

THE AUSTRONESIAN SETTLEMENT OF MAINLAND SOUTHEAST ASIA

Robert Blust

1. The linguistic macro-history of mainland Southeast Asia. Five distinct language families are historically attested in mainland Southeast Asia: 1. Austroasiatic (AA), 2. Sino-Tibetan (ST), 3. Tai-Kadai (TK), 4. Hmong-Mien (HM; in the older literature: Miao-Yao), and 5. Austronesian (AN). In this section I will briefly summarize the subgrouping and geographical distribution of languages belonging to each of these families, and from this information draw a set of inferences about the relative antiquity and primary center of dispersal of each genetic group within the region.¹ These inferences follow from a fundamental principle: areas of greater diversity are more likely to have been settled longer than areas of lesser diversity. This principle has a variety of applications. It is perhaps best known to linguists in its application to the determination of primary centers of dispersal, or homelands, as first formulated by Sapir (1916 [rptd: 1968]: esp. 452ff), and later formalized by Dyen (1956).² In this context it applies to collections of languages which have a single common ancestor, that is, to members of the same language family.

In considering the distribution of languages belonging to *different* language families, the diversity principle permits inferences about the most likely centers of dispersal of each family, but cannot tell us the order of priority by which these families arrived in their attested locations. To justify inferences about the relative antiquity of different language families in a given geographical region we need information not only about the internal subgrouping of each language family (relative diversity), but also about the separation times of the primary branches in each family (absolute diversity).

The latter consideration applies, for example, in situations where a continuous distribution of related languages, A, intersects an interrupted distribution of related languages, B. There are two subtypes of such a distribution: (1) A has greater absolute diversity than B, (2) B has greater absolute diversity than A. Type

(1) is most simply explained by a hypothesis that A was in place prior to B, and that the latter distribution is a product of migration around (or through) A without dividing the territory of the latter. On the other hand, Type (2) implies that B was in place prior to A, and that the latter split the territory of B through a more recent migration. For convenience we can call the type of migration inferred from distribution type (1) an indirect, or circuitous migration, and that inferred from distribution type (2) a direct, or splitting migration. To judge from the small number of reasonably clear-cut cases in the published literature, splitting migrations are the more common type. Such a situation appears to be found, for example, in aboriginal California, where the scattered pockets of distantly related Hokan languages contrast markedly with: 1. a solid block of more closely related Penutian languages which separates many of the Hokan languages north of San Francisco Bay from each other, and from the Esselen, Salinan, and Chumashan branches of Hokan further to the south, and 2. a solid block of even more closely related Uto-Aztecan languages (all belonging to the Takic branch of that family) which separates Esselen, Salinan, and Chumashan from the distantly related Yuman branch still further to the south (Shipley 1978).

1.1. Austroasiatic. According to Ruhlen (1987) there are some 155 AA languages distributed from northeastern India in the west to Vietnam in the east, and from Yunnan in the north to the Malay Peninsula and the Nicobar Islands in the south. By universal agreement the primary split within AA separates the Munda languages of northeastern India from the typologically dissimilar and distantly related Mon-Khmer languages of mainland Southeast Asia. Ruhlen (1987:148) describes Munda and Mon-Khmer as 'subfamilies' of Austroasiatic. Within Mon-Khmer he recognizes three primary branches: North (including Khasi, Palaungic-Khmuic, and Viet-Muong), East (including Katuic, Bahnaric, Khmer, and Pearic), and South (including Monic, Aslian, and the languages of the Nicobar Islands). As the useful map accompanying Lebar, Hickey and Musgrave (1964) shows clearly, the distribution of AA languages is very fragmentary; only in Vietnam and the interior of the Malay Peninsula are there relatively unbroken blocks of AA speakers. Given this

subgrouping picture, the most economical hypothesis would be that Proto-Austroasiatic was spoken in the region of present-day northern Burma or Assam.

1.2. Sino-Tibetan. Ruhlen (1987) lists 258 ST languages, which fall into two primary divisions: Sinitic, with 12 languages, and Tibeto-Karen, with 246 languages. The latter division in turn contains two primary branches: Karen, with 14 languages, and Tibeto-Burman, with 232 languages. Tibeto-Burman further divides into 1. Tibetic (75 languages in the Himalayas), 2. Baric (16 languages in the Himalayan foothills), and 3. Burmic (141 languages mostly in the hill country of the Assam-Burma borderland). In striking contrast with AA, the Tibeto-Karen languages extend in a relatively narrow, but solid band from north of the 25th parallel in the Naga Hills of Assam, to south of the 11th parallel in peninsular Burma and Thailand. Unfortunately, information on absolute separation times for AA and Tibeto-Karen languages is lacking. However, impressionistically the separation of Munda from Mon-Khmer appears to be greater than that of Karen from Tibeto-Burman, and far greater than the separation of the Burmic languages from their immediate common ancestor.³

Given the above observations it appears simplest to assume that the AA languages already occupied the Assam-Burma borderland (and adjacent regions to the east and west) at the time that Tibeto-Karen languages began to expand southward from the eastern Himalayas. As a result of the latter splitting migration the Munda languages and Khasi in the west became separated from the remaining members of the AA family further to the east and south.

1.3. Tai-Kadai. According to Ruhlen (1987) there are some 57 TK (called 'Daic') languages, divided into two primary subgroups: 1. Lati-Gelao, and 2. Li-Kam-Tai. Almost all of the most divergent members of this group are located in southern China (Guizhou, southwest Hunan, Guangxi, Guangdong, border region of Yunnan and Vietnam). The closely related Central and Southwestern Tai languages occupy an almost continuous block of lowland ricelands from southern China to the northern half of the Malay Peninsula, with an extension into the Shan states of

Burma. This distribution strongly suggests a relatively recent expansion from southern China, and the inferential evidence from language is closely supported by the evidence of documentary history (cf. Hall 1955, Chapter 7, or Wyatt 1984: Chs. 1, 2, who sketch a picture of gradual Tai infiltration of the Salween, Menam, and Mekong basins over the past millenium).

1.4. Hmong-Mien. Ruhlen (1987) lists four HM (called 'Miao-Yao') languages, of which three are in the Yao group. These are scattered in small pockets over much of southern China (Hunan, Guangxi, Yunnan) and northern mainland Southeast Asia. They are closely related, and appear to have moved southward from southern China over the past several centuries.

1.5. Austronesian. The AN languages of mainland Southeast Asia fall into three discrete groups: 1. Chamic, 2. Malay, and 3. Moken-Moklen.

There are ten reported Chamic languages, as follows: 1. Utset/Utsat (also called 'Huihui' = 'Moslem' in the Chinese sources), on Hainan Island, China, 2. Jarai, on the Darlac Plateau of Vietnam, with extensions into eastern Kampuchea, 3. Rhade, on the Darlac Plateau south of the Jarai, also with extensions into eastern Kampuchea, 4. Haroi, east of Cheo Reo in Vietnam (hence between the Jarai and the coast), 5. Eastern Cham, mostly along the south-central coastal plain of Vietnam from about Phan Rang to Phan Thiet, 6. Western Cham, mostly around the Tonle Sap in central Kampuchea, 7. Chru, near the valley of Dran, between Dalat and Phan Rang, Vietnam, 8. Southern Raglai, inland from Phan Rang, Vietnam, 9. Northern Raglai, in the mountains inland from Nhatrang, Vietnam, and 10. Caccia Raglai, near the Pacific Ocean, in Ninh Thuan Province, Vietnam. Lebar, Hickey and Musgrave (1964) also mention Bih, Krung, Noang, and Rai, which are not included by Ruhlen, and Lee (1966) mentions Bih and Rai, but not Krung or Noang. The Chamic languages are closely interrelated, and have undergone far-reaching typological adaptations to their neighbors. Both Utset (Benedict 1984, Thurgood and Maddieson 1992), and Eastern Cham (Edmondson and Gregerson to appear) have developed contrastive contour tones, the former probably through contact with one or both of the indigenous TK languages of