Once they had split up each group forgot the past customs they had enjoyed together and developed different languages because some had short tongues and others long tongues. Each group found a new name for itself.
(From a story related by a Penan headman, Sarawak: Arnold 1958)

The modern traveler who has the good fortune to wander at will through the Indo-Malaysian region will quickly observe that there are many traditional varieties of culture and subsistence economy. This remains true even when one considers the tremendous impacts of modern urbanization and industrialization. Furthermore, there are great variations in language, although these are often not noticed by outsiders owing to the increasing strengths of the national languages, Bahasa Melayu and Bahasa Indonesia (these are actually the same language, originally Malay, with small dialect differences). Linguists in particular have developed precise techniques for drawing inferences about the histories of language families. I consider that any general statements concerning the prehistory of the Indo-Malaysian region within the past 5,000 years must take the linguistic evidence seriously into account. This is especially true for any discussion of recent human population expansion. One cannot work purely from the archaeological data alone.
I. LANGUAGE FAMILIES IN SOUTHEAST ASIA AND THE WESTERN PACIFIC

Almost all the peoples of Indo-Malaysia speak languages termed Austronesian (Wurm and Hattori 1983; Pawley and Ross 1993; Bellwood et al. 1995). There are two small exceptions: the Orang Asli of interior Peninsular Malaysia, who speak languages (called Aslian) in the Austroasiatic family; and some peoples of eastern Indonesia who speak Papuan languages, a group of diverse families centered on New Guinea. The Aslian languages are considered in some detail in Chapter 8, while the Papuan languages will be considered at the end of this chapter. When looking at Austronesian linguistic prehistory and the relationships of Austronesian to outside families, research cannot be restricted to Indonesia and Malaysia alone. I will paint a broad picture to begin, and then focus in detail on more localized questions.

Estimates of the number of Austronesian languages now range up to about 1,200, thus making it perhaps the largest in the world in terms of language number and certainly the largest in terms of precolonial extent: over half way around the world, from Madagascar to Easter Island (Fig. 4.1). Over half of these languages are spoken in Oceania, from New Guinea eastward. Present Austronesian speakers are roughly distributed as follows: Indonesia 190 million, Malaysia 18 million, Philippines 66 million, Taiwan 330,000, interior Vietnam 500,000, Madagascar 12 million, and Oceania about 2 million (SBS 1992; Tryon 1995a). Most of the dispersed languages in the vast region of Oceania clearly have very small speech communities.

The Austronesian languages have a geographical distribution that is relatively unbroken except for the outliers in Madagascar and southern Vietnam, the latter having been isolated by recent Vietnamese expansion. There are four language groupings that abut directly on the Austronesian sphere of distribution (Ruhlen 1987) (Fig. 4.2). The first is the Papuan complex, which has about 750 languages in many separate phyla (to use the terminology most favored by linguists working in this region), in New Guinea, western island Melanesia, and a few islands in eastern Indonesia (Wurm 1982; Foley 1986). The second is the Austroasiatic family of Mainland Southeast Asia, a scattered group of about 150 languages (Diffloth 1979; Parkin 1991) that includes Aslian of Peninsular Malaysia, Mon-Khmer, Vietnamese, Nicobarese, some languages in Assam, and the Mundaic languages of northeastern peninsular India. The third comprises the Tai (or Daic) family (Thai, Lao, and related languages) of the central and northern mainland, extending up into southern China (Lebar et al. 1964; Ruhlen 1987). The fourth is the large Sino-Tibetan family (Matisoff 1991), although the main contact between this and Austronesian
occurs in Taiwan and is of historically recent origin (mainly post-seventeenth century).

When considering geographical distribution alone, the extent of the Austronesian languages is indeed very impressive. It is my view that the expansion of this language family has involved an actual expansion of Austronesian-speaking founder communities through this vast area (Bellwood 1991, 1995a). This may seem self-evident, especially for Oceania, although the possibilities of adoption of Austronesian languages by members of unrelated linguistic groups cannot be entirely disregarded, especially in western Melanesia and amongst the Negritos of the Philippines (Ross 1994; Reid 1994a). If such an expansion has occurred, then it would appear to have reduced the area of prior distribution of the Papuan and Austroasiatic languages, with the exception of Vietnamese. The questions for Austronesian expansion are: When did it occur? How did it occur? and Why did it occur?
II. SOME LINGUISTIC CONCEPTS

According to Swadesh (1964:575), a linguist who has made great contributions to prehistory:

There are three main ways in which linguistics can illuminate prehistory: (a) by establishing facts concerning the common origin and subsequent divergence of languages, implying the earlier unity and subsequent separations of peoples; (b) by discovering diffused features (of phonetics, structure and vocabulary) among languages which bear evidence of prehistoric culture contact; and (c) by reconstructing the vocabulary of old stages of languages so as to bring out suggestions of the physical environment and content of prehistoric cultures.

A number of terms linguists use in their deliberations now require a brief introduction (for a more detailed explanation, see Crowley 1992). Dialects of a language share a common basic vocabulary—usually over 80 percent—sufficient to remain mutually intelligible. If they diverge enough to become no longer mutually intelligible, they become separate languages. This separation is of course gradual as intelligibility does not come or go with one percentage point, and the development of two separate daughter languages from one parent can take a millennium or more, especially if contacts continue (the degree of continuing contact versus separation is essential in this regard).²

Languages considered to be related are grouped into families (or phyla in the terminology favored by Papuan linguists), such as Austronesian or Indo-European. All languages are probably related in the final resort, since the roots of language must go back with increasing convergence through all levels of anatomically modern human evolution. But, for comparative purposes, relationship has to be observable in phonology, structure, and vocabulary. When languages have been evolving apart for something over 10,000 years, it seems that such observations can no longer be made with confidence.

Languages in a family can be divided into subgroups with ordered implications for time depth; closely related languages (closely related, that is, in terms of shared innovations) have more recent common ancestors than those more distantly related. Hence, subgroups are hierarchical in terms of geographical space and time depth. A subgroup can be defined as "any group of languages which have passed through a period of common development exclusive of the other languages of the same family and during which period some linguistic change has occurred" (Grace 1959). Uniquely innovated shared cognates of restricted geographical occurrence help to define subgroups, while those that occur within and beyond a particular subgroup are defined as "shared inheritances" (not innovations), and as such refer to a deeper level of subgrouping. A
cognate is a word deemed through correspondences in sound and meaning to have been inherited by two or more languages from a common ancestor, rather than borrowed from an outside language.

Members of a subgroup also share a common proto-language, an entity reconstructed by identifying cognates (commonly inherited forms) found in daughter languages; the more widespread the cognate, the farther back the proto-language for which it can be reconstructed. Like subgroups, proto-languages are also hierarchical in space and time, each one corresponding to the base of a subgroup branch in the family tree. Minor proto-languages on the "outer branches" of Austronesian can be shown to have been single languages or dialect chains in many cases, but when moving toward the main trunk and back in time the picture becomes more diffuse and complex. An entity such as Proto-Austronesian (Table 4.1), which existed perhaps 5,000 years ago, may or may not have been a single language as opposed to a chain of related dialects—it is possible that we will never know for certain. However, we can rule out the possibility that it comprised a group of completely unrelated languages; whatever its precise nature, Proto-Austronesian must have been a relatively homogeneous linguistic entity of somewhat restricted geographical extent.

I must also emphasize that the entity reconstructed as Proto-Austronesian is in no real sense a "beginning" for Austronesian; it was not created de novo. What the concept refers to is that point at which the language or chain of languages ancestral to all modern Austronesian languages underwent an initial divergence into two subgroups whose descendant languages have survived essentially apart from the present. Proto-Austronesian is preceded by a phase I will call Initial Austronesian, which in turn goes back to the point at which the ancestral language for the whole Austronesian family split off from its contemporary sister languages. The latter have either developed along different lines to become other modern language families (the Austroasiatic, Tai, and perhaps even Sino-Tibetan families may have this kind of relationship with Austronesian; see below), or some might have become extinct. If they were not written we may never know about this second possibility, but language termination is a common event in linguistic history.

The ancestry of Austronesian thus goes well back beyond the proto-stage, and the family may be visualized as a very large and multibranched limb on the total linguistic tree of mankind. Where the limb departs from another bigger limb (which will in turn join back through others to the ultimate trunk), it is convenient to place the label "Initial Austronesian" (Fig. 4.3). We can only get back beyond the start of this limb by comparing Austronesian languages with those in other families, and many linguists have claimed significant results from such comparisons. Some of these results will be discussed later.
<table>
<thead>
<tr>
<th></th>
<th>PAN</th>
<th>Rukai</th>
<th>Tagalog</th>
<th>Javanese</th>
<th>Fijian</th>
<th>Samoan</th>
<th>Rapanui</th>
</tr>
</thead>
<tbody>
<tr>
<td>two</td>
<td>*DuSa</td>
<td>dosa</td>
<td>dalawa</td>
<td>lo-ro</td>
<td>rua</td>
<td>lua</td>
<td>rua</td>
</tr>
<tr>
<td>four</td>
<td>*Sepat</td>
<td>sepate</td>
<td>āpat</td>
<td>pat</td>
<td>vā</td>
<td>fā</td>
<td>hā</td>
</tr>
<tr>
<td>five</td>
<td>*limaH</td>
<td>lima</td>
<td>lima</td>
<td>limo</td>
<td>lima</td>
<td>lima</td>
<td>rima</td>
</tr>
<tr>
<td>six</td>
<td>*enem</td>
<td>eneme</td>
<td>ānim</td>
<td>nem</td>
<td>ono</td>
<td>ono</td>
<td>ono</td>
</tr>
<tr>
<td>bird</td>
<td>*manuk</td>
<td>(aāaāme)</td>
<td>manok</td>
<td>manu?</td>
<td>manumanu</td>
<td>manu</td>
<td></td>
</tr>
<tr>
<td>head louse</td>
<td>*kuCuH</td>
<td>koco</td>
<td>kūto</td>
<td>kutu</td>
<td>kutu</td>
<td>?utu</td>
<td>kutu</td>
</tr>
<tr>
<td>eye</td>
<td>maCa</td>
<td>maca</td>
<td>mata</td>
<td>moto</td>
<td>mata</td>
<td>mata</td>
<td>mata</td>
</tr>
<tr>
<td>ear</td>
<td>*Calɪja</td>
<td>caliqa</td>
<td>tēŋa</td>
<td>(kupiŋ)</td>
<td>daliqa</td>
<td>taliqa</td>
<td>tariqa</td>
</tr>
<tr>
<td>liver</td>
<td>*qaCey</td>
<td>aŋay</td>
<td>atay</td>
<td>ati</td>
<td>yate</td>
<td>ate</td>
<td>?ate</td>
</tr>
<tr>
<td>road</td>
<td>*Zalan</td>
<td>ka-dalan-ane</td>
<td>daan</td>
<td>dalan</td>
<td>sala</td>
<td>ala</td>
<td>ara</td>
</tr>
<tr>
<td>pandanus</td>
<td>*paŋuDaN</td>
<td>paŋodale</td>
<td>pandan</td>
<td>pandan</td>
<td>vadra</td>
<td>fala</td>
<td>—</td>
</tr>
<tr>
<td>coconut</td>
<td>*niuR</td>
<td>(abare)</td>
<td>niyog</td>
<td>nior</td>
<td>niu</td>
<td>niu</td>
<td>(haŋari)</td>
</tr>
<tr>
<td>sugarcane</td>
<td>*tebuS</td>
<td>cubusu</td>
<td>tubo</td>
<td>tebu</td>
<td>dovu</td>
<td>tolo</td>
<td>toa</td>
</tr>
<tr>
<td>rain</td>
<td>*quZaN</td>
<td>odale</td>
<td>ulan</td>
<td>udan</td>
<td>uca</td>
<td>ua</td>
<td>?ua</td>
</tr>
<tr>
<td>sky</td>
<td>*laŋiC</td>
<td>(sobelebeleŋ)</td>
<td>laŋit</td>
<td>laŋit</td>
<td>laŋi</td>
<td>laŋi</td>
<td>raŋi</td>
</tr>
<tr>
<td>stone</td>
<td>*batu</td>
<td>(leneg)</td>
<td>bato</td>
<td>watu</td>
<td>watu</td>
<td>(maŋa)</td>
<td>(maŋea)</td>
</tr>
<tr>
<td>cooking pot</td>
<td>*kuDen</td>
<td>(paloŋ)</td>
<td>(baŋa?)</td>
<td>(kuwali)</td>
<td>kuro</td>
<td>?ulo</td>
<td>(pani)</td>
</tr>
<tr>
<td>canoe</td>
<td>*awaŋ</td>
<td>avaŋe</td>
<td>baŋka</td>
<td>(prau)</td>
<td>waga</td>
<td>vaŋa</td>
<td>vaka</td>
</tr>
<tr>
<td>eat</td>
<td>*kaŋen</td>
<td>kane</td>
<td>kāŋin</td>
<td>ma-ŋan</td>
<td>kan-ia</td>
<td>?ai</td>
<td>kai</td>
</tr>
</tbody>
</table>

*Table 4.1* Some widespread Austronesian cognates (excluding forms in brackets). PAN = Proto-Austronesian; Rukai is a Taiwan language; Tagalog is the national language of the Philippines; Rapanui is Easter Island. Courtesy: Malcolm Ross.
Fig. 4.3 A “family tree” for the Austronesian languages, derived from the subgrouping of Robert Blust. AN = Austronesian; MP = Malayo-Polynesian; WMP = Western Malayo-Polynesian; CEMP = Central-Eastern Malayo-Polynesian; OC = Oceanic.
Lexical cognates are also utilized in the subdiscipline of lexicostatistics, which subgroups languages hierarchically in terms of percentages of cognates between all pairs of languages within a group. I will return below to the assumptions of lexicostatistics and its chronological offshoot (giottochronology) when I come to consider the crucial question of the rate of linguistic change. I should, however, make it clear that languages are not subgrouped purely by comparing words; indeed, the technique of lexicostatistics is regarded with suspicion as a classificatory device by many linguists. Our basic knowledge of Austronesian linguistic prehistory comes from much deeper comparisons in the complex fields of grammar (for example, word order in sentences, occurrences of prefixes and suffixes, pronoun forms, verb structures) and phonology (the study of the sound changes through which different languages have progressed). Lexicostatistics alone can in fact lead to a quite erroneous view of Austronesian linguistic prehistory. This is because great lexical diversity alone does not automatically correspond with great time-depth, although there is certainly much truth in the view that great overall linguistic diversity does (Dyen 1975).

III. THE MAJOR SUBGROUPS OF AUSTRONESIAN

Comparative linguists are now in general agreement about the basic shape of the Austronesian family tree. Most today use the classification developed by the linguist Robert Blust (1977, 1978, 1982, 1993, 1995a). This classification is based on the occurrences of shared innovations, especially in phonology and in pronoun forms, at the lower levels of the family tree (for details see Ross 1995a; Tryon 1999b). It divides the Austronesian family into at least two major subgroups, of which one, Malayo-Polynesian, includes all Austronesian languages not located in Taiwan (see Fig. 4.3). The Formosan languages themselves include at least one major subgroup, and perhaps between three and six according to recent research (Reid 1982; Li 1985; Blust 1995b). Arguments against the view that the Formosan languages represent one or more primary subgroups have been put forward by Dyen (1995) from a lexicostatistical viewpoint, and by Wolff (1995) from a counterperspective on pronoun forms, but in my view Blust has presented convincing rebuttal (see also Pawley and Ross 1993; Ross 1999b, for support on the Blust tree from verb morphology).

Since the first edition of this book was published, the Blust classification has become almost universally adopted by those linguists who use the comparative method of reconstruction at all levels of language—rather than simply lexicostatistics—and I have decided to delete some of the historical discussion included in that edition. The Blust classification also fits well with the evidence of Island Southeast Asian archaeology (see Chapter 7), although this independent evidence cannot be used directly in its support against other possible sub-
grouping schemes. But certainly, at the present time, there is no other convinc-
ing competitor.

The major subgroups of Austronesian are therefore as follows:

1.1. Formosan (possibly three or more primary subgroups)
1.2. Malayo-Polynesian (all Extra-Formosan languages according to Blust, although Reid [1982] excludes some Northern Philippine languages from Malayo-Polynesian and places them in a separate subgroup intermediate between 1.1 and 1.2; see also Starosta [1995] for an even more complex tree at this level)
2.1. Western Malayo-Polynesian (Philippines, Vietnam, Madagascar, Peninsu-
lar Malaysia, Sumatra, Java, Borneo, Sulawesi, Bali, Lombok, western Sumbawa, and two languages—Palauan and Chamorro—of western Micronesia)
2.2. Central-Eastern Malayo-Polynesian (3.1 plus 3.2)
3.1. Central Malayo-Polynesian (Lesser Sundas from eastern Sumbawa east-
ward, Moluccas except Halmahera)
3.2. Eastern Malayo-Polynesian (South Halmahera and all the Pacific Island
Austronesian languages; Melanesia, Micronesia, and Polynesia: 4.1 plus 4.2)
4.1. South Halmahera–West New Guinea
4.2. Oceanic (all Eastern Malayo-Polynesian languages except 4.1)

By translating this family tree into an account of the expansion of the Austro-
nesian speakers, the following founder moves can be listed in chronological
order (Blust 1984–1985, 1996):

1. Initial Austronesian and Proto-Austronesian should be located in Taiwan.
The founders of the Malayo-Polynesian subgroup then moved into rela-
tive separation (maybe even absolute separation) in the Philippines, via
Luzon, leaving the other speakers of Formosan languages behind.
2. Founders moved south through the Philippines, then west into Borneo and Sulawesi, and later toward Java, Sumatra, Peninsular Malaysia, and Vietnam. These regions contain the Western Malayo-Polynesian (WMP) languages, which do not comprise a single subgroup (the internal structure of the totality of WMP has not yet been completely defined; some discussion on this follows later).
3. Other founders moved east and south into the Moluccas and Lesser Sundas, giving rise to Central-Eastern Malayo-Polynesian (Blust 1993).
4. Further founder movements gave rise to a separation between Central Malayo-Polynesian and Eastern Malayo-Polynesian, the latter spreading rapidly onward via Halmahera and the Bismarck Archipelago into Oceania.
It should be noted from the above that the Austronesian languages did not spread into Indonesia via the Malay Peninsula, even though this route was favored at one time (e.g., by Heine Geldern 1932).

**A. The Location of Proto-Austronesian**

I have already made it clear that the island of Taiwan is central to the question of the location of Proto-Austronesian. This conclusion has been reached with increasing frequency in recent years, for instance by Dahl (1973), Shutler and Marck (1975), Foley (1980), Harvey (1982), Reid (1984-1985), and the more recent authorities listed above. Many earlier authors, using the results of traditional comparative linguistics, had concluded that the location of Proto-Austronesian would have to lie somewhere in or close to the western region of the Austronesian distribution (excluding Madagascar). For example, Kern (1889/1917) favored the Vietnamese coastal region, Haudricourt (1954) favored the coastal region between Hainan and Taiwan, and Benedict (1975) has for the past forty years been strongly in favor of linking Austronesian through Taiwan with the Thai and Kadai languages of southern China. Blust (1982) has shown that words for placental mammal species (pangolin, bovids, monkeys) reconstructed for the Proto-Austronesian vocabulary make a location west of Huxley’s Line—in Taiwan or Sundaland—a virtual certainty, although it should be noted this view does depend upon acceptance of Blust’s family tree. Indeed, all proto-language vocabularies can only be reconstructed with certainty if the precise family tree of the language family concerned is also known with certainty, and this is a field where certainty may always be a little elusive.

Nevertheless, taking all these views into account and noting recent trends, there now seems little reason to doubt Taiwan as by far the most likely location for Proto-Austronesian. The most important arguments against Taiwan in recent years have been those by Dyen, whose lexicostatistical classification of the Austronesian languages (Dyen 1962, 1965a) showed that lexical (but not phonological) diversity was higher in western Melanesia than in any other Austronesian region (Fig. 4.4). Dyen’s theories on unvarying overall rates of lexical change through time led him to suggest western Melanesia (particularly the Bismarck Archipelago) as the location of Proto-Austronesian—a claim supported by Murdock (1964), but seen today by most authors (including myself) to be impossible, for reasons not connected with linguistics alone (Bellwood 1978: 131–132). I will be returning to Dyen and the question of Melanesia later, but it should be noted that he, as a linguist of broad and respected knowledge, has long been ambivalent about the Formosan languages and has at times admitted their qualifications to represent a primary subgroup of Austronesian (e.g., Dyen 1965b).
Fig. 4.4 Isidore Dyen's family tree for the Austronesian languages.
Table 4.2. Some terms for material culture reconstructed for early Austronesian societies. After Blust 1976.

<table>
<thead>
<tr>
<th>Class of material culture</th>
<th>Reconstruction code</th>
<th>Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>House and contents</td>
<td>NWE</td>
<td>house/family dwelling</td>
</tr>
<tr>
<td></td>
<td>WE</td>
<td>ridgepole, rafter, thatch, house-post, storage rack above hearth, notched log ladder, hearth, public building</td>
</tr>
<tr>
<td>Tools, utensils, weapons</td>
<td>NWE</td>
<td>bow, shoot an arrow, rope/cord</td>
</tr>
<tr>
<td></td>
<td>WE</td>
<td>putty/caulking substance, comb, conch shell trumpet, cooking pot, nail, pillow/wooden headrest, digging stick, bamboo trail- or pitfall-spikes, torch, hew/plane</td>
</tr>
<tr>
<td>Arts and crafts</td>
<td>NWE</td>
<td>needle</td>
</tr>
<tr>
<td></td>
<td>NW</td>
<td>loom (early Western Malayo-Polynesian only?)</td>
</tr>
<tr>
<td></td>
<td>WE</td>
<td>plait/weave, draw, whet/sharpen, sew</td>
</tr>
<tr>
<td>Adornment</td>
<td>NW</td>
<td>tattoo</td>
</tr>
<tr>
<td>Refreshment</td>
<td>NW</td>
<td>drunk (adjective)</td>
</tr>
<tr>
<td></td>
<td>WE</td>
<td>lime for betel quid, betel nut</td>
</tr>
<tr>
<td>Hunting and fishing</td>
<td>NW</td>
<td>hunt, go hunting</td>
</tr>
<tr>
<td></td>
<td>WE</td>
<td>bait, bamboo basket trap for fish, kind of fishnet, fishhook, fish drive, dragnet, derris root fish poison, bird lime, snare</td>
</tr>
<tr>
<td>The canoe</td>
<td>NW</td>
<td>canoe/boat</td>
</tr>
<tr>
<td></td>
<td>WE</td>
<td>canoe, canoe paddle, outrigger, rollers for beaching a canoe, sail, canoe bailer, to paddle, rudder/steer, raft, cross-seat in a boat, punting pole.</td>
</tr>
<tr>
<td>Domesticated animals</td>
<td>NWE</td>
<td>domesticated pig, dog</td>
</tr>
<tr>
<td></td>
<td>WE</td>
<td>cock/rooster</td>
</tr>
</tbody>
</table>

continued
<table>
<thead>
<tr>
<th>Class of material culture</th>
<th>Reconstruction code</th>
<th>Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Garden and field</td>
<td>NWE</td>
<td>rice (E cognates not beyond west New Guinea), pestle, garden/cultivated field, sugarcane</td>
</tr>
<tr>
<td></td>
<td>NW</td>
<td>husked rice, mortar, cooked rice, winnow, rice straw, millet</td>
</tr>
<tr>
<td></td>
<td>WE</td>
<td>to weed, Alocasia sp. (an aroid), breadfruit, ginger, citrus fruit, banana, yam, sago, taro, fallow land, to plant, melon</td>
</tr>
<tr>
<td>Food preparation</td>
<td>NWE</td>
<td>to smoke meat or fish</td>
</tr>
<tr>
<td></td>
<td>NW</td>
<td>salt</td>
</tr>
</tbody>
</table>

N (north) = Formosan; W (west) = Western Malayo-Polynesian; E (east) = Eastern Malayo-Polynesian.
In coming down strongly on the side of Taiwan as the location of Proto-Austronesian, I should make it clear that I do not regard this small island as the absolute “homeland” of the Austronesians. It is simply the place where Initial Austronesian was established and where the first split within reconstructed Proto-Austronesian occurred. The true homeland of the ancestral Austronesians (before the Initial stage) was without any reasonable doubt on the mainland of southern China.

**B. The Material Culture of the Proto-Austronesians**

Words and meanings can only be reconstructed for Proto-Austronesian if cognates are found in the languages of two or more of the primary subgroups (Formosan and Malayo-Polynesian) and if borrowing can be positively excluded. The catch here, of course, is that the shape of the family tree must be known with certainty in order to reconstruct a proto-language with total conviction, and as I have indicated above there are still disagreements about this, although I have chosen the tree suggested by Blust as the most convincing one currently available. Nevertheless, despite disagreements, it is generally accepted by all linguists that words having cognate forms in at least one language in the Formosan, Western Malayo-Polynesian, and Eastern Malayo-Polynesian major divisions are almost certain candidates for Proto-Austronesian (see Table 4.1).

A moment’s thought will suggest that words reconstructed from Western and Eastern Malayo-Polynesian alone without Formosan cognates cannot automatically be considered as Proto-Austronesian; they can only truly be regarded as Proto-Malayo-Polynesian. Since Proto-Austronesian and Proto-Malayo-Polynesian may be separated by half a millennium in time and a subtropical versus a tropical latitude (i.e., Taiwan versus the Philippines), I believe this distinction to be quite significant. Words for truly tropical plants such as sago and bread-fruit can only be reconstructed for Proto-Malayo-Polynesian, exactly as expected. Words for colder climate species such as rice, millet, and sugarcane go back to Proto-Austronesian, and so these clearly were grown in Taiwan. The fact that there is quite an expanse of latitude here across an important zone of change in world climate clearly makes interpretation a little complex. In Table 4.2, items marked NWE or NW are potentially Proto-Austronesian (NE never occurs), and items marked WE are potentially Proto-Malayo-Polynesian and only possibly Proto-Austronesian. To simplify matters I have not added the Proto-Austronesian reconstructions themselves, and I will leave reconstructions of aspects of society for discussion in the next chapter.

It will be noted from Table 4.2 that most reconstructions are defined as WE, and few have cognates in Formosan. This no doubt reflects the limited information available on some of the Formosan languages (most Formosan aboriginal
populations now survive inland and have been heavily acculturated to the
dominant Chinese population), but another reason may be that Austronesian
cultures underwent marked change as they moved southwards into tropical
regions, as I will demonstrate later. This becomes rather obvious when one
examines the garden and field section, and also canoe terminology (on which
see Pawley and Pawley 1994).

I will be making further observations on this list in due course, but it should
be noted that the domesticated animals do not include any herbivores (cattle,
sheep, or goats). Another point, made by Pawley (1981) and by Blust (1976),
is that sound correspondences suggest strongly that material culture traditions
(potting, agriculture, fishing, and so forth) were continuous—they were never
lost by any widespread groups of Austronesians and later regained through rein-
novation or external borrowing. The main point to be noted, however, is that
Proto-Austronesians clearly were agriculturalists and had domestic pigs and dogs.
In addition, Proto-Malayo-Polynesians certainly made pottery and sailed between
islands in canoes (see also Zorc 1994). In archaeological terms, these were Neo-
lithic communities, albeit still fishers, gatherers, and hunters wherever these
activities were profitable.

C. The Antecedents of Proto-Austronesian

The cultural relationships of the Initial Austronesians who crossed from main-
land China to Taiwan are universally agreed to be with the mainland, rather
than with the island areas to the south, which at this early date were still inhab-
ited by hunters and gatherers who presumably spoke non-Austronesian lan-
guages now extinct. Only the Philippine Negritos survive as a distinct biological
population from this early phase in the islands directly south of Taiwan, and
they have now universally adopted Austronesian languages related to those of
their Filipino neighbors (see Reid 1994a, 1994c for a possible non-Austronesian
substratum incorporated in Philippine Negrito languages).

One proponent of Austronesian links with mainland languages has been
Paul Benedict, who first suggested in 1942 that Austronesian, Thai, and a small
group of isolated languages in Hainan and southern China (which he called
Kadai) could be classified into one large macrofamily. Since 1942, Benedict has
added the scattered Hmong-Mien (formerly called Miao-Yao) languages of south-
ern China and northern Mainland Southeast Asia to the Thai-Kadai-Austrones-
ian family and has coined the term Austro-Thai (now better spelled Austro-Tai)
to refer to the whole group (Benedict 1975). He has also reconstructed a large
Proto-Austro-Tai vocabulary, which includes terms for field, wet field (for rice or
taro), garden, plow, rice, sugarcane, the betel nut complex, cattle, water buffalo,
bow and arrow, axe, and canoe. Some of these items are also found in Proto-
Austronesian, while others, such as cattle and water buffalo, are not, although this may simply mean that these animals were not taken off the mainland during the early period of Austronesian expansion.

If Benedict is right, then these reconstructions apply to Neolithic societies in southern China, whose archaeological remains will be described later. They have always seemed to me to be quite reasonable from a prehistorian's perspective, and Benedict's general views have been supported in part by other linguists such as Dahl (1973) and Reid (1984–1985). However, since the first edition of this book was published, in which I accepted Benedict's views without provisos, the concept of Austro-Tai as a macrofamily of genetically related languages (i.e., languages sharing a common origin) has come under attack. In the past few years there have been several detailed reviews of macrofamily hypotheses involving Austronesian, and one that is now more favored than Austro-Tai is the "Austric" hypothesis, which suggests that Austronesian and Austroasiatic are genetically related at a deep level (Blust 1996; Diffloth 1994; Reid 1994b). This hypothesis was first suggested by Schmidt in 1906 and has had rather a checkered career since that time, but Blust's support is currently quite strong, and Austric has recently received archaeological support from Charles Higham (1996). For Austro-Tai, one current hypothesis (Thurgood 1994) is that the resemblances relate to early borrowing rather than true genetic relationship. Other suggestions exist to the effect that Austronesian may be related to Chinese (Sagart 1994), and even Japanese receives periodic attention as having an-cient Austronesian affiliations, although the evidence here is not widely accepted (Vovin 1994).

From the viewpoint of prehistory, these macrofamily hypotheses are quite important (Bellwood 1994). Whether Austronesian relates to Tai or Austroasiatic and whether the relationship is one of common origin or ancient borrowing, the conclusions are almost the same in terms of historical significance. Ancestral Austronesian languages, prior to the colonization of Taiwan and the period of Proto-Austronesian, were part of a geographical network of languages on the mainland of southern China. These languages were still perhaps undifferentiated into the clear and separate ancestors of existing families, and they formed a network that also included the seeds of early Austroasiatic and early Tai languages, and perhaps also early Sino-Tibetan languages—especially ancestral Chinese according to Sagart, with apparent support from Egerod (1991). The ultimate roots of Austronesian therefore lie in the Neolithic cultures of southern China, a topic to be explored further in Chapter 7. The earliest Austronesians grew rice and probably expanded into the Southeast Asian islands through processes of agricultural demographic growth, as well as through a culturally motivated desire to find new lands. A few archaeologists still refuse to accept this view of a Chinese regional origin for Austronesian, overlooking the details of
the linguistic evidence (Meacham 1984–1985), but for many years now I have been unable to take such objections very seriously (Bellwood 1984–1985, 1991). Austronesian prehistory is now becoming well founded from a multidisciplinary perspective.

IV. DATING THE AUSTRONESIAN FAMILY TREE

Before commencing a closer look at linguistic prehistory within Indonesia and Malaysia, I want to digress into the important question of linguistic dating. How can we date reconstructed entities such as Proto-Austronesian? Dating is one of the great scaffolds of prehistory and one that linguists cannot erect easily by themselves. In theory, if the packages of basic vocabulary items (i.e., words for noncultural terms—such as *man*, *woman*, and *sky*—usually compared in lists of 100 or 200 meanings) used for lexicostatistical classifications can be shown to have changed at a known and constant rate through time, then proto-languages should be datable by a mathematical comparison of the shared cognate percentages between selected pairs of their daughter languages. Thus, in theory, all the major splits in the Austronesian family tree can be dated—a feat of no small importance for prehistory—assuming of course that one can know the rate of vocabulary change over a fixed period of time, such as a millennium. This is the basis of the technique of glottochronology, to be discussed further below.

But do such constant rates really exist? Let us start with some simple assumptions about how dialects start to diverge amongst populations who use only verbal communication methods (i.e., without literacy, mass communication media, and so forth). For instance, it may be assumed that people of one ethnic group who live contiguously and interact freely with each other throughout the whole group will maintain a single language over time, regardless of changes that occur within the language itself as a result of the passage of time or borrowing. If a few families move off to a new area and maintain only diminished contact with the main group, they will slowly develop a different dialect, but as long as mutual intelligibility continues to be required, the new group will tend not to drift across the boundary into a separate language (as defined by lack of mutual intelligibility)—unless they become effectively cut off. In a sense, this process of becoming "cut off" may be regarded as optative and as reflecting social structure; in atomizing tribal polities, one may expect it to be more frequent than in large, integrated civilizations.

It may also be assumed that the rate of divergence between two communities speaking an initial common language will be somewhat faster if they are totally separated than if they maintain contact, and much faster if the speakers of one or both languages are in frequent bilingual contact with speakers of other unre-
lated languages. This introduces an important factor: Linguistic differentiation entirely from within a network of closely related dialects, as might occur on small isolated islands with no external contact or bilingual skills in totally "foreign" languages, will inevitably proceed more slowly than in cases where such isolation is lacking. Such slow differentiation seems to have been the case on large landmasses settled by essentially one language community, such as Madagascar and New Zealand (Pawley and Ross 1993). As for bilingualism, it does not homogenize language or cause individual languages to spread great distances, but it does lubricate diversification; more on this below in the Melanesian context.

In situations where communication over a large region is slowed by distance or difficult terrain but not broken, a sufficiency of time will often produce a dialect chain. In such situations, dialects at distant ends of the chain can often be mutually unintelligible (and thus actually different languages), but the differences will only be very gradual as one moves along the chain, unless it is broken up by expansion of groups within it or by the moving in of unrelated language groups. The existence of such chains does not stop linguistic change, but it may well act to slow it down because linguistic innovations will tend to be widely shared and overlapping in distribution—not localized and concentrated in one community. Many tribal regions of Southeast Asia and Oceania exhibit chains of this kind.

We may conclude from this discussion that the rate of linguistic change following the expansion or splitting of a language community will depend a great deal on rates of subsequent communication, as well as on individual sociolinguistic situations and the presence of other linguistically unrelated populations in the vicinity. Such variables, undocumented for prehistoric communities, can cause problems with chronological calculations. Lexicostatistical studies per se, such as the major Austronesian study by Dyen (1965a) already referred to, do not concern themselves directly with time, but only with classification and subgrouping. However, the time factor is central to an offshoot of lexicostatistics called glottochronology (Gudschinsky 1964). Analyses of languages (mostly Indo-European and west Asian) with long written histories have led to suggestions that basic vocabularies will change at constant rates: 19.5 percent per millennium for a standard 200-word list and 14 percent for a 100-word list. Hence simple glottochronological formulae can, in theory, be derived for calculating how long ago the common ancestor of two or more related languages was in existence.

Two basic assumptions of glottochronology are that the basic vocabularies of all languages change at the same rate and that basic vocabularies in turn are more stable than "cultural" vocabularies and more resistant to borrowing. Certain basic vocabulary items (such as the words for two, five, eye, and louse in the
Austronesian languages) can be shown to be more resistant to change than others (see Table 4.1), but taken as a whole the rate of overall basic vocabulary change is stated to be constant. The idea of such a constant rate, quite apart from the stability of individual vocabulary items, has of course long been challenged (e.g., Teeter 1963; Bergsland and Vogt 1962). Some languages—such as Icelandic—can be shown to have been extremely conservative, while others—as in Melanesia—can be shown to have changed their vocabularies quite rapidly. Although there exists an insufficient written record for Austronesian languages from which variant rates can be demonstrated directly, it is possible to throw light on this matter by using another method. Theoretically, all daughter languages of a proto-language should be equally different in basic vocabulary from the proto-language if the overall rate of change of the basic vocabulary is constant through time. A detailed study using this reasoning has been carried out by Blust (1981a), who has compared sixty-two Austronesian languages outside Taiwan with a reconstructed 200-word list for Proto-Malayo-Polynesian.

The results are quite dramatic: Western Malayo-Polynesian languages retain an average of 41 percent of cognates with Proto-Malayo-Polynesian (for example, Malay 59 percent, Minangkabau 50 percent, Tagalog 46 percent, Makassarese 38 percent, Sundanese 35 percent), while Oceanic languages retain only 25 percent on average (for example, Motu 37 percent, but most other New Guinea languages retain only 16 to 30 percent). These differences are significant, although there is much overlap, with some Western languages having low percentages (for example, Yogyakarta Javanese 30 percent) and some Oceanic languages having moderately high percentages (for example, Fijian and Motu 37 percent). Nevertheless, one can hardly escape the conclusion that Western Malayo-Polynesian languages have changed, on the average, much more slowly than Oceanic.

Are there any observations that can be made to throw light on this situation? It could be significant that certain languages, such as Atayal of Taiwan and Enggano of western Sumatra, have possibly been subjected to rapid changes due to "word taboos" (i.e., certain sounds, such as those occurring in the name of a dead chief, may acquire a taboo against further utterance; Dyen 1965a:53; Kahler 1978). Other languages, especially in the more stratified Austronesian societies, have different vocabularies for use by or when addressing persons of high status (Grimes and Maryott 1994). Others, such as Atayal again, have slightly different vocabularies for male and female speakers (P.J-K. Li 1983). But all these customs are restricted in occurrence and can hardly explain the overall picture. They are individual idiosyncrasies.

Clearly, there are other more fundamental sociolinguistic factors that are likely to affect rates of linguistic change in preliterate societies. Such variables as the size of the speech community, its internal integrity (the quantity of dialect
variation), and—most importantly—the external relationships between speakers of a given language and other communities with varying degrees of linguistic relationship, are all sure to be important. The latter factor has already been highlighted above. Different rates of linguistic diversification or convergence (languages do not always grow farther and farther apart) for large urban communities, mobile hunter-gatherers, and small groups of tribal gardeners may perhaps be expected. There are reasons for suggesting that the latter, particularly when they are atomized into small independent groups and have occasion for interaction with speakers of quite different languages, can have very rapid rates of vocabulary change. This has been shown especially for neighboring societies who speak Papuan and Austronesian languages in western Melanesia, where bilingualism, heavy interlanguage borrowing, and even complete language shift—a process that can often carry over major substratum aspects of an abandoned language—are all attested in great detail (e.g., Ross 1994; Dutton 1995; the significance of bilingualism in speeding language change is highlighted by Grace 1985 for New Caledonia).

We have now reached a point where further generalizations will serve little purpose. I have tried to show that the rate of linguistic change will be affected by the degree of communication maintained between related language communities and by various sociolinguistic factors. Small independent populations of tribal horticulturalists, especially if they inhabit a region of existing linguistic diversity, will tend to develop a fast rate of change. Large, integrated civilizations should change more slowly, especially if they have literacy. Basically, the only way to date nodes in the tree of a language family such as Proto-Austronesian is to reconstruct the words for those items of material culture that are likely to appear in the archaeological record, and then to search for those items in the dated stages of that record. For instance, early Austronesians made pots and kept domestic pigs and dogs. Such items occur widely in the Indo-Malaysian Archipelago after about 4,000 years ago but are completely absent before. When they first appear we have a good case for suggesting an Austronesian presence, although this is obviously not a foolproof way of identifying early Austronesians in space and time—other people made pots and kept pigs too, particularly Austroasiatic speakers. Yet this is the best method that we are ever likely to have. To claim greater certainty in tying together the archaeological and linguistic records is to daydream; we can only refine hypotheses.

The time has now come to look at some of the more detailed relationships of the Indo-Malaysian Austronesian languages and to suggest some archaeological correlations (which will be expanded in later chapters) for time-depth. I will not be able to raise glottochronology to the status of a reliable and useful tool, but I will try to give some understanding of the linguistic history of the two regions and to suggest why different rates of linguistic change have occurred.
V. INDO-MALAYSIAN LINGUISTIC PREHISTORY: SOME POSSIBILITIES

In the first edition of this book I wrote that there was no coherent information about the linguistic situation in Indonesia and Malaysia prior to Austronesian expansion, simply because this expansion was so successful and so complete that no traces of anything older had survived (excluding the Papuan languages in the east). Since then we have two important claims, one by Adelaar that there are traces of an Austroasiatic (specifically Aslian of Peninsular Malaysia) substratum in western Sarawak (Adelaar 1995), and another by Reid (1994a) that some of the Negrito languages of the Philippines carry traces of a non-Austronesian substratum, perhaps reflecting an ancient origin as pidgins but since deacreolized by continuing close contact with Austronesian populations.

In the case of Sarawak, there is a possibility that early rice agriculturalists might have occupied some regions before Austronesians arrived, according to radiocarbon dates for rice approaching 2500 BC from the cave of Gua Sireh (Bellwood et al. 1992; Ipoi and Bellwood 1991) (see Chapter 7, Section IIID). In addition, it has long been known that Malay, as spoken on the Malay Peninsula, has many loans from the Aslian subgroup of Austroasiatic. Aslian is presumably a descendant of a major pre-Austronesian language grouping of the region. It can thus be assumed that before the breakup of Proto-Austronesian and the expansion of its daughter subgroups southward, the unknown languages of the Indo-Malaysian Archipelago were flanked to the west (on the Southeast Asian mainland and in part of Borneo) by ancestral Austroasiatic languages, and to the east (in eastern Indonesia and New Guinea) by ancestral Papuan languages. The affinities of the Philippine Negrito substratum are unknown.

Turning now to the Austronesian languages themselves, I have already made it clear that the location of Proto-Austronesian is best placed in Taiwan. Internal relationships of Formosan languages need not be of concern here, but the rather limited archaeological evidence from the island suggests a date between 4000 and 3000 BC for Initial Austronesian settlement. My reasoning here is simply that pottery, a known cultural and linguistic marker of at least Proto-Malayo-Polynesian communities (although not strictly reconstructable to Proto-Austronesian, perhaps because of cognate attrition), first appears in Taiwan at about this time—perhaps 1,000 years before its appearance in any of the islands to the south.

As I have noted, however, the date of Proto-Austronesian does not correlate with the date of Initial Austronesian settlement in Taiwan. Instead, it correlates with the point at which the Initial Austronesian communities of the island, whether speaking one language or several, split irrevocably in such a way that two subgroups were formed, both of which have survived separately to the
Fig. 4.5 Approximate dates (derived from archaeological correlations) for the expansion of Austronesian settlement.
present day. Such separation could have occurred within Taiwan (and must have if Taiwan has more than one primary subgroup of Austronesian), or when people first moved from Taiwan into relative isolation in the northern Philippines. In the same way, Proto-Malayo-Polynesian correlates with a later separation that probably began as people fanned out through the Philippines, before moving south toward Borneo, Sulawesi, and the Moluccas. Proto-Austronesian and Proto-Malayo-Polynesian are not in themselves datable, as discussed above, but taking into account the expansion of relevant forms of material culture in the archaeological record, I would suggest a date of 3000 BC for the breakup of the former and 2500 BC for the breakup of the latter. These dates are a little younger than those preferred by linguists (e.g., Pawley and Ross 1993 suggest 4000–3000 BC for the breakup of Proto-Austronesian), but in my view the archaeological record is likely to have the last say on this.

At this point, I will return again to the observations made by Blust (1981a) on varying retention rates in the Malayo-Polynesian languages. If the average Oceanic retention figure (25 percent) is used to calculate a date for Proto-Malayo-Polynesian by using the standard formula of glottochronology (80.5 percent retention per language per millennium), then the answer will be around 4000–5000 BC. If the much higher average Western Malayo-Polynesian retention figure (41 percent) is used, then the answer will be about 2000 BC. Such a difference is no small matter, and there can be little doubt from the archaeological record that the Western Malayo-Polynesian rate, in general terms, is the most accurate one for the overall nonwestern-Melanesian prehistory of the Austronesian languages. The Oceanic retention rate is obviously highly depressed and may be considered unrepresentative of the general Austronesian situation. I will look further at the Melanesian languages in due course.

A. After Proto-Malayo-Polynesian

The dates I am suggesting for the Philippines indicate an Austronesian colonization at about 2500 BC, followed by fairly rapid movements to Borneo and Sulawesi by about 2000 BC. Whether these movements took place before the Philippines were wholly settled we may never know. The archipelago is highly fragmented, and settlement of remote parts may have continued long after other Austronesian settlers had expanded southward. If this did happen there may be a lesser linguistic diversity and time-depth for some parts of the Philippines than expected, as perhaps documented by Blust (1991) for his Greater Central Philippines subgroup (although Blust relates much of this language expansion to sociocultural dominance patterns after 500 BC). The Malayic languages have expanded through sociocultural dominance since about 2,000 years ago, as we will see below.
We now come to the two major divisions within Malayo-Polynesian: Western Malayo-Polynesian and Central-Eastern Malayo-Polynesian. The former has never been established as a true subgroup with rigor, but most of its members share quite high cognate percentages between themselves (over 38 percent according to Dyen 1965a, who termed this subgroup the Hesperonesian linkage). Many languages have large numbers of speakers (approaching 100 million in the case of Javanese) and are spoken over large areas. The overall internal relationships of Western Malayo-Polynesian are unclear, and the lexicostatistical clustering of these languages may be more an artifact of high retention rates from Proto-Malayo-Polynesian than of a true subgrouping relationship. Furthermore, many of the languages of this group have long histories of association with civilizations and have borrowed from each other so intensively that true shared innovations are often very hard to detect. Nevertheless, within Western Malayo-Polynesian a number of localized subgroups can be detected. Most of the Philippine languages form a single subgroup, although there is some disagreement about the details (Reid 1982; Zorc 1986). The languages of northern Sulawesi (Minahasa and Sangihe-Talaud) also clearly relate to those of the Philippines.

Another important subgroup, first proposed as the Javo-Sumatra Hesion by Dyen (1965a), has been examined in detail by Nothofer (1975, 1985, 1988) under the name Malayo-Javanic. It is stated by Nothofer to include Javanese, Sundanese, Madurese, and the "Malayic" languages (see below), plus Acehnese of northern Sumatra. This subgroup clearly occupies a huge geographical area and as the languages all share quite high cognate percentages, they give the impression of a recent major expansion—perhaps during the first millennium BC—although archaeological correlations in support are almost nonexistent for much of this region. This impression of recency could be tempered by high retention rates and undetected borrowings, but I think we can hardly escape the working hypothesis that many of the Austronesian languages of western Indonesia, Peninsular Malaysia, and Vietnam were not established prior to 1500–1000 BC, or perhaps even later in some areas.

The internal details of Nothofer’s classification have been challenged by Blust (1981b, 1988, 1995b), who doubts the existence of a single subgroup containing Javanese, Madurese, and Malay, and proposes instead a closer subgroup relationship between Malay, Acehnese, and the Chamic languages of southern Vietnam (the Malayo-Chamic hypothesis). Javanese in this scheme is believed to relate most closely to Balinese and the Sasak language of Lombok (Blust 1984–1985). Blust believes the Malayo-Chamic languages spread from a common homeland in western Borneo after about 300 BC and that Austronesian languages were once continuous on the Southeast Asian mainland between Peninsular Malaysia and southern Vietnam, a continuum later broken up by
historical Mon (Benjamin 1987), Khmer, Thai, and Vietnamese expansion. Acehnese is regarded as a late back-migration from the continent to Sumatra, carrying traces of borrowing from Austroasiatic sources. The close links between Malay and Cham are attested by comparisons of written forms of Old Malay and Old Cham that have survived in inscriptions from the first millennium AD (Marrisson 1975). There seems little reason to doubt that both these mainland groups share a close if not immediate common origin.
For the Malayic languages (Fig. 4.6), the most recent detailed analyses have been by Adelaar (1992), who proposes a basic division between Malay plus its closest relatives (including Minangkabau and Kerinci of Sumatra and the various Malay dialects spoken in Sumatra, western Java, coastal Borneo, and the Moluccas) and a "Malayic-Dayak" group of related languages on Borneo (Iban, Selako, Kendayan; Hudson 1970). Adelaar and Blust (1988) both agree that the spread of the Malayic languages has taken place from western Borneo since about 2,000 years ago, with much impetus doubtless being given by the growth of the trading empire of Srivijaya focused in southern Sumatra after AD 670. By AD 1500, Malay was a lingua franca over much of the archipelago, especially amongst the Islamic ruling families and their retainers.

Within Western Malayo-Polynesian we therefore see a widespread group of languages—some with many millions of speakers—in continuous contact in historical times and spread over much of Vietnam, Sumatra, Borneo, and all of Java and the Malay Peninsula (Blust 1994b). Much of this language spread has occurred since 1500-1000 BC, according to the linguists—who clearly disagree amongst themselves on details of subgrouping (see Nothofer 1991 for an overall review). There is no clear archaeological evidence from this region that strongly supports this date range or any other (Bellwood 1993), but what is also apparent from the linguistic record is that the other Western Malayo-Polynesian languages in the Philippines and Sulawesi, plus Batak and Gayo in northern Sumatra and the languages of the "Barrier Islands" off western Sumatra (Nias, Enggano, and Simeulue), do not form a part of this recent-spread situation, a circumstance recognized by Nothofer (1986, 1994). Nothofer refers to these languages as "Palaeo-Hesperoronesian," his implication being that the languages of the north Sumatran and Barrier Island regions in particular have been isolated from the cut and thrust of language expansion in the rest of western Indonesia. The linguistic prehistory of Indonesia is thus "layered," with considerable language replacement happening in centrally placed regions long after initial Austronesian dispersal. We cannot take the view that all the languages of the region have merely evolved in situ since Austronesian dispersal began.

The languages of Madagascar provide us with another interesting example of a late expansion, albeit in this case to a seemingly uninhabited island. The fairly uniform languages here probably originated in a settlement of Austronesians from southern Borneo, where the most closely related languages are Maanyan and Ngaju (Dahl 1951; Dyen 1971). According to Dahl, the presence of certain key Sanskrit loans in the languages of both Madagascar and southern Borneo suggest that the migration postdates AD 400, which is when the first evidence for the influence of Sanskrit on the Western Malayo-Polynesian languages appears. A date during the middle or late first millennium AD seems likely and is supported by Adelaar (1989, 1995), who believes, from the study of loan vocabulary, that the southern Borneo people might have been serving as
crews on ships captained by Malay- or Javanese-speaking overlords. On the issue of late expansion, it should not be forgotten that other Austronesians also reached New Zealand and much of eastern Polynesia, on the opposite side of the world from Madagascar, at about the same time.

On turning to the languages of the Central Malayo-Polynesian subgroup, we move into a very different linguistic situation from that of most of Sundaland. This region has a very large number of languages—about 100 according to Pawley (1974)—and most of them have small numbers of speakers and occupy only small geographical areas. There are linguistic indications that the Central Malayo-Polynesian languages spread very quickly once colonization started (Blust 1993), and the region as a whole has no clear signs of the type of secondary language radiation so visible in Sundaland and in the central Philippines. Blust's analyses indicate that the Central Malayo-Polynesian languages have a higher average retention rate (about 36 percent) from Proto-Malayo-Polynesian than do the Oceanic languages. Thus, while many of these societies are relatively small in scale, the higher retention rate surely reflects a much weaker influence from the pre-Austronesian inhabitants of the region than was the case in western Melanesia (see below).

To date, rather little can be said about internal relationships within Central Malayo-Polynesian, but it is apparent from archaeology that Austronesian settlers had reached Timor by perhaps 2000 BC, although the dating here is rather uncertain. The formation of the Eastern Malayo-Polynesian subgroup, which includes the Austronesian languages of southern Halmahera, parts of West New Guinea, and all of the vast Oceanic subgroup, seems to have commenced with the initial Austronesian colonization of Halmahera or the Bird's Head region of West New Guinea. The Austronesian languages of Halmahera have probably been spoken there since 1500 BC, and have here undergone much contact with the Papuan languages in the northern part of the island (Voorhoeve 1988, 1994; Bellwood in press). At around 1500 BC the Oceanic languages underwent their very rapid expansion through parts of Melanesia from an immediate homeland in the Bismarck Archipelago. The Oceanic languages form a well-defined subgroup with a common origin in Proto-Oceanic, as established by Dempwoff, Milke, and other more recent authors such as Grace, Pawley, and Ross (see Pawley 1981; Ross 1988, 1989; Pawley and Ross 1995).

In the first edition of this book I felt obliged to provide a fairly long argument in favor of equating Proto-Oceanic with the beginnings of the Lapita archaeological culture of island Melanesia at about 1500 BC. Today, this equation is so firmly accepted by linguists and archaeologists alike that it no longer needs lengthy justification. In New Guinea and adjacent parts of western Melanesia, the Oceanic speakers were apparently preceded by Papuan speakers, who may well have had systems of arboricultural food production in coastal regions and quite sizeable populations before Austronesians first arrived. Since
that time, the Melanesian Austronesian languages have undergone more rapid diversification than their cousins in Indonesia or Polynesia owing to intense contact-induced change and even language shift. Melanesia may be regarded as an area where small, independent, and relatively egalitarian communities, in frequent contact with a large array of different linguistic communities and practicing bilingualism as a major method of communication for trade and social interaction, have become subjected to very high rates of lexical diversification (Dutton 1995).

The same situation does not hold for most of Sundaland and may never have done so, although the relative linguistic homogeneity seen there now may be partly due to the integrating effects of 1,500 years of Hindu-Buddhist and Islamic civilization. Those islands that were little affected by these civilizations, such as Sulawesi, the Philippines, and the islands of eastern Indonesia, do show much more linguistic diversity, both in terms of the large numbers of languages with small numbers of speakers and the degrees of difference between them.

VI. THE PAPUAN LANGUAGES AND THEIR RELATIONSHIPS WITH INDONESIA

The Papuan languages have been mentioned above in connection with contact-induced change in the Austronesian languages, but they also deserve a brief consideration in themselves because of their reflection of the prehistory of the large island of New Guinea—a prehistory that has been of considerable importance for the eastern part of Indonesia. Within Indonesia (excluding Irian Jaya) they are spoken in central and eastern Timor, Alor, Pantar, and on Morotai and the northern half of Halmahera. Elsewhere in the Moluccas and Lesser Sundas all the languages are Austronesian.

A superficial glance at a map might suggest that these Papuan enclaves are simply remnants of an earlier and larger distribution overrun by Austronesian speakers, but recent research on the Papuan languages as a whole may make such a view rather simplistic. In the case of northern Halmahera, for instance, all the Papuan languages are closely related and clearly record recent radiation, although their first establishment on the island could well be pre-Austronesian (this is simply not clear; see below and Voorhoeve 1988, 1994 for discussion). Here I simply want to take an overall look at the prehistory of the Papuan languages, especially that espoused in a number of papers by Wurm (1978, 1982, 1983).

The main documented phases of expansion in the prehistory of the Papuan languages, according to Wurm, are as follows (and here I am simplifying what Wurm presents as a rather more complex story):
Fig. 4.7 The distribution of the Papuan languages, using the phylum divisions favored by Stephen Wurm and Don Laycock.
1. New Guinea was first settled about 60,000 years ago (according to inference based partly on Australian archaeological dates), but no present-day Papuan languages descend directly from this early linguistic phase.

2. About 15,000 and 10,000 years ago the first Papuan speakers, possibly in two separate groups according to pronoun forms, settled the New Guinea region, including the eastern Indonesian islands mentioned above. The dates may be regarded as little more than guesses. The languages of Halmahera and Morotai, which are classified in the West Papuan Phylum, descend directly from one of these two periods of linguistic expansion (although Voorhoeve does not record specific agreement with this view; see above).

3. About 3000 BC a major expansion of the Trans New Guinea Phylum of languages took place. This expansion began west of New Guinea, according to Wurm, and occurred after Austronesians had already arrived in the general region, according to the evidence of loan words (a suggestion that perhaps makes the claimed date of 3000 BC a little too old). The Trans New Guinea Phylum languages spread initially along the northern coast of New Guinea and also to Timor, Alor, and Pantar, where they replaced the earlier West Papuan languages.

4. Since 1500 BC, the Trans New Guinea Phylum languages, strongly influenced by Austronesian loan words acquired in the vicinity of the Markham Valley, have expanded in the highlands of New Guinea. The Phylum now contains about 500 languages and covers about four-fifths of the Papuan linguistic area. It is perhaps an excellent example of how a successful linguistic expansion can wipe out traces of earlier diversity.

The basic conclusions relevant for Indonesia that can be drawn from Wurm’s accounts are clear. According to him, Halmahera was settled by Papuan speakers long before any Austronesian presence in the area. Timor, Alor, and Pantar might have been settled by two separate groups, one long before and the other contemporary with an Austronesian presence. New Guinea has a far more complex history.

There are aspects of Wurm’s reconstruction that I think need some comment. First, in a very detailed review of Papuan linguistics by Foley (1986), the view is taken that the Trans New Guinea Phylum is not a valid construct and that the Papuan languages as a whole are, in reality, highly fragmented into about sixty separate families. On the other hand, Pawley (1995) and Voorhoeve have searched and found widespread lexical cognates across the area of the Trans New Guinea Phylum, thus suggesting that it could have some genetic basis as a true language family. Whatever the ultimate verdict on the Trans New Guinea Phylum, there is still the problem that Wurm derives all his strata of migration from Indonesia at separate times. It seems just as likely that many could represent expansions from within the large pool of population supported
by the large size and varied resources of New Guinea. For instance, there is now some excellent evidence indicating that New Guinea Highlanders developed their own localized form of horticulture prior to 6,000 years ago, long before Austronesians could have arrived on the New Guinea coasts (Bellwood 1996b). This would be sufficient to support a considerable increase of population over that possible with a hunting and gathering economy, and it would be sufficient to explain the expansionary success of the Trans New Guinea Phylum. It could also explain the remarkable resilience that the Papuan languages have shown in holding their domination of the New Guinea region, despite Austronesian settlement. Foley (1986) also believes that some of the language phyla he recognizes have resulted from agricultural population expansion. Unfortunately, there are no easy answers here, but as with Austronesian, it is important to recognize that not everything in the Papuan linguistic pattern has developed in situ through all time.

So far, most of the linguistic opinions I have presented in this chapter have been what I term "standard"; despite quibbles over details, they are all generally acceptable to the majority of linguists. One exception to this generalization is of course the hypothesis of Dyen that the Austronesian languages developed in and spread from Melanesia. The past twenty years of research in linguistics and archaeology leave little hope for a hypothesis that was originally so clearly presented and yet can now be shown to be incorrect. It was therefore with some surprise that I once read an article by Terrell (1981) that seemed to be harking back to Dyen's viewpoint, albeit from a totally different theoretical perspective. Terrell suggested that the Papuan and the Austronesian language families in Melanesia could have had a common origin and that the divergences between them evolved within Melanesia with the passage of time and a fast rate of linguistic change. As Terrell (1981:251) stated: "In short, in the interests of parsimony, do not invoke migrations beyond necessity."

The point I wish to make to close this chapter is that I believe the linguistic reconstruction I have presented fits the available facts in the most convincing way. Migrations can, in fact, be quite parsimonious explanations for patterns of human variation, although I would agree with Terrell that they are not always the answer for every problem in prehistory. Fortunately, few archaeologists today take the strong antimigration stance that dominated the discipline in the early 1980s. Indeed, a debate about the processes of cultural and biological evolution, with a partial focus on the issue of migration, has just erupted in the anthropological literature (Moore 1994; Bellwood 1996d). Terrell clearly adopted a totally different stance from mine; one that was reasoned and logical, but one that (in my view) was heavily dependent on one rather peripheral region of the Austronesian world (western Oceania) for its facts and theories. We may never know which answer, if any, best explains the Austronesian dispersal, but some answers surely have more interdisciplinary support than others.