchardt, H. (1890) *Kreolische Studien IX. Über das Malaioportugiesische von Batavia und Tugu*, Wien: (Tempsky:) Kaiserliche Akademie der Wissenschaften.
- Shorto, H.L. (1975) 'Achinese and Mainland Austronesian', *Bulletin of the School of Oriental and African Studies*, 38:81–102.
- Skinner, G.W. (1963) 'The Chinese minority', in R.T. McVey (ed.), *Indonesia. Survey of World Cultures* (Southeast Asia Studies, Yale University), 97–117, New Haven (CT): Human Relations Area Files.
- Sneddon, J.N. (1993) 'The drift towards final open syllables in Sulawesi languages', *Oceanic Linguistics*, 32:1–44.
- Starosta, S., Pawley, A.K. and Reid, L. (1982) 'The evolution of focus in Austronesian', in A. Halim, L. Carrington and S.A. Wurm (eds), *Papers from TICAL* vol.2, 201–216, Canberra: Pacific Linguistics.
- Steinhauer, H. (1986) 'Austronesian geographical prospects', *Bijdragen tot de Taal-, Land- en Volkenkunde*, 142:296–313.
- Suthiwan, T. (1992) 'Kata pijaman Melayu dalam bahasa Thai', *Dewan Bahasa* 8/8:695–700.
- Taber, M. (1993) 'Towards a better understanding of the indigenous languages of Southwestern Maluku', *Oceanic Linguistics*, 32:389–441.
- Tadmor, U. (1995) *Language contact and systemic restructuring: the Malay dialect of Nonthaburi, Central Thailand*, PhD thesis, University of Hawaii.
- Thurgood, G. (1994) 'Tai-Kadai and Austronesian: the nature of the historical relationship', *Oceanic Linguistics*, 33:345–368.
- (1999) *From ancient Chamic to modern dialects, 2000 years of language contact and change*, Oceanic Linguistics Special Series, Honolulu: University of Hawaii Press.
- Tideman, J. (1936) *Hindoe-invoede in het noordelijk Batakland*, Uitgaven van het Bataksch Instituut nr. 23, Amsterdam.
- Tryon, D.T. *et al.* (eds) (1995) *Comparative Austronesian dictionary. An introduction to Austronesian studies* (5 vols), Trends in Linguistics Documentation 10 (W. Winter and R.A. Rhodes (eds)), Berlin: Mouton de Gruyter.
- Tryon, D.T. (1995) 'Introduction to the Austronesian comparative dictionary', in D.T. Tryon *et al.* (eds), 1–44.
- Tsang, Cheng-hua (2001) 'Recent discoveries of the Tapengkeng culture in Taiwan: implications for the problem of Austronesian origins', paper: Seminar Perspectives on the phylogeny of East Asian Languages, Périgueux (France), 29–31 August 2001.
- Tsuchida, Shigeru (1976) *Reconstruction of Proto-Tsouic phonology*, Study of languages and cultures of Asia and Africa monograph series No. 5, Institute for the study of Languages and Cultures of Asia and Africa, Tokyo.
- (1995) 'Alienable and inalienable distinction in Puyuma?' in P. J.-K. Li *et al.* (eds), *Austronesian studies relating to Taiwan*, 793–804, Taipei: Institute of History and Philology, Academia Sinica.
- Tsuchida, S. and Yamada, Y. (1991) 'Ogawa's Siraya/Makatao/Taivoan (Comparative Vocabulary)', in S. Tsuchida, Y. Yamada and T. Moriguchi (eds), *Linguistic materials of the Formosan sinicized populations I: Siraya and Basai*, 1–94, Tokyo: The University of Tokyo, Department of Linguistics.
- Urry, J. and Walsh, M. (1981) 'The lost "Macassar language" of Northern Australia', *Aboriginal History*, 5/2:90–108.
- Van den Berg, R. (1996) 'The demise of focus and the spread of conjugated verbs in Sulawesi', in H. Steinhauer (ed.), *Papers in Austronesian Linguistics* No.3, 89–114, Canberra: Pacific Linguistics.
- (in press) 'The place of Tukang Besi and the Muna-Buton languages', in Lynch, J. (ed.).
- Van der Vlis, C.J. (1842) *Formosaansche woorden-lijst, volgens een Utrechts Handschrift*, Verhandelingen van het Bataviaasch Genootschap 18:437–488.
- Van Engelenhoven, A. (1995) *A description of the Leti language (as spoken in Tutukei)*, PhD thesis, Leiden University.
- Van Ronkel, P.H.S. (1902) 'Het Tamil-element in het Maleisch', *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* (Koninklijk Bataviaasch Genootschap voor Kunsten en Wetenschappen, Batavia) 45:97–119.
- Vischer, M.P. (in press) *Precedence: processes of social differentiation in the Austronesian world*, Canberra: RSPAS, Dept. of Anthropology, Australian National University.
- Walker, A. and Zorc, R.D. (1981) 'Austronesian loanwords in Yolngu-Matha of Northeast Arnhem Land', *Aboriginal History*, 5/2:109–134.
- Williams-Van Klinken, C., Hajek, J. and Nordlinger, R. (2002) *Tetun Dili. A grammar of an East Timorese language*, Canberra: Pacific Linguistics.
- Wolff, J.U. (1973), 'Verbal inflection in Proto-Austronesian', in A.B. Gonzalez (ed.), *Parangal Kay Cecilio Lopez*, 71–91, Quezon City: Philippine Journal of Linguistics (Special monograph 4).
- (1976) 'Malay borrowings in Tagalog', in C.D. Cowan and O.W. Wolters (eds), *Southeast-Asian history and historiography: Essays presented to D.G.H. Hall*, 345–367, Ithaca: Cornell University Press.
- (1988) 'The PAN consonant system', in R. McGinn (ed.), *Studies in Austronesian linguistics*, Ohio University Monographs in International Studies, 125–147, Athens: Ohio University.
- (1995) 'The position of the Austronesian languages of Taiwan within the Austronesian group', in P.J.-K. Li *et al.* (eds), 521–583.
- (1996) 'The development of the passive verb with pronominal prefix in Western Austronesian languages', in B. Nothofer (ed.), *Reconstruction, classification, description – festschrift in honor of Isidore Dyen*, 15–40, Hamburg: Abera Verlag.
- (1999) 'The monosyllabic roots of Proto-Austronesian', in E. Zeitoun and P.J.-K. Li (eds), 139–194.
- (2002) 'Final words: the development of the focus system', in F. Wouk and M. Ross (eds), 437–449.
- Wouk, F. and Ross, M. (2002) *The history and typology of Western Austronesian voice systems*, Canberra: Pacific Linguistics.
- Wurm, S.A. and Hattori, S. (eds) (1981–3) *Language atlas of the Pacific area*, Canberra: Pacific Linguistics.
- Wurm, S.A. and Mühlhäusler, P.; Tryon, D.T. (eds) (1996) *Atlas of languages of intercultural communication in the Pacific, Asia and the Americas*, Berlin and New York: Mouton de Gruyter.
- Wurm, S.A. and Wilson, B. (1975) *English finderlist of reconstructions in Austronesian languages (post-Brandstetter)*, Canberra: Pacific Linguistics.
- Zeitoun, E. and Huang, L.M. (2000) 'Concerning ka-, an overlooked marker of verbal derivation in Formosan languages', *Oceanic Linguistics*, 39:391–414.
- Zeitoun, E. and Li, P.J.-K. (eds) (1999) *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, Symposium Series of the Institute of Linguistics (Preparatory Office), Taipei: Academia Sinica.

- Zobel, E. (2002) 'The position of Chamorro and Palauan in the Austronesian family tree: evidence from verb morphosyntax', in F. Wouk and M. Ross (eds), 405–434.
- (to appear) *A diachronic study of Austronesian verb morphosyntax*, PhD thesis, Universität Frankfurt.
- Zorc, R.D. (1986) 'The genetic relationships of Philippine languages', in P. Geraghty, L. Carrington and S.A. Wurm (eds), *Focal II: papers from the Fourth International Conference on Austronesian Linguistics*, Canberra: Pacific Linguistics.

CHAPTER TWO

LANGUAGE SHIFT AND ENDANGERMENT

Margaret Florey

1 INTRODUCTION

In recent years there has been a marked upsurge internationally in concern about the issue of language endangerment – by communities of speakers of endangered languages (ELs), the media, and among linguists. This increased attention has stemmed from a growing realization that the world's linguistic resources are rapidly shrinking. Estimates such as the widely quoted prediction that perhaps only 10% of the approximately 6000 languages in the world today could be classified as 'safe', while 10% are nearly extinct and 20% are moribund (Krauss 1992:6) highlight diminishing linguistic and cultural diversity and the loss of indigenous knowledge.

The various topics relating to language endangerment and language obsolescence have been the subject of much academic discourse through the past decade. Researchers have addressed issues from the perspectives of evaluating or describing programs aimed at reversing language shift or maintaining or renewing ELs (Fishman 1991, Hartman and Henderson 1994, Hinton 1994, Hinton and Hale 2001, Mugler and Lynch 1996), analyzing the issues facing a particular country or language family (Brenzinger 1992, Schmidt 1990, Robins and Uhlenbeck 1991), discussing community responses to language endangerment (Grenoble and Whaley 1998), offering suggested focus questions for the study of ELs (Dorian 1989), critiquing the role of specialists vis-à-vis language endangerment (Ostler 1998), producing tools for working with ELs (Thieberger 1995), placing language endangerment within the broader framework of the loss of biodiversity (Maffi 2001), and providing a survey accessible to non-specialists of the issues involved in language obsolescence (Crystal 2000).

Yet the concern and the initiatives which are emerging in other parts of the world have not been matched in the Austronesian region, which, despite its tremendous linguistic resources and the rapid pace at which minority languages are becoming endangered, remains remarkable for the lack of detailed information about many of the languages and their linguistic vitality. To date, there has been no comparable sense of urgency among linguists – or among speakers or community groups – to work towards the documentation and maintenance of endangered Austronesian languages. Indeed, Austronesian languages have received scanty attention in the volumes which purport to provide a regional overview of language endangerment. The *Atlas of the world's languages in danger of disappearing* (Wurm 1996) includes no maps or analysis of language endangerment for Austronesian languages – even for those languages for which this information has been reasonably well-documented, such as in Taiwan and Maluku (the Moluccas). Discussion of Austronesian languages in Robins and Uhlenbeck (1991) is limited to several pages focused largely on Indonesia. It is still too often the case that in many of the grammars

ROUTLEDGE LANGUAGE FAMILY SERIES

Each volume provides a detailed, reliable account of every member language, or representative languages of a particular family. Each account is a reliable source of data, arranged according to the natural system of classification: phonology, morphology, syntax, lexis, semantics, dialectology and socio-linguistics. Each volume is designed to be the essential source of reference for a particular linguistic community, as well as for linguists working on typology and syntax.

The Austronesian Languages of Asia

and Madagascar

*Edited by Alexander Adelaar &
Nikolaus P. Himmelmann*

The Bantu Languages

*Edited by Derek Nurse &
Gérard Philippson*

The Celtic Languages

Edited by Martin Ball & James Fife

The Dravidian Languages

Edited by Sanford B. Steever

The Germanic Languages

*Edited by Johan van der Auwera &
Ekkehard König*

The Indo-Aryan Languages

*Edited by George Cardona &
Dhanesh Jain*

The Indo-European Languages

*Edited by Paolo Ramat &
Anna Giacalone*

The Iranian Languages

Edited by Gernot Windfuhr

The KhoeSan Languages

Edited by Rainer Vossen

The Manchu-Tungusic Languages

Edited by Alexander Vovin

The Mongolic Languages

Edited by Juha Janhunen

The Oceanic Languages

*Edited by John Lynch, Malcolm Ross &
Terry Crowley*

The Romance Languages

*Edited by Martin Harris &
Nigel Vincent*

The Semitic Languages

Edited by Robert Hetzron

The Sino-Tibetan Languages

*Edited by Graham Thurgood &
Randy LaPolla*

The Slavonic Languages

*Edited by Bernard Comrie &
Greville B. Corbett*

The Turkic Languages

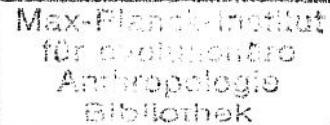
*Edited by Lars Johanson &
Eva Csato*

The Uralic Languages

Edited by Daniel Abondolo

THE AUSTRONESIAN LANGUAGES OF ASIA AND MADAGASCAR

*Edited by Alexander Adelaar and
Nikolaus P. Himmelmann*



PL 5026 AUS 2005

CONTENTS

First published 2005
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Simultaneously published in the USA and Canada
by Routledge
270 Madison Ave, New York, NY 10016

Routledge is an imprint of the Taylor & Francis Group

© 2005 Selection and editorial matter, Alexander Adelaar and
Nikolaus P. Himmelmann; individual chapters, the contributors

Cartography by Chandra Jayasuriya, The University of Melbourne

Typeset in Times New Roman by
Newgen Imaging Systems (P) Ltd, Chennai, India
Printed and bound in Great Britain by
MPG Books Ltd, Bodmin

All rights reserved. No part of this book may be reprinted or reproduced
or utilised in any form or by any electronic, mechanical, or other means,
now known or hereafter invented, including photocopying and recording,
or in any information storage or retrieval system, without permission in
writing from the publishers.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data

A catalog record for this book has been requested

ISBN 0-7007-1286-0

<i>List of illustrations</i>	x
<i>List of contributors</i>	xv
<i>Preface</i>	xvi
<i>List of abbreviations</i>	xix

1 The Austronesian languages of Asia and Madagascar: a historical perspective Alexander Adelaar	1
1 Introduction	1
2 Important reference works concerning Austronesian comparative historical linguistics	2
3 Writing	3
4 PAn and PMP linguistic history: some basic information	5
5 The internal classification of Austronesian languages	8
6 The Austronesian homeland and Austronesian migrations	26
7 Language contact	29
2 Language shift and endangerment Margaret Florey	43
1 Introduction	43
2 Factors in assessing language endangerment	44
3 Ethnolinguistic vitality of Austronesian languages	47
4 Reviewing research priorities	59
3 Colonial history and language policy in Insular Southeast Asia and Madagascar Hein Steinhauer	65
1 Introduction	65
2 Indonesia	66
3 Malaysia	69
4 Brunei Darussalam	72
5 East Timor	73
6 The Philippines	75
7 Madagascar	78
8 Language policy in other Southeast Asian countries	80
9 Concluding remarks	84
4 Ritual languages, special registers and speech decorum in Austronesian languages James J. Fox	87
1 Introduction	87