

# KALENJIN

**2,500,000 SPEAKERS**

*Kenya*

**O**ne of the NILO-SAHARAN LANGUAGES, Kalenjin is spoken by a group of traditionally pastoralist peoples in north-western Kenya.

*Kalenjin*, meaning ‘I tell you’, is a term coined by the speakers themselves in the 1940s to emphasise their linguistic and cultural unity. They were earlier grouped as the *Nandi dialects*: the separate dialects include Nandi, Kipsigis, Päkot and others (see map at TUR-KANA).

The migration that brought a language of the Nilo-Saharan family to this region may have taken place in the 14th century. In the 19th century the Nandi were the best known of the Kalenjin-speaking peoples. They were then a fiercely independent people, defeating the Masai in the 1870s after long warfare, and successfully preventing Swahili caravans from taking Nandi slaves and even from travelling through their country. They resisted British administration for fifteen years, from 1890 to 1905.

The savagery of the final British conquest and deportations left a legacy of lasting hostility.

A. C. Hollis’s book *The Nandi: their language and folk-lore* (Oxford: Clarendon Press, 1909; revised by G. W. B. Huntingford, 1969) is still cited by anthropologists, but Hollis wisely found his informants outside the borders of Nandi country. Indeed, ‘his principal informant, Arap Chepsiet, sometimes introduced himself to visiting officials as “the man who wrote Bwana Hollis’s book”’.

Kalenjin has ten vowels and three tones. The language is unusual in having a three-way distinction of vowel length. It has no distinction between voiced and unvoiced consonants. Nouns have separate indefinite (‘primary’) and definite (‘secondary’) forms, and sex or gender prefixes, *ki-*, *ce-*: *kípnándiín* ‘Nandi man’, *çèpnándià* ‘Nandi woman’; *kibléngwa* ‘hare’, *çepkoikòs* ‘tortoise’.

The first ten numerals are: *akenge*, *aeng*, *somok*, *angwan*, *mut*, *kullo*, *tisap*, *sisiit*, *sokol*, *taman*.

*Atinye cheptánnýō nepiyyonyi mutai, korukut lakat*  
– ‘I have a daughter who gets plenty to eat every morning and goes to bed hungry every night.’ The answer? A broom.

# KAMBA

**2,500,000 SPEAKERS**

*Kenya*

**A**lthough clearly related to its neighbours MERU and KIKUYU (see map there), Kamba may, to judge from its well-established dialect divisions, be among the oldest established of the BANTU LANGUAGES of northern Kenya.

The local name of the language is *kiKamba*, of its speakers *aKamba*, and of their country *uKamba*. They were called *waringao*, 'naked people', in Swahili.

Kamba speakers were once an important trading people of inland Kenya: their colony in Rabai, near coastal Mombasa, was perhaps established for trade reasons and certainly kept up trade links with the homeland. Swahili traders, too, came to the edge of Kamba country but did not enter it. Kamba shared a local market system with the

Masai and Kikuyu areas to the west. In the 1920s travellers still found Kamba a particularly useful language, known as a lingua franca by many Kikuyu and Masai speakers and by the smaller peoples around the foothills of Kilimanjaro. For Kamba speakers themselves Swahili was the most useful second language, followed later by English.

The British East Africa Company set up a trading post at Machakos in 1892. The country became a British 'protectorate' in 1895 and some land was taken by English and Afrikaans speakers from 1903 onwards.

Kamba has a seven-vowel system, sometimes written *i ē e a o ũ u*, though the accents marking open *i* and *u* are often omitted in print. The first ten numerals are: *ĩmwe*, *ilĩ*, *itatũ*, *inya*, *itaano*, *thanthatũ*, *mũonza*, *nyanya*, *kenda*, *ĩkũmi*.

# KAMTOK

## PERHAPS 2,000,000 SPEAKERS OF KAMTOK AS A LINGUA FRANCA

### *Cameroun*

**O**ne of the ENGLISH CREOLES AND PIDGINS, this language has had many names in its short life: *Cameroons Pidgin*, *Wes Cos* ('West Coast'), *Broken English*. *Kamtok* ('Cameroun talk') has now caught on.

At the eastern end of the continuum of English creoles that begins with KRIO of Sierra Leone, there is little difference between Kam-

tok and the West African Pidgin English of Nigeria: the two are in general mutually intelligible. Both serve principally as an everyday medium of trade and interchange among speakers of different local languages – but with the linguistic fragmentation of coastal Cameroun there is a growing number of first-language speakers of Kamtok in the larger towns.

#### Kamtok and Krio proverbs

##### Kamtok

alata neva hapi wen  
pusi bɔn pikin

trɔki wan fait  
bot i sabi se i han sɔt

wan han no fit tai  
bɔndu

Mouse is never pleased when  
Cat has kittens

Tortoise wants to fight but  
(he knows) his arms are short

You can't tie a bundle with one  
hand

##### Krio

arata nɔba gladi we  
pus bɔn pikin

trɔki wā  
bot i an shɔt

wan an nɔ ebul tai  
bɔndul

From Loreto Todd, 'Cameroonian' in *Readings in creole studies* ed.  
Ian F. Hancock (Gent: Story-Scientia, 1979) pp. 281–94

# KANNADA

27,000,000 SPEAKERS

*India*

**K**annada (Canarese, to give its older English name) belongs to the DRAVIDIAN LANGUAGES of southern India. It is the state language of Karnataka.

As a language of the Rashtrakuta and Yadava dynasties in Maharashtra, used in inscriptions along with Sanskrit, the influence of Kannada once reached further north than it does now. Indeed, MARATHI shows strong evidence of co-existence with its Dravidian neighbour. In turn Kannada has been influenced by Sanskrit, the learned language of India, and later by Portuguese, radiating from nearby Goa, the metropolis of Portugal's eastern possessions.

Known from inscriptions of the 6th century and after, Kannada has a surviving literature from the 9th century onwards. Some early texts are lost – their existence known or surmised through such sources as Kēśirāja's 13th-century grammar ('Jewel mirror of grammar') and the 9th-century textbook on poetry, *Kavirājamarga*. The *campu* epic, of mixed prose and verse, was a major genre in the 10th to 12th centuries. Lyric poetry came later. Kannada was a major language of literature under the Empire of Vijayanagar, from 1336 to 1575. The beginnings of a modern literary language can be traced to the middle of the 19th century.

Great numbers of Indo-Aryan loanwords, especially from Sanskrit, have brought into educated Kannada speech a contrast between aspirated and unaspirated stops (e.g., *ph bh* contrasting with *p b*) which is foreign to Dravidian languages. The typical Dravidian rolled *r* and fricative *ɻ* are no longer distinguished in modern Kannada. For a table of numerals see

TELUGU.

## Kannada, Telugu and Tulu

Standard spoken *Kannaḍa* is based on the colloquial language of Bangalore and Mysore City. The Dharwar dialect of the north of Karnataka, beyond the former borders of Mysore, forms a regional standard of its own. The coastal dialects are also quite distinct.

### Languages of the Nīlgiri Hills

Besides the languages of the Todas and the Kōtas, two other languages are vernacular on the Neilgherry Hills – viz., the dialect spoken by the Burghers or Baḍagars (the northern people), an ancient but organised dialect of the Canarese; and the rude Tamil spoken by the Iḷulars ('people of the darkness') and Kuruburs (Can. *Kurubaru*, Tam. *Kurumbar*, shepherds), who are occasionally stumbled upon by adventurous sportsmen in the denser, deeper jungles, and the smoke of whose fires may occasionally be seen rising from the lower gorges of the hills.

Robert Caldwell, *Comparative grammar of the Dravidian or South-Indian family of languages*, 3rd edn (London: Kegan Paul, 1913) p. 34

Baḍagaga baḍāse, The Baḍaga wants mutton,  
Kōtaga pōtāse, The Kōta wants beef,  
Todavaga hālāse, The Toda wants milk,  
Kurumaga jēnāse. The Kurumba wants honey.

Paul Hockings, *Counsel from the ancients: a study of Baḍaga proverbs, prayers, omens and curses* (Berlin: Mouton De Gruyter, 1988)

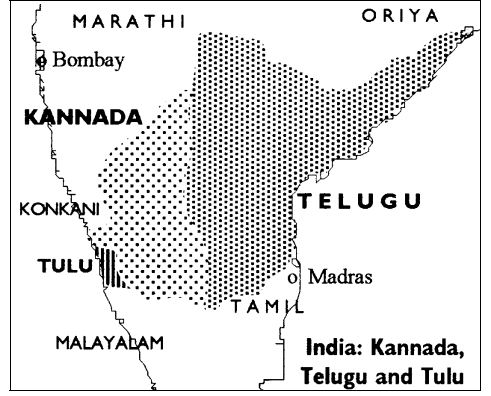
Kannada has two close relatives which are sometimes counted as dialects. *Badaga* is a language of the Nilgiri Hills, now with 100,000 speakers, descendants of 16th-century emigrants from Mysore.

*Kōdagu* or Coorg, with 90,000 speakers, is the language of the old hill state of Coorg, independent until British annexation in 1834.

*Telugu* dialects are usually classed in four groups: Northern, Southern, Eastern and Central. The last-named, the speech of East and West Godavari, Krishna and Guntur, has contributed most to the modern standard.

*Tulu* is spoken in a coastal region of southern Karnataka, around Mangalore.

*Toda* and *Kōta*, mentioned in the *Badaga* proverb (see box), are Dravidian languages of small tribal groups, with less than a thousand speakers each.



### Writing in Kannada

The Kannada alphabet is in all essential features identical with that of Telugu, though it has its own typical printed style. Consonants are combined with following vowels to form a single symbol. Most compound consonants involve the writing of one element below the line.

#### Kannada script: the consonant characters

ಕಖಗಘಙ ಚಛಜಞ ಟಠಡಢ ತಥದಧನ ಪಫಬಭಮ ಯರಳ ಶಷಸವಳ

k kh g gh ṇ c ch j jh ṇ ṭ ṭh ḍ ḍh ṇ t th d dh n p ph b bh m y r ḷ v ś ṣ h !

# KANURI

**4,000,000 SPEAKERS**

*Nigeria, Niger, Chad and Cameroun*

**K**anuri is a language of the Saharan group of NILO-SAHARAN LANGUAGES.

*Kanurí* is the local name of the language, *Kanúrí* that of the speakers: their homeland is *Kanem*, east of Lake Chad. *Bornu*, name of a second traditional state, is the name of the Kanuri language in some 19th-century English records.

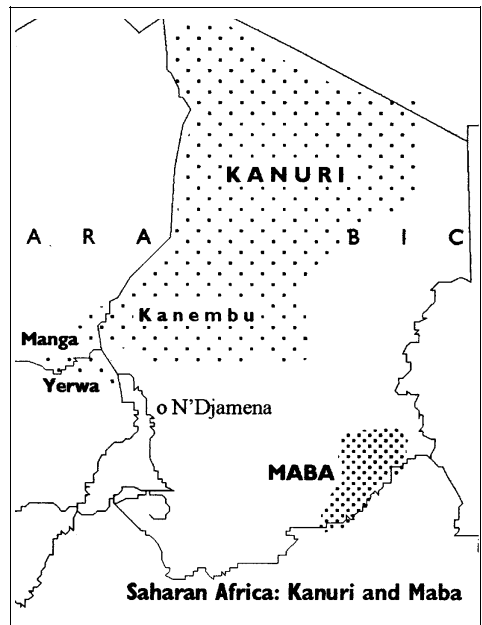
Legends tell of Kanuri origins in a migration from Arabia, the migrants intermarrying with the indigenous people of the shores of Lake Chad. In fact the history of the Kanem-Bornu Empire may go back to the 8th century, but it became an Islamic state in the 11th and remained a major force in central Africa for eight hundred years. Originally centred on Kanem, east of Lake Chad, internal power shifted to Bornu (north-eastern Nigeria, west of the lake) in the 14th century. The empire's power reached its zenith under Prince Idris Aloma (ruled 1571–1603), whose conquests and long distance diplomacy encouraged trade and cultural development.

The names of the old tribes incorporated in the empire are still remembered, but their languages disappeared over the centuries as Kanuri, the prestigious language of court and government, took their place. Across southern Chad the territories of vassal states of the later empire can be traced – Logone, where the native language is SARA; Wandala in the northern salient of Cameroun, where they speak Mandara, a Niger-Congo language; Bagirmi in central Chad, where the main language, Bagirmi, is Nilo-Saharan; Kanem itself; Damagherim; and even Wadai, where the local language is MABA. In these, the provincial

languages remain in use, though for some centuries Kanuri was the language of diplomacy, education and long distance trade. Arabic, the language of the Qur'ān and of religious education, is widely known in the lands of the former empire.

Kanuri has six vowels and a complex tone pattern based on two main tones, high and low. Suffixes, such as the plural *-a*, alter the tone of preceding syllables: *fērò* 'girl', *fērò-á* 'girls'. The first ten numerals in Kanuri are: *tílò*, *ndī*, *yásqo*, *dēgu*, *ūgu*, *rásqo*, *tullur*, *usqu*, *legár*, *mēgū*.

The first printed record of Kanuri, a list of numerals, appeared in 1790. A fuller printed vocabulary, collected from a Bornu traveller by



J. L. Burckhardt in Cairo, was published in 1819; but the best early work on Kanuri was by Sigismund Koelle, author of the *Polyglotta Africana*. In 1854, the same year in which this great work appeared, he also published a *Grammar of the Bornu or Kanuri language* and a collection of folklore, entitled *African native literature, or proverbs, tales, fables and historical fragments in the Kanuri or Bornu language, to which are added a translation of the above and a Kanuri-English vocabulary*.

### Kanuri and MABA

The three dialect groups of Kanuri are: *Yerwa Kanuri* (or Bornu) of north-eastern Nigeria, the centre of the empire; *Manga Kanuri* of south-western Niger, the old northern province; and, *Kanembu*, the language of the ancient heartland of Kanem, now in Chad.

MABA, once the language of the state of Wadai, is also a member of the Nilo-Saharan family, but is only distantly related to Kanuri.

# KAREN LANGUAGES

**PERHAPS 4,000,000 SPEAKERS**

*Burma, Thailand*

**T**he largest ethnic minority in east central Burma are the speakers of Karen languages, which form a sub-group of the Tibeto-Burman Languages, or, some would say, a branch of the family of SINO-TIBETAN LANGUAGES on an equal level with Chinese on one side and the Tibeto-Burman languages on the other. This is a measure of the long independent history of the Karen languages, which must stretch back several thousand years, but not of the current number of speakers. There are no more than a few million.

*Kareang*, *Karieng*, are names used in Mon and Thai for Karen speakers and their languages: the English name derives from these. In Burmese they are called *Kayin* and in Shan *Yang*. The old-fashioned name *White Karen*, in Burmese *Kayin byu*, in Thai *Yang khao*, was often used for the lowland Karen, speakers of Pwo and Sgaw.

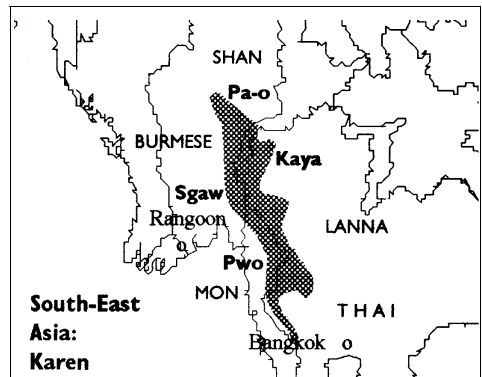
*Karyan* 'Karens', and *Cakraw* perhaps 'Sgaw', are found in Burmese records of the 13th century, and there is no reason to suppose that they were new arrivals even then. Little is known of their history till the early 19th century, when missionaries in British-occupied Tenasserim began to work with speakers of Sgaw and Pwo, devising scripts (based on Burmese) for these two languages. Translations and other publications were soon appearing. A vast *Thesaurus of the Karen knowledge*, by Sau Kau-too and the Rev. J. Wade, was published at Tavoy in 1847–50.

Karen speakers now occupy two quite distinct kinds of terrain: the flat coastal plains of the Irrawaddy, Sittang and Salween deltas, and the abrupt hills and valleys of the Tenasserim range

that marks the border between Burma and Thailand.

Lowland Sgaw, and most Pwo speakers, are merged – administratively and to some extent culturally – with Burmese. They tend to be Buddhists, and to be bilingual in Burmese: indeed, the Karen languages may well be in steep decline in these areas, but no recent information is available. Karen State, established in 1952, includes many of the speakers of Sgaw who are hill dwellers. Kayah State marks the territory of Kaya (Karenni) speakers. With continuing warfare and oppression in Burma the numbers of Karen language speakers in Thailand has increased massively in recent years.

Karen languages – like Chinese – are overwhelmingly monosyllabic. Pwo and Sgaw both have four tones in open syllables. The object normally follows the verb, and modifiers (adjectives and adverbs, in our terms) follow the word that is modified, much as in Thai. There are a great many Burmese loanwords in Karen, mostly quite recent loans of cultural terms. Some loans from Mon and Thai are also found.





### A lullaby

'When playing with the chickens, children sometimes catch one of them and pretend to rock it to sleep, droning the while:

Hsaw po, mi, mi –  
n'mo n'pa leh hsu Yo.

Heh ke so ne na  
p'theh tha wa ko lo.

Aw gha lo gha lo –  
me aw, hsaw hpo.

Sleep, sleep, little chick,  
mother and father have gone to Shanland.

They will come back  
and bring you plenty of betel nuts.

You can eat them one by one –  
sleep, little chick.'

Harry Ignatius Marshall, *The Karen people of Burma* (Ohio State University, 1922) p. 174

### The Karen languages on the map

*Sgaw* has possibly 2,000,000 speakers. They call themselves *Kanyaw*; in Burmese they are often known as *Bamā Kayin*, 'Burmese Karen' because their territory is close to the central areas of Burmese speech.

*Pwo* has possibly 1,250,000 speakers, often called in Burmese *Talaing Kayin*, 'Mon Karen': they are lowlanders in regions that are or were Mon-speaking.

*Pa-o* has perhaps 500,000 speakers in southern Shan State. In Burmese its speakers are called *Kayin net* 'Black Karen' or *Taungthu*, 'hill people', and in Shan *Tongsu*.

*Kayā* speakers, of whom there may be about 300,000, used to be called *Kayin ni* in Burmese with the equivalents *Red Karen* and *Karen-ni* in English. *Kayah* is now the usual English form.

### Numerals in Karen languages

	Sgaw	Pwo	Pa-o
1	ta'	là'	tà'
2	khí	nì	nì
3	thō`	thàn	thòm
4	lwì	lí	lít
5	yè	yé'	ngát
6	khū'	khù	thù
7	nwí	nwè	nō't
8	khò'	kho'	thōt
9	khwí	khwí	kút
10	shí	shì	chì

Numerals follow the noun and are followed by classifiers: thus Pwo *li' la bi* 'one book', literally 'book one flat-item'; *ghi ni phlo* 'two houses', literally 'house two round-items'.

Paul K. Benedict, *Sino-Tibetan: a conspectus* (Cambridge: Cambridge University Press, 1972) pp. 129, 131 and Robert B. Jones, *Karen linguistic studies* (Berkeley: University of California Press, 1961)

# KASHMIRI

3,000,000 SPEAKERS

*India, Pakistan*

**K**ashmiri is the best-known member of the Dardic group of INDO-ARYAN LANGUAGES – the only one that has unequivocally reached the status of a literary language.

The Kashmir Valley has probably spoken Kashmiri for a long time: it is reasonable to suppose that the Dardic languages have been here as long as the main group of Indo-Aryan languages has been spreading from north-western India, that is, for three thousand years or more. There is a venerable literary tradition, going back to the 13th century, when the poet Sitikantha chose ‘the language of the people, understood by all’ for his manual of Tantric worship, *Mahanaya-prakasa*. From the next century come the verses of the great woman poet of Kashmir, Lalla Ded. Later, Persian literature had a pervasive influence: there is even a Kashmiri translation of Firdausi’s vast epic *Shah-nama*.

Through all this time the government of Kashmir has been conducted in other languages. For many centuries it was Sanskrit, and the early history of Kashmir is recorded in the Sanskrit chronicle compiled by Kalhana in 1148, *Rājatarāṅgiṇī* ‘Ocean of Kings’. After Akbar conquered Kashmir in 1589, Persian became the language of government (as it was for the whole of northern India), and remained so until the early 20th century. It was then replaced by Urdu, which still functions as the government language both in Pakistani Kashmir, where it is also the national language, and in Indian Kashmir. Many speakers of Kashmiri use Urdu as their literary language.

There is, however, literature and a press in

Kashmiri, which is normally printed in a form of Perso-Arabic script similar to that used for Urdu. The official post-1950 version of this script (which some writers have not fully adopted) includes diacritical marks for all the vowels of Kashmiri. The *Sharada* alphabet, a descendant of ancient Brahmi, is not so often seen.

Numerals in the Dardic languages and Khowar			
Kashmiri	Shina	Kalami	Khowar
1 akh	ek	ā, ak	ī
2 zəh	dū	dū	jū
3 trih	trē	ṭhā	troi
4 chōr	cār	cōr	cōr
5 pānch	push	panj	pōnj
6 sheh	shah	shō	choi
7 sath	sat	sat	sot
8 aiṭh	ach	aṭh	osht
9 nav	nau	num	nəoh
10 dah	dāī	dash	josh

The Dardic languages include several of those spoken in the valleys of Kashmir (see map at INDO-ARYAN LANGUAGES). Their distinctive feature, among the Indo-Aryan group, is precisely the extent of their differences from the rest of the group and from one another – a sign of their long history of separate development in a region where communications are far more difficult than in the Indian plains. Shina, with about 300,000 speakers, is relatively close to Kashmiri but much less influenced by the languages of ‘civilisation’ such as Persian and San-

<p>skrit. It is spoken to the north, west and east of the Kashmir Valley. Kalami is also known as</p>	<p>Garwi or Kohistani. For KHOWAR see separate entry.</p>
---	---

# KAZAKH

**8,000,000 SPEAKERS**

*Kazakhstan, Russia, China*

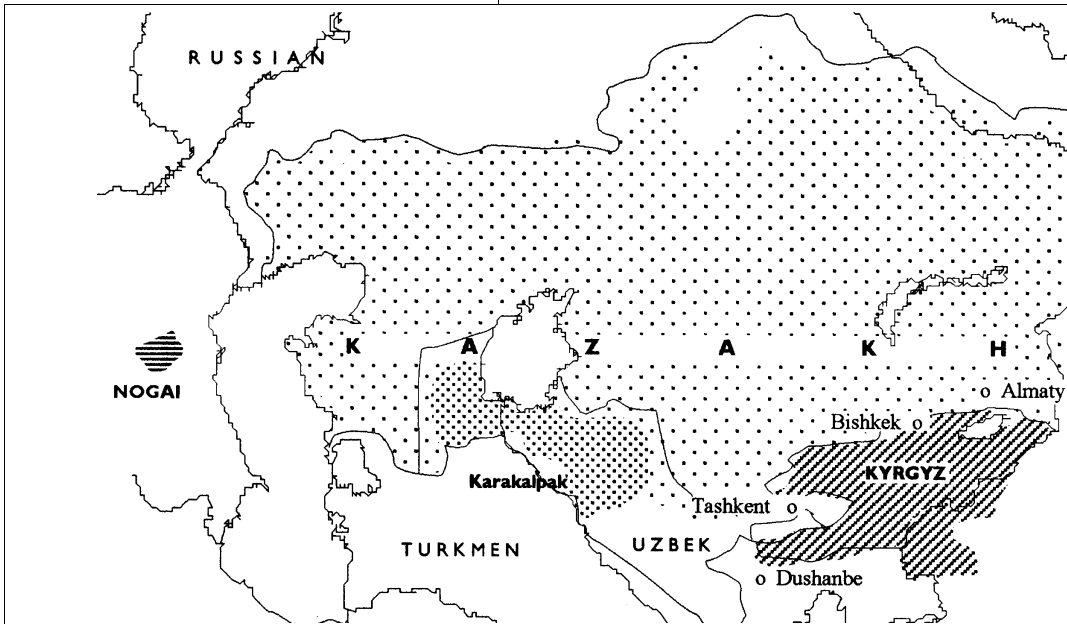
**O**ne of the **TURKIC** LANGUAGES, Kazakh is spoken by a very widespread, traditionally nomadic people of the Central Asian steppe. There are over five million speakers in Kazakhstan and about a million each in neighbouring areas of Russia and the Chinese province of Xinjiang.

The vast steppes of Kazakhstan were the scene of successive migrations of Turkic and other peoples for many centuries. Historical atlases show one government after another holding sway here, but none – at least until Soviet rule in this century – really ruled the nomadic population, whose name *Qazak* ('vagabond', see below) suggests their miscellaneous origins and rebellious character. If a single origin is to be assigned to the Kazakh and their language, they should perhaps

be considered as a more widely ranging offshoot of the already nomadic **KYRGYZ**.

Until the 18th century, Chagatai (see **UZBEK**) was the written language used in the region. Russian political domination of the Kazakhs developed in the early 19th century. At the same time Islamic **TATAR** missionaries, speakers of a related Turkic language, began to set up religious schools among the Kazakhs.

Kazakh is a language with a long tradition of oral epic. But the first published Kazakh-language poetry was that of the *Zar Zaman* 'Time of Trouble' poets in the 1840s: notable among them was Muhammad Utemis uli (1804–46). Educated in the Islamic schools, they naturally wrote Kazakh in Arabic script – and incited revolt



against Russia. One eventual response was the devising of a Cyrillic writing system for Kazakh to be used in Russian-organised secular schools. Ibrahim Altynsarin (1841–89), Russian-educated, led this trend in Kazakh culture. In the short term it failed. Kazakh in Cyrillic script was not widely accepted, and by the time of the Revolution even schools and local government were using Arabic script once more.

In 1927, as the short-lived Kazakh nationalist movement was being rooted out, the Latin alphabet was imposed and Arabic was outlawed. As Soviet policy unrolled, Cyrillic script returned, to replace Latin, in 1940. But in this, the largest of the Soviet Republics after Russia itself, centralisation of language policy faltered in the last decades of the old Soviet Union. The media gained some slight independence, and there were moves to reduce the number of Russian (as well as Persian and Arabic) loanwords in Kazakh. Kazakhstan became an independent state in 1991.

For a table of numerals see KYRGYZ.

### Kazakh and Kyrgyz on the map

A people called *Kyrgyz* is recorded in very early Chinese and Greek sources. They appear to

have moved southwards from the upper Yenisei valley, in Siberia, to occupy their present territory in the Tien Shan mountains.

#### Wandering name

*Qazaq*, in Turkic languages including Chagatai and Tatar, means 'adventurer, vagabond, free man, bachelor'. It became the name of the most persistently independent population group of central Asia, the *Kazakhs*. Kazakhs were once to be found among the elite of the Mughal Empire in India. In 18th-century India, as reported by English travellers, *kuzzauk* was the local name for predatory bands of adventurers, little better than highwaymen. Far to the west in the plains of Ukraine, and gradually eastwards as well, Russia's borders were settled and guarded by autonomous villages of horse-riding warrior-farmers, the *Cossacks* (Russian *Kazak*). At least partly Turkic in origin, Cossacks long remained central to rural Russian society and to Russian politics.

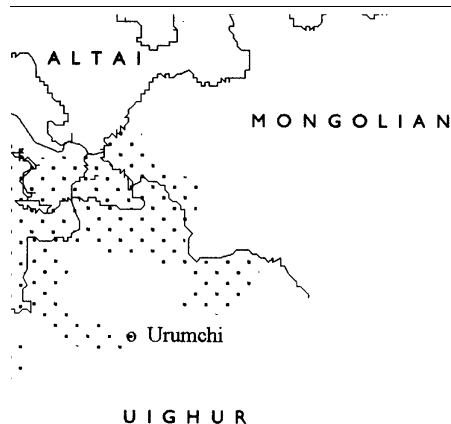
The horse-riding nomads of central Asia had a typical garment, a cloak: French *casaque*, English *cassock*, 'soldier's greatcoat; priest's gown'.

*Kazakh* and Russian are mixed in a complex mosaic of settlement along the northern steppe from Orenburg to Semipalatinsk. Forced migration of peasant families under Stalin, followed by the 'Virgin Lands' colonisation under Khrushchev, led to a huge increase of Russian speakers in this zone. South of it lies the 'hungry steppe', sparsely populated.

There are about a million speakers of Kazakh in China, and one sizeable community in the Hovd basin in western Mongolia.

*Karakalpak* is a dialect of Kazakh which has been given the status of a literary language. There were recently 300,000 speakers, settled near the mouth of the Amu Darya in Uzbekistan – in a region faced with ecological catastrophe by the drying up of the Aral Sea.

After the collapse of the once powerful Nogay Horde, whose warriors were of mixed Mongol and Turkic origin, hundreds of thousands of



Central Asia: Kazakh and Kyrgyz

<p><i>Nogai</i> speakers settled in what is now northern Dagestan in the 17th century after extensive wanderings. They were conquered by Russia in the early 19th century, and after a twenty-five-year uprising in 1834–59 the majority fled to</p>	<p>Turkey, where the numbers of identifiable Nogai (or ‘Tatars’) are now in decline as they slowly assimilate to the Turks among whom they live. Only around 55,000 speakers of Nogai now live in their ancestral lands.</p>
--	--

# KHASI

500,000 SPEAKERS

## *India and Bangladesh*

**K**hasi is one of the main Mon-Khmer group of AUSTROASIATIC LANGUAGES, spoken well to the north-west of all other members of the group. It is the state language of the Indian state of Meghalaya, whose capital is Shillong (see map at MUNDARI).

Medieval Indian legends of a land of female magicians, Kudali, may very well have something to do with the matrilineal and, some would say, matriarchal society in which Khasi speakers still live, and with the sorcery which some Khasi women still practise.

When they are first heard of in historical records, in the 16th century, Khasi speakers already made up twenty-five chiefdoms, which persisted through British times into the period of Indian independence. The Kingdom of Jaintia, more outgoing than the rest, paid tribute to the Ahom kingdom to the north and made conquests and raids into the Sylhet region (in modern Bangladesh) to the south. Jaintia came under British rule in the 1850s, but the other Khasi chiefdoms remained nominally independent, and as such were transferred to the suzerainty of the Governor of Assam in 1947 as the United Khasi-Jaintia Hills District, later to be joined with the Garo Hills in the state of Meghalaya.

Yet Shillong, in the temperate Khasi hills, had made an ideal provincial capital for all of British Assam. There was also early interest in Khasi speakers on the part of Welsh Presbyterian missionaries, who devised a Latin orthography for Khasi in 1842, on the basis of the dialect of Cherrapunji, which preceded Shillong as a radiating point for British influence. The missionaries also introduced primary and secondary

education and founded a theological seminary. Thus, though politically independent, Khasi speakers in fact underwent significant English-speaking cultural and linguistic influence. Meanwhile, Bengali-speaking Sylhet had long been the centre of trade penetration into the Khasi hills. Thus Khasi has numerous loanwords from Bengali, Urdu and English. It has also been influenced by neighbouring Sino-Tibetan languages.

The system of eight-day markets is so central to Khasi rural life that in each country district the days of the week are named after the eight accessible market villages. Markets are centres for recreation (notably archery contests), courtship and the transmission of news.

### **Welsh missionaries and Khasi spelling rules**

The Latin alphabet as used for Khasi has some most unusual features – which turn out to correspond with features of WELSH spelling, because the system for writing Khasi was devised by Welsh missionaries. In the Khasi words *dap* ‘full’ and *dab* ‘bullock’, there is no difference in the final consonant sound: it is a voiceless *-p* in both. The difference is in the vowel sounds: a long *a* in *dab*, a short *a* in *dap*. The missionaries thought they heard different final consonants because, in Welsh, a voiced consonant like *-p* is preceded by a shorter vowel sound, while a voiceless consonant like *-b* is preceded by a longer one. Unique among South Asian scripts, this odd way of writing the difference between short and long vowels is perfectly efficient: it is easier than writing a diacritic and shorter than doubling

the vowel letter. Many Khasi words begin with a two- or three-consonant cluster. Where these did not match Welsh or English consonant groups, the missionaries pronounced and wrote the words with two syllables, inserting the neutral vowel *shwa* (the vowel of English *the*) where they thought they heard it. This vowel is written *y* in

Welsh, as in *mynydd*, ‘month’: hence Khasi *bndi* ‘mortgage’ is usually spelt *byndi*, and *bna* ‘hear’ is often spelt *byna*.

After Eugénie J. A. Henderson in *Austroasiatic languages: essays in honour of H. L. Shorto* (London: School of Oriental and African Studies, 1991) pp. 61–6



# KHMER

8,000,000 SPEAKERS

*Cambodia*

**K**hmer belongs to the Mon-Khmer group of languages within the Austroasiatic family. Khmer and Vietnamese are the only two AUSTROASIATIC LANGUAGES that are national languages in a modern state.

The language name is *Khmae* in the standard language but is still *Khmer* in 'Northern Khmer', the dialect spoken in eastern Thailand.

*Cambodia* is the Latinised form, used internationally, of Sanskrit *Kamboja*. Other modern variants of the name are French *Cambodge* and Khmer *Kampuchea*. The medieval Hindu and Buddhist kingdoms of south-east Asia often labelled themselves with classical names out of Sanskrit literature: this was one. Two thousand five hundred years ago Kamboja was the name of a minor kingdom somewhere in north-west India.

Khmer has a history traceable to the 7th century. It is the language of the great culture that built the sacred capital of Angkor, centre of a powerful kingdom between the 9th and 15th centuries. Its bilingual culture was a strong, continuing influence on Thailand. The monuments were abandoned to the jungle, but the language continued to be a vehicle of Buddhist and Hindu scripture and poetry. For ninety years from 1864 the weakening Cambodian kingdom was a French protectorate: it regained independence in 1954, but collapsed in anarchy in the early 1970s after American air attacks.

The oldest dated Khmer inscription goes back to AD 611. Hundreds of inscriptions from the following centuries allow the development of the language to be divided into periods: 'pre-

Angkorian Old Khmer' to 802, 'Angkorian Old Khmer' from then till 1431, 'Middle Khmer' until the 18th century, then 'Modern Khmer'. As the monumental civilisation of Angkor fell into senescence, Middle and Modern Khmer were recorded less in inscriptions than in palm leaf manuscripts – and eventually in printed books. The traditional literature of Khmer shows early Indian inspiration transformed by its south-east Asian context. The story of Rama, the *Ramakirti* or *Reamker*, is an enduring classic. Royal chronicles are an important historical source.

Khmer words have one or two syllables: if two, the first is unstressed and may be skipped in colloquial speech. But Pali and Sanskrit loanwords can have several syllables. These loanwords are found in large numbers, with other special vocabulary, in the 'monks' language' and the now obsolescent 'royal language', special speech registers used in polite address.

Linguists find many similarities between the structures of Khmer and of Thai, two unrelated languages which coexisted for many centuries, exchanging literary and cultural influences. In modern times Khmer has borrowed scientific and cultural terms from French and has continued to build neologisms on the basis of Sanskrit and Pali.

The first ten numerals in Khmer are: *muəy*, *pīi*, *bəy*, *buən*, *pram*, *prammuəy*, *prampīi*, *prambəy*, *prambuən*, *dəp*.

## Register: clear and 'breathy' vowel quality

The concept of 'register', distinct from 'tone' and now considered central to the phonetic

study of many south-east Asian languages, was first introduced and defined in an innovative study of Khmer.

'The Cambodian 'registers' differ from tones in that pitch is not the primary relevant feature. The pitch ranges of the two registers may sometimes overlap, though what I shall call the *Second Register* tends to be accompanied by lower pitch than the *First Register*.

'The characteristics of the first register are a 'normal' or 'head' voice quality, usually accompanied by a relatively high pitch.

'The characteristics of the second register are a deep, rather breathy or 'sepulchral' voice, pronounced with lowering of the larynx, and frequently accompanied by a certain dilation of the nostrils. Pitch is usually lower than that of the first register in similar contexts . . .

'In sentences the word registers are modified according to intonation and by emotional factors. Register may be used, as in many other languages, to express emotion, and when this happens the emotional register may overlie the lexical register, much as in many tone-languages intonation may overlie lexical tone.'

Eugénie J. A. Henderson, 'The main features of Cambodian pronunciation' in *Bulletin of the School of Oriental and African Studies* vol. 14 (1952) pp. 149–74  
According to one recent phonemic analysis, the concept of register is valid for earlier forms of Khmer but not for the 20th-century language, in which, instead, there is a huge number of vowels – as many as thirty.

**Mon and Khmer**

Modern MON can be divided into three dialect groups: Pegu, Martaban and Moulmein, and Ye. Mon villages in central Thailand – resulting from settlements in the last three or four hundred years – speak dialects close to those of Martaban and Moulmein.

*Khmer* is the majority language of Cambodia.

In eastern Thailand there are 500,000 speakers of 'Northern Khmer'

**Khmer vowels**

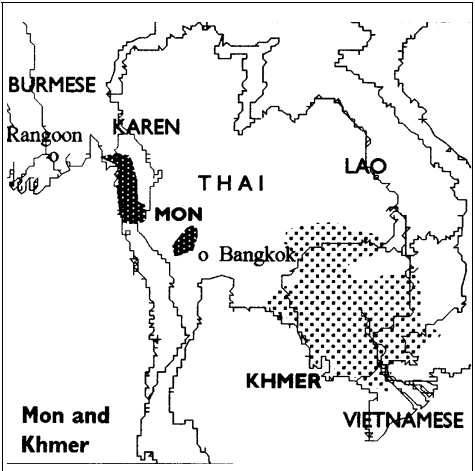
អា អិ អឺ អី អ៊ី អ៊ុ អ៊ួ អើ អឿ អៀ អេ អែ  
ឺ ឿ ឺវា ឺវា ឺវា ឺវា ឺវា ឺវា ឺវា ឺវា ឺវា ឺវា

Twenty-three written Khmer vowels in traditional order are shown here with the character *a*. Fewer vowels are found in Old Khmer texts.

**Khmer script**

Khmer script, like others of south and south-east Asia, descends from the Brahmi of ancient India – in this case by way of the *Pallava* script which was used both in south India and in Indochina in the 5th and 6th centuries AD. It has continued to develop through the centuries.

From the beginning Khmer script was used both for the classical tongues – Sanskrit and Pali – and for the contemporary local language, Old Khmer. It provided the immediate inspiration for the invention of the THAI alphabet, which was specifically designed for the very different sound patterns of Thai. In the Thai kingdom (until a few decades ago) Khmer script continued to be used, in the traditional way, for writing and printing Pali.



# KHOISAN LANGUAGES

**T**hree language families remain in southern Africa as a reminder that its linguistic history is more complex than the story of the Bantu expansion of the first millennium AD.

For convenience these three are grouped together as Khoisan or Khoesan languages, after the Khoe names for themselves, *Khoe* (plural *khoe-khoe*), and for their longstanding enemies, *San*. The word *khoe-i* ‘person’, origin of the language name, also survives as a loanword in Xhosa, *ikwayi* ‘commoner, deposed chief’. *San*, like *Bushman* (Dutch *Bosjeman*), is a derogatory term for a despised people.

Modern *Namibia*, covering part of what has been known for longer as *Namaland*, bears the name of the *Nama* people and their language, which is spoken by larger numbers than any other survivor of the Khoisan families.

Until their modern decline, Khoisan speakers were highly important transmitters of culture in southern Africa. The word for cattle in the Nguni languages (Zulu, Xhosa and others) is borrowed from a Khwe language: was the agricultural practice of cattle-keeping borrowed, along with the word, by Bantu speakers from Khwe speakers? Was the keeping of sheep – older here by some centuries than that of cattle – introduced to the southern edge of Africa by Khwe speakers, two thousand years ago, and from whom had they got the idea? Alongside these subsistence necessities, luxuries may also be traced to Khoe and its relatives. For ‘smoking of tobacco or cannabis’ southern Bantu languages use a verb *-daka* (Zulu), *-dzaha* (Tsonga) apparently borrowed from a Khwe language, which itself must have borrowed from Arabic *dakkhana* ‘to puff’.

Khoe and Nama, the two best-known members of the Khwe family, have a long history in

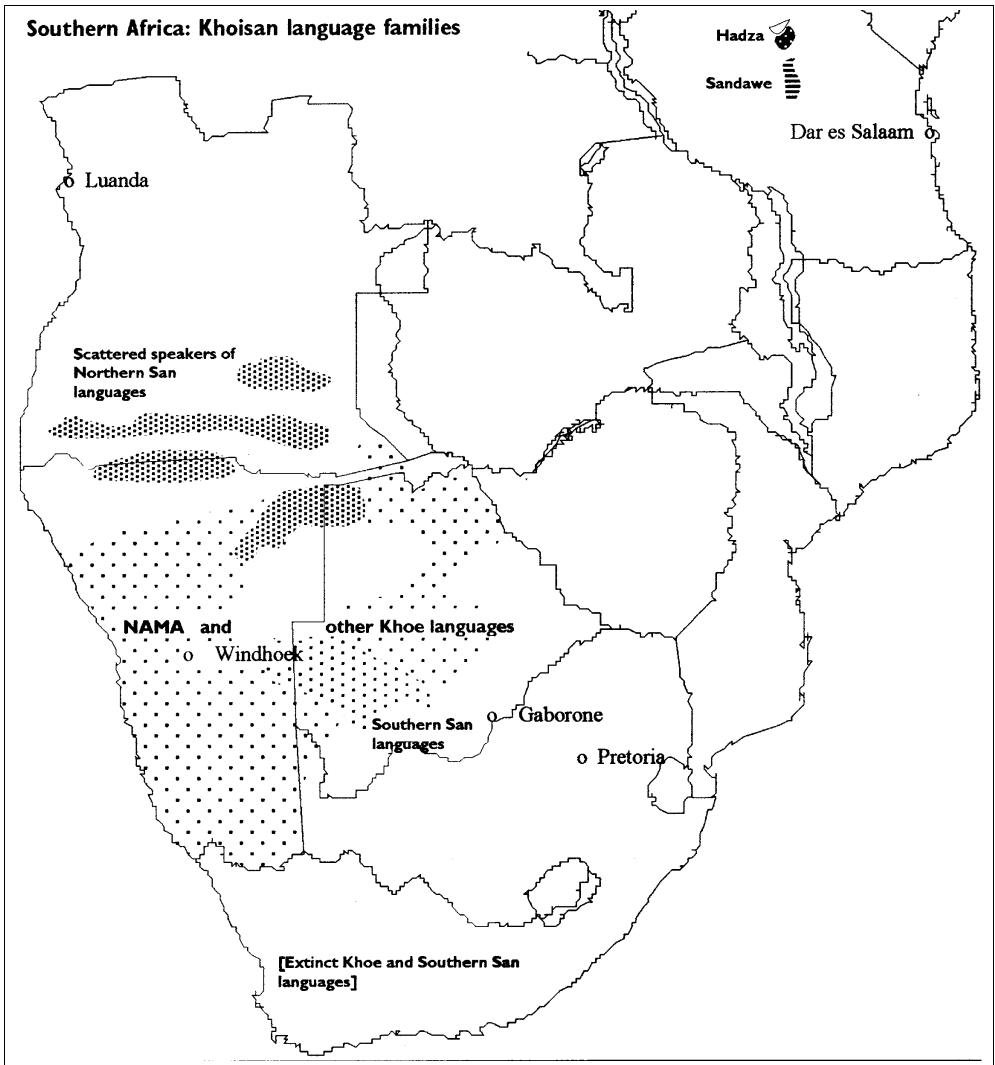
what is now South Africa. It has been calculated that the Khoe dialect chain (of which practically nothing now remains) had been developing *in situ* for at least fifteen hundred years. Khoe and Nama arrived where they now are as the result of an expansion or migration through south-eastern Africa. Kwadi speech probably reached southern Angola by way of a westward movement to the Atlantic coast. These movements may have begun about 2,500 years ago somewhere near the Limpopo valley – and a further Khwe language, ‘Limpopo Khoi’, extinct and unrecorded, can be resurrected from loanwords in proto-South-eastern Bantu, linguistic ancestor of Sotho, Tsonga, Zulu and other modern languages.

Kwadi is recently extinct. The use of Nama seems to be in rapid decline. Khoe, a language of what is now Cape Province, was gradually replaced by *Khoe Dutch*, and eventually by AFRIKAANS: place names and plant and animal names survive as reminiscences of Cape Khoe.

Oblique strokes are usually used to write the click sounds that are typical of Khoisan languages. The first ten numerals in Nama are: /*gúí*, /*gàm*, /*nòna*, /*hàga*, /*góro*, /*náni*, /*hū*, /*khéisa*, //*khòise*, /*disi*. If they have worked in the South African mines, !Kung speakers count with Fana-galo (see ZULU) numerals, ultimately from English: *wán*, *thú*, *trí*, *fóri*, *fáifi*, *síkisi*, *sébhèni*, *éti*, *náini*, *téni*. The traditional numeral systems of Khoe and !Kung are more complicated than this: ‘3’ literally ‘it’s a few’, ‘4’ literally ‘it’s many’ or ‘it’s two and two’, ‘5’ literally ‘it finishes the hand’, ‘10’ literally ‘both hands dead and finished’.

## The Khoisan language families on the map

The Khwe or Khoe or Hottentot or Central Khoisan languages are spoken by small groups in Angola and South Africa, and by as many as



150,000 speakers of *Nama* in Great Namaland, southern Namibia. *Khoe*, a language of what is now Cape Province, once had an equal number of speakers, and a chain of dialects stretching from the Cape itself to the Kei river and beyond. A great number of Khoe speakers died in small-pox epidemics in the 18th century. *Kwadi* and other Khoe languages have ceased to be spoken in the course of the 20th century.

The Southern Bushman or Southern San languages are now almost extinct; there are a few hundred speakers in Namibia and Botswana. The */Xam* or 'Cape Bushman' language was

wiped out between 1750 and 1900 by Khoekhoe and Europeans, who gradually enslaved and exterminated its remaining speakers.

The *!Kung* or Ju or Northern Bushman or Northern San languages have a few thousand speakers in northern Namibia and southern Angola.

*Sandawe*, with as many as 70,000 speakers in far-off Tanzania, may be a long-separated relative of the Khwe languages. Its neighbour *Hadza* may be a linguistic isolate. It has only 200 speakers, hunter-gatherers in inhospitable country near Lake Eyasi.

# KHOTANESE

## Extinct language of Central Asia

**B**elonging to the multilingual history of the Silk Road in Central Asia, Khotanese is the name given to a group of Eastern Iranian dialects, now extinct, rediscovered in documents and literary texts from the neighbourhood of Khotan in Chinese Xinjiang, dated between the 7th and 10th centuries.

*Khotana* was the Prakrit (Indo-Aryan) name for this city-state. The people themselves called it *Hvatana*; its Chinese name now is *Hotan*.

Since it was located well to the east of all other recorded IRANIAN LANGUAGES, scholars guessed that Khotanese was related to the language of the nomadic Iranians of the steppes, the *Saka*, *Skythoi* and *Scythi* of Persian, Greek and Latin history – and the guess appears to have been right.

Some Khotanese vocabulary seems to go back

to a period when its speakers were Zoroastrians, like *urmaysde* ‘sun’ from the name of the Zoroastrian supreme deity *Ahuramazda*. But the Khotanese were early converts to Buddhism, and a large proportion of the vocabulary comes from Prakrit dialects, like *ṣṣamana* ‘monk’, or from Buddhist Sanskrit. In late Khotanese, political terms are borrowed from Tibetan: *ṣkyaisa* ‘official gift’. The Tumshuqese dialect of Khotanese has borrowings from TOCHARIAN B – business words such as *kapci* ‘finger seal’.

The first ten numerals in Khotanese are: *ṣṣau* / *ṣṣa*, *duva* / *dvi*, *drai*, *cahora*, *paṃśa*, *ksətə*, *hauda*, *haṣṭa*, *nau*, *dasau*. The script – alongside other influences – came from northern India. It is a variant of the early Brahmi script.

Based on R. E. Emmerick, ‘Khotanese and

‘Tumshuqese’ in *Compendium linguarum iranicarum* ed. Rüdiger Schmitt (Wiesbaden: Reichert, 1989) pp. 204–29 and other sources

# KHOWAR

**PERHAPS 100,000 SPEAKERS**

*Pakistan*

**K**howar, the everyday language of the remote valley of Chitral, is one of the INDO-ARYAN LANGUAGES (see map there). Its closest linguistic relationships are not clear, because – though influenced by several of its neighbours, Iranian, Nuristani, Dardic and Burushaski, and by languages of culture such as Sanskrit, Persian and Urdu – it also shows many striking differences

from neighbouring languages. A table of numerals is given at KASHMIRI.

Alongside the Urdu that is Pakistan's national tongue, a place has been made for Khowar as a standard local medium of communication. There is now some writing and publishing in the language, and a strong tradition of oral poetry.

## A fawn and its mother

Nan-ei, nan-ei, af hera mosh goyan –  
*Nano zhan-ei, ano pazhal no-a?*  
 Nan-ei, nan-ei, thuek lapheika prai –  
*Nano zhan-ei, yoro zahri no-a?*  
 Nan-ei, nan-ei, ta pazo lei goyan –  
*Nano zhan-ei, tambuso xel no-a?*  
 Nan-ei, nan-ei, coghuwan pon kuri?  
*Nano zhan-ei, Phureto an no-a . . .*

Mother, mother, there's a man coming –  
 'Mother's love, is it not the mountain shepherd?'  
 Mother, mother, a rifle's flashing –  
 'Mother's love, is it not the sun's rays?'  
 Mother, mother, blood runs down your breast –  
 'Mother's love, is it not the sweat of summer?'  
 Mother, mother, where is the orphan's road?  
 'Mother's love, is it not the Purit Pass . . .'

The passes from the Chitral to the Panjkora valley are the only land link between Chitral and the rest of Pakistan.

'Some Khowar songs' collected by Wazir Ali Shah, translated by Georg Morgenstierne, in *Acta orientalia* vol. 24 (1959) pp. 29–58

# KIKUYU

4,300,000 SPEAKERS

*Kenya*

**K**ikuyu, which belongs to the BANTU LANGUAGES, is one of the major languages of Kenya, spoken in a swathe of territory between Nairobi and Mount Kenya and by many inhabitants of the capital.

The local names for the people, *aGikuyu*, the country, *uGikuyu*, and the language, *kiGikuyu*, demonstrate a local sound change, 'Dahl's Law', that is found in several Bantu languages. Some linguists use the form *Gikuyu* or *Gekoyo*, but most prefer *Kikuyu*, which is usual among speakers of other Kenyan languages.

Historically Kikuyu speakers were a fiercely independent, indeed militaristic, people, growers of sorghum and millet, keepers of goats and sheep, and collectors of honey. The Swahili slave trade route from the Uganda region and Lake Victoria crossed the southern tip of Kikuyu country, but the Kikuyu themselves were too formidable to be seriously threatened by slavers. In the late 19th century they resisted the British invasion violently, but in vain: the best land was seized for large scale farming, and many Kikuyu, classed as 'squatters', took work as farm labourers or in the colonial capital, Nairobi. The attempt by the invaders to rule them through appointed 'chiefs' had little success.

Kikuyu speakers were prominent in the Mau Mau and other resistance movements. Jomo Kenyatta, president of independent Kenya, was Kikuyu in origin and was the author of the best available study of Kikuyu society (*Facing Mount Kenya*, 1938).

Between Kikuyu and English speakers Up-Country SWAHILI was the usual lingua franca, and

Swahili remains essential in modern Kenya, though English is also widely known. Kikuyu has numerous English loanwords, sometimes hard to recognise at first sight, such as *njanji* 'judge'. Kikuyu itself is necessary as a second language to speakers of the smaller languages of the neighbourhood, such as Chuka, Embu and

## Kikuyu, Kamba and Masai

'The Kikuyu and Masai were constantly at war, raiding for cattle. The Masai were invincible on their own terrain, the plains, but were easy victims for the Kikuyu if they ventured into the forests, where they were killed by the arrows and staked war-pits which awaited them. At times, however, peace was made and relationships established by intermarriage. Trade also occurred between the tribes, women being able to pass freely from one to the other [even when they were at war]. Stigand states that Kikuyu often "lent" their cattle to be tended by the Masai, and Eliot writes that in the great famine of 1882 Masai settled in Kikuyu and took Kikuyu wives, sometimes entering the service of Kikuyu as "a sort of mercenaries" . . .

'The Kikuyu have various myths of origin which reflect their relations with other peoples. One tells that the three sons of Mumbere, the creator of the world, Masai, Gikuyu and Kamba, were given a choice by their father of a spear, bow, and digging-stick: Masai chose the spear, Kamba the bow, and Gikuyu the digging-stick.'

John Middleton, *The central tribes of the North-Eastern Bantu* (London: International African Institute, 1953) pp. 13, 15

Tharaka. There has been long term mutual interaction between Kikuyu and MASAI (see box).

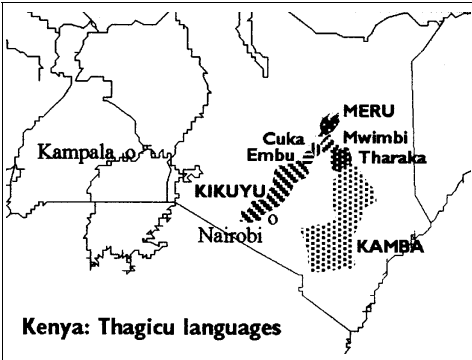
The first ten numerals in Kikuyu are: *ĩmwe*, *igĩrĩ*, *ithatũĩ*, *inya*, *ithano*, *ithathatũĩ*, *mungwanja*, *inyanya*, *kenda*, *ikunmi*.

**The Thagicu languages**

*Thagicu* is the name of the region from which speakers of Kikuyu and related languages may have dispersed. It was suggested as a name for the language group by P. R. Bennett in 1967.

The *West Thagicu* branch includes Kikuyu and three smaller Kenyan languages, *Embu*, with 250,000 speakers, *Cuka* or Chuka, with 70,000, and *Mwimbi*, with 70,000.

The Thagicu group also includes three other languages, MERU (1,250,000 speakers), *Tharaka* (100,000) and KAMBA (2,500,000). Meru serves as



a second language for some speakers of Chuka and Tharaka. Kamba has two main dialects, *Machakos* and *Kitui*, each subdivided: people of *Ulu*, the ‘high country’, call those of Kitui *aDaisu*, ‘lowlanders’. There are Kamba speakers in Rabai near Mombasa.



# KIRIBATI

**70,000 SPEAKERS**

*Kiribati*

**O**ne of the Micronesian group of AUSTRONESIAN LANGUAGES, Kiribati takes its name from the chain of atolls that form the republic of Kiribati. As British possessions these were known as the Gilbert Islands.

*Gilbertese* and *Kiribati* are two versions of the same word – for Kiribati is the local language form of ‘Gilbert’. Captain Thomas Gilbert put the main island of the group on the map in 1788: its older name is Tarawa. The people now call themselves *I-Kiribati*, with the same ethnic prefix that is seen in *I-Aotia* ‘Australian’. In modern sources *I-Kiribati* is often wrongly given as the name of the language.

The Gilbert Islands were British-ruled from 1892 until 1979, interrupted by Japanese occupation during the Second World War. English was and is the main foreign influence on the Kiribati language.

Kiribati (see map at MARSHALLESE) is spoken in the Kiribati chain itself, in Banaba (Ocean Island) and now also in the Line Islands, one of which is Kiritimati (Christmas Island). All these form part of the Republic of Kiribati. The

language also extends to Niu, which belongs to Tuvalu. Resettlement and work migration have taken Kiribati speech to Gizo and Wagina in the Solomon Islands, to Rabi in Fiji, and to Nauru.

Kiribati is unusual among Micronesian languages: it has only five vowels, where most have seven to twelve; it retains final vowels; its usual word order is verb-object-subject. These details have been seen as showing Polynesian influence – but Sheldon Harrison argues that they are conservative features, retained from the proto-Oceanic which is the ancestor of both Micronesian and Polynesian languages.

Yet linguistics shows that Kiribati was once under strong Polynesian – perhaps TONGAN – influence. There are numerous Polynesian loanwords, suggesting trade and cultural penetration, such as *moa* ‘chicken’, *baurua* ‘sailing canoe’, *rongorongo* ‘news’.

Examples from S.P. Harrison, ‘Linguistic evidence for Polynesian influence in the Gilbert Islands’ in *Language contact and change in the Austronesian world* ed. Tom Dutton, Darrell T. Tryon (Berlin: Mouton De Gruyter, 1994) pp. 321–49

# KITUBA

**PERHAPS 4,000,000 SPEAKERS**

*Congo (Kinshasa), Congo (Brazzaville)*

**A** creole based on KONGO (see map there), Kituba originated in the 16th century as a language of trade on the routes that led to the kingdom of Kongo. Kongo traded with the Portuguese of Angola, whose demand for ivory, slaves and other products of the centre of Africa was insatiable.

*Kituba* is often known in Congo (Kinshasa) as *Kikongo ya letá* or *Kileta*, 'language of government' (from French *l'état* 'the state'), or *Kisodi*, 'language of soldiers', or in French as *Kikongo véhiculaire* 'trade Kongo' or *Kikongo simplifié*. In Congo (Brazzaville) its usual name is *Munukutuba*.

Although the kingdom of Kongo and its overland trade have disappeared, Kituba remains – in three distinct regions radiating from Malebo Pool. South-east of Kinshasa, along the southern tributaries of the lower Congo, in a region where many languages cluster together in a small area, Kituba is the means of communication, the most-used second language, known in town and country alike. Missionaries use it for preaching and teaching. In Brazzaville, Kituba is an urban language, used by the majority of inhabitants and important in broadcasting. Downriver in Dolisie and Pointe-Noire in Congo (Brazzaville), and in Boma, Matadi and Thysville in Congo (Kinshasa), Kituba is the mother tongue of many speakers.

In the surrounding country, however, Kituba has not spread. Kongo, Kituba's parent tongue, remains vigorous in and around these lower Zaire towns and is favoured in education, though local officialdom uses Kituba. In Kinshasa itself and

its suburbs, those who have moved from the surrounding countryside tend to use Kongo while those who come from upriver use LINGALA: Kituba is less known.

Kituba has been influenced by French and Lingala, the other lingua francas of the lower Congo. It is also influenced by Zaire Swahili and by Portuguese. But it grew up as a trade language for speakers of Kongo and other Bantu languages, and remains closer to its roots than many other creoles. It even retains a limited set of noun class prefixes.

## Verb tenses in Kongo and Kituba

I am eating	ndiá	móno kéle diâ
I will eat	nidiá	móno áta diâ
I ate	ndiidi	móno diâ-áka
I have eaten	nádiâ	móno méne diâ

In Kongo persons and tenses are built into single-word verb forms by internal changes of sound and tone. This kind of word-building is also typical of the older Indo-European and the Semitic languages, called by linguists 'synthetic'.

In Kituba, as in many other lingua francas and creoles, separate, unchanging words are combined to indicate person and tense alongside a verb. This structure, also familiar from English and Chinese, is called 'isolating'.

Example from S. S. Mufwene in *Substrata versus universals in creole genesis* ed. P. Muysken, N. Smith (Amsterdam: Benjamins, 1986) pp. 129–62

# KOMI

360,000 SPEAKERS

*Russia*

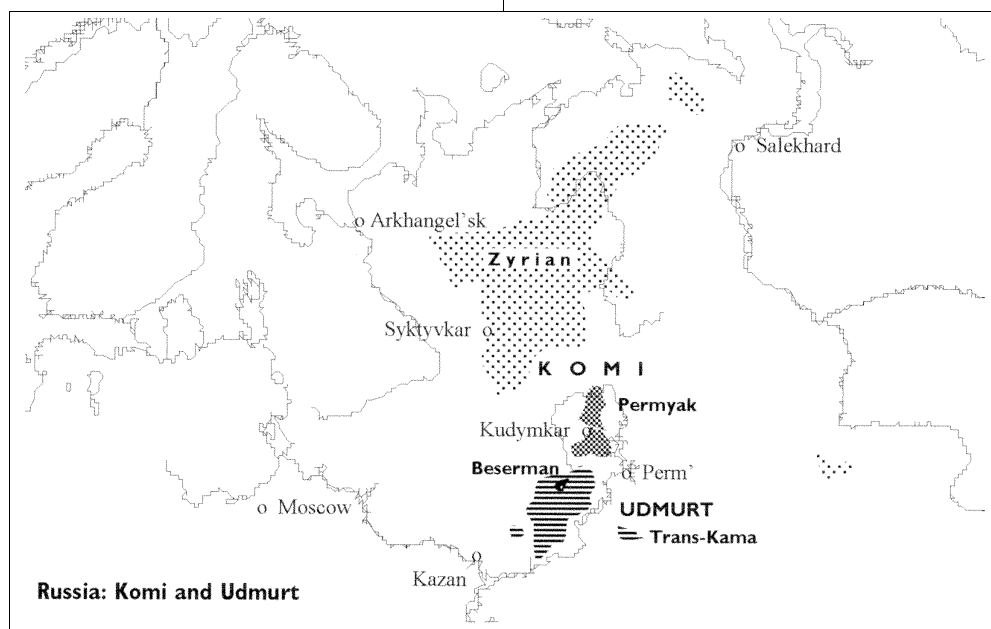
**O**ne of the URALIC LANGUAGES, Komi is spoken in the Komi Republic in north-eastern European Russia along river valleys that drain into the Barents Sea.

Spreading northwards in prehistoric times from the territory of the related UDMURT, the speakers of Komi once traded in fur all the way from the old city of Perm' to the White Sea and the Arctic Ocean. They were known as *Beormas*, 'Permians', to the Norwegian Ohthere, who described his travels to the English King Alfred around AD 900 (see box at SAMI). Loanwords in Komi are evidence of early contacts with Karelian and Veps dialects of FINNISH. They were also in close contact, across the northern Ural Mountains, with the Khanty and Mansi of the lower Ob valley (see HUNGARIAN),

notably at the old market city of Salekhard where the 'Permians' traded with Russia too.

In the Soviet period there were several very large prison camps in Komi territory. The social problems of the modern Komi Republic come partly from the large number of former camp inmates, Russian-speaking and with a criminal history, who have settled locally after release.

Komi was first written as long ago as the 14th century, when the missionary St Stephen of Perm' devised the Abur alphabet, used for three hundred years. Little remains of early Komi writings, but there is a rich oral literature. The most typical form of Komi traditional poetry is the *berdedchankiv* or lament, clearly related to the *yarabts* sung by the Nenets of the northern tundra.



Komi nouns have seventeen cases and a singular/plural distinction. With possessive suffixes (*pi* ‘son’; *piey* ‘my son’; *piid* ‘your son’; *piis* ‘her son’) the theoretical number of noun forms rises to 136, most of which can actually be heard in everyday speech. The first ten numerals are: *ömi*, *kık*, *kuim*, *nyol’*, *vim*, *kvaim*, *sizim*, *kök’yamis*, *ökmis*, *das*.

**Komi and Udmurt on the map**

Since the Russian Revolution in 1917 there have been two Komi literary languages. *Zyrian* (per-

haps meaning ‘sea people’) is based on the middle Vychegda dialect of Syktyvkar, the capital of the extensive, very sparsely populated Komi Republic. *Permyak* (‘Permian’) is based on the southern dialect of Kudymkar. The Southwest and Trans-Kama dialects of *Udmurt*, now far separated from the rest, have many Bashkir and Tatar loanwords. So has the distinctive Besermen dialect, spoken in the north of the Udmurt Republic by people of Tatar origin who long ago took to speaking the local language.

**At a wedding**

Syoyim da yuvim  
(pasyibe da pomesyibe)  
keres shär shobdise,  
shobdä shär tusyse,  
tusy shär yadrese,  
yadre shär pirogse.  
Syoyim da yuvim  
tsyäskid ireshse,  
kurid vinase,  
tsyäskid yaya shıdse.  
Pasyibe da pomesyibe  
Syoyim da yuvim!

We have eaten and drunk  
(Thanks to our host)  
The best wheat of the hill,  
The best ear of the wheat,  
The best grain of the ear,  
The best cakes of the grain.  
We have eaten and drunk  
Sweet light beer,  
Strong brandy,  
Sweet meat soup.  
Thanks to our host  
We have eaten and drunk!

Once sung by girls at a Komi wedding, this song was collected by Robert Lach, a folklorist who worked with Russian prisoners of war around 1918.

Robert Lach, *Gesänge russischer Kriegsgefangener*, Vol. 1 pt 1:  
*Udmurt and Komi songs* (Vienna, 1926) pp. 59, 115–16

# KONGO

**3,200,000 SPEAKERS**

*Congo (Kinshasa), Angola and Congo (Brazzaville)*

**O**ne of the BANTU LANGUAGES, Kongo attained significance as the state language of the kingdom of Kongo, already the greatest power on the south-western coast of Africa at the time of the early European explorations.

Political units became gradually larger, in the region that is now northern Angola, Congo (Brazzaville) and western Congo (Kinshasa), until, in the 14th and 15th centuries, the substantial and powerful kingdoms of Kongo, Loango and Tio were established. Loango occupied much of the coast between the mouths of the Ogooué and the Zaire, controlling the road from Malebo Pool (Stanley Pool) through a copper-mining region to the coast at Loango city, long a port of call for Dutch traders. Kongo straddled the lower Zaire and held the Atlantic coast and hinterland to the south, where Portuguese ships dominated trade. Inland, from the present sites of Kinshasa and Brazzaville northwards, was the kingdom of Tio, which lasted till the 19th century.

Language spread with government and trade: there was a four-day week with weekly markets. But Kongo became economically dependent on the increasing European demand for slaves, and collapsed in 1665, when the king fell in battle against the Portuguese at Ambwila. The small successor kingdoms eventually succumbed to the European adventurers of the 'Congo Free State' in the late 19th century. After a period of Belgian rule, this has now become an independent state, Congo (Kinshasa).

The first information about the Kongo language to reach Europe was F. Pigafetta's 'Report of the Kingdom of Congo', published in Italian

in 1591, largely based on the reports of the Portuguese mariner Oduardo Lopez, who had visited Luanda in 1578. In 1624 a bilingual catechism in Portuguese and Kongo, *Doutrina christãa*, was published in Lisbon – probably the oldest printed book in any African language. A Latin–Spanish–Kongo dictionary was compiled in manuscript about 1650.

## **The oldest African grammar**

'The year 1659 saw publication in Rome of the first known grammar of a sub-Saharan African language, a 98-page study of Kongo by Giacinto Brusciotto. This Italian Capuchin is known also by the Latin and Portuguese renderings of his name as Hyacinthus Brusciottus a Vetralla and Jacinto Brusciato de Vetrilha. The title of his book is *Regulae quaedam pro difficillimi Congensium idiomatis faciliiori captu ad grammaticae normam redactae* [Some rules compiled in the form of a grammar for the easier understanding of the very difficult language of the Congolese]. Brusciotto recognised and described the system of noun classes, which he termed "principiations", and of concordial agreements. It is significant that his division into "principiations" was determined by the concordial agreements and not by the form of the noun prefix, thus revealing deeper linguistic insight than some of his successors up to the present day!'

Desmond T. Cole, 'The history of African linguistics to 1945' in *Current trends in linguistics* vol. 7 ed. Thomas A. Sebeok (The Hague: Mouton, 1971) pp. 1–29

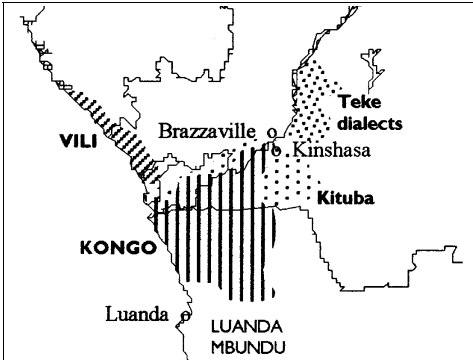
Kongo dialects are spoken on both banks of the lower Congo, west of Kinshasa, and a creole of Kongo origin, Kituba, is used in part of the region as a lingua franca. After massive recent migrations Kongo is spoken by about three-quarters of the vast population of Kinshasa itself. Lingala, however, remains the single most important lingua franca of the city.

For a table of numerals see LINGALA. Examples of Kongo verb forms are given at KITUBA.

**Kongo: languages and kingdoms**

*Kongo* was once the language of the Kingdom of Kongo. *Kituba*, a creole based on Kongo, initially spread with Kongo trade through regions not originally Kongo-speaking.

*Vili* is now the usual name of the language of coastal Congo (Brazzaville) (the former Kingdom of Loango) and its speakers. It once meant ‘caravan



trader’, evidence of the centrality of long distance trade to this pre-colonial state. *Luanda Mbundu* was the language of the southern region of the old Kingdom of Kongo (see map at MBUNDU).

*Teke* is a group of dialects, with about 300,000 speakers, whose present location helps to show the old extent of the kingdom of Tio.

# KONKANI

1,500,000 SPEAKERS

*India*

**K**onkani is the southernmost of the contiguous INDO-ARYAN LANGUAGES. Its chief centre is the old Portuguese possession, Goa (an Indian state since 1987) – and this is the clue to the separate identity of Konkani. Linguistically close to the Marathi of the state of Maharashtra, it is a language of the western coastal strip of India: not of the whole of it, but of the northern third – coastal Maharashtra – spreading southwards to many large and small harbour towns. It grew up alongside the Portuguese development of Indian coastal traffic, and is now in decline as the reason for its separateness has disappeared.

*Konkan* is the coastal strip between the Western Ghats of Maharashtra and the Indian Ocean, running southwards as far as Goa. *Konkaṇī* (Portuguese *Concani*, a name already used in the oldest, 16th-century publications in the language) is thus the speech of this coastline: later, more specifically, the speech of Goa and its daughter settlements, when that came to be seen as a distinct language.

From south to north (see map at MARATHI) eight Konkani dialects can be identified along the coast of Kerala, where the surrounding language is MALAYALAM. Of these the major one is at the long term Portuguese colony of Cochin. On the coast of Mysore, whose majority language is KANNADA, Mangalore has a large Konkani community. The territory of Goa itself is the largest enclave of all. From Kerala northwards to Goa, then, Konkani is surrounded by Dravidian speech. Onwards again, the Konkani of coastal Maharashtra gradually – north of Ratnagiri –

shades into the Konkani dialects of Marathi.

Goa was captured by the Portuguese in 1510, becoming the metropolis of their Eastern possessions, the seat of a Viceroy and of an archbishop. Konkani literature begins in the 16th century, as early as the language itself. It is Christian in inspiration, with religious texts, Bible translations and grammars. Konkani served the Portuguese as a missionary language, while administration was carried on in Portuguese. Traditionally Konkani speakers – here and in East Africa, where many are settled – have had a knowledge of Latin, the religious language of Catholic Christianity. Latin is still used in some religious rituals, though worshippers in general no longer understand it.

Recently there has been an attempted revival in writing in modern Konkani dialects, alongside an effort to create a unified literary language. For all that, most Konkani speakers consider Marathi, Kannada and Malayalam to be their literary languages.

In keeping with its multicultural context, Konkani has been written in four scripts: Latin, Devanagari, Malayalam and Kannada (now the commonest). The Latin orthography, based on Portuguese, was devised in the 16th century and is still sometimes used. Retroflex consonant sounds are marked by a double letter (in Portuguese a doubled *r* has this same meaning): this leaves double length consonant sounds to be marked with a hyphen, rather as they are in Catalan. There is no indication of vowel length.

The first ten numerals in Konkani are: *yek, don, tīn, cār, pānz, sa, sāt, āt, nou, dhā*.

# KOREAN

**63,000,000 SPEAKERS**

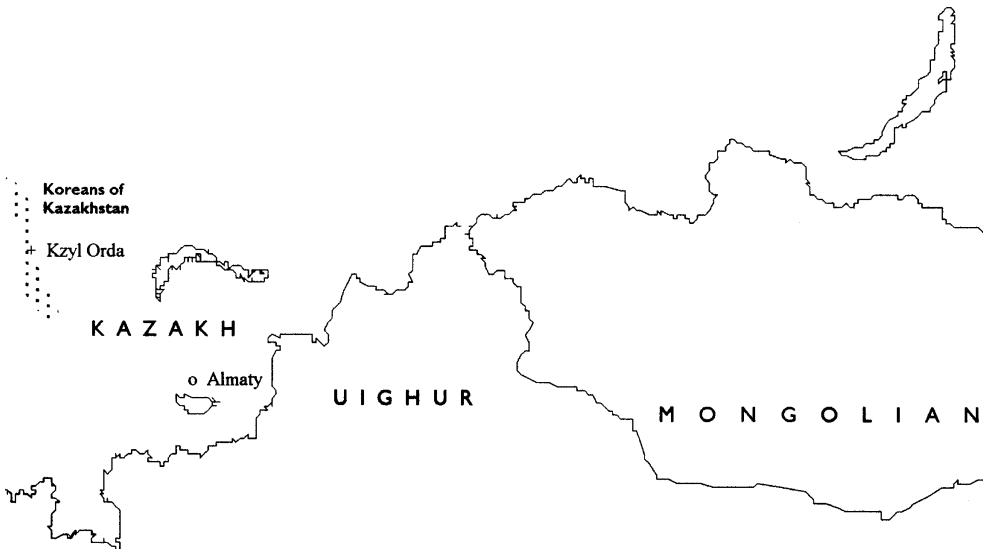
*South Korea, North Korea,  
China, Japan, Uzbekistan, Kazakhstan, Russia*

**K**orean is thought by some linguists to be a member of the ALTAIC family of languages. If so, it must have separated from the remainder of the family many thousands of years ago. It is now the national language of North and South Korea. There are considerable Korean minorities in Manchuria and (as a result of recent migration) in Japan and central Asia.

Before the 15th century, when Korean was first written phonetically, little in the history of

the language is certain. Scholars generally suppose that an early form of Korean was already spoken in Korea and parts of Manchuria two thousand years ago. The small states in the region, often subject to China, were beginning to merge. In the 7th century the southern kingdom of Silla succeeded in unifying the peninsula, and Korean apparently descends from the old 'Han' dialect spoken in Silla.

Since then, Korean has undergone strong po-



**Northern Asia: the Korean diaspora**



litical, religious and literary influence from China. This is reflected in huge numbers of Chinese loanwords in the language. They have been customarily written in Chinese characters, which thus appear interspersed with Korean script in every older text.

*Koryŏ* (whence *Korea*) and *Chosŏn* (which, as *Chosen*, is the Japanese name for Korea) are in origin the names of two successive dynasties that ruled Korea. The *Koryŏ* capital, Songdo, is near modern Seoul.

A collection of ancient songs is dated to the Silla period, which ended in 935 when the *Koryŏ* began to rule the country. The literature of *Koryŏ* included new genres, notably the dramatic poetry that celebrated the Buddha and the local gods of Korea. *Sijo*, a poetic form consisting of three-line stanzas, is traced to the 12th century and is still current. Narrative prose first became popular in the 17th century.

Korea fell under Japanese rule from 1910 to

1945. With the ensuing division of the country between regimes of opposing political philosophies, the written languages of North and South have diverged. In the North the Korean alphabet is mandatory. Chinese characters have not been used officially since 1949, and Chinese loanwords are officially discouraged – yet many remain, and expressions modelled on Chinese ('calques') permeate writing and political discourse. In the South, Chinese characters continue to be used, though less than before. Schoolchildren still need to learn nearly two thousand of them. Meanwhile American English influence on the language increases.

The first ten numerals in Korean are: *hana, tul, set, net, tasŏt, yŏsŏt, ilgop, yŏdŏl, ahop, yŏl*.

### Korea and the Korean diaspora

Korean emigration is one of the less-known diasporas of recent history. Already extending across the Korean borders into Manchuria, Korean farmers began to settle the Pacific coast of far eastern Russia after 1863. From 1924 other pioneers were encouraged to colonise parts of Uzbekistan and Kazakhstan. With a sudden change of Soviet policy under Stalin, the Pacific coast settlers were deported en masse to central Asia in 1937. During the Second World War, Koreans were recruited to settle and farm the island of Sakhalin, then under Japanese rule.

Today substantial Korean minorities are to be found. There are 2,000,000 in Chinese Manchuria, and Korean counts as one of the 'major nationalities' of China. Recent migration to Japan has resulted in a minority of 700,000 there. The 100,000 *Korŏi* of northern Sakhalin are now under Russian rule. There are 200,000 Koreans in Uzbekistan, most of them at the Amu Darya delta, and 150,000 in Kazakhstan, concentrated near Tashkent, Kuygan, Taldy-Kurgan and Kzyl Orda. At Kzyl Orda there is a Korean newspaper, radio station and theatre. Many of these Korean farmers developed desert lands with water drawn from the Amu Darya and Syr Darya, and are now threatened by their own success, which is leading to the drying up of the Aral Sea.



## The Korean alphabet

k kk n t tt l m p pp s ss ng ch tch ch' k' t' p' h

ɪ ʌ ʔ ʌ ɪ ɪ ɪ ɪ ʊ ʊ ʊ ʊ ʊ ʊ ʊ ʊ ʊ ʊ ʊ ʊ  
 a æ ya yæ ø e yō ye o wa wæ oe yo u wō we wi yu ũ ü i

## Hangŭl – Chosŏngŭl

All writing in Korea was in the Chinese script – whether the author was writing Korean, or whether the Chinese language itself, culturally and politically dominant, was being recorded – until 1443/4. That was when King Sejong, of the Chosŏn dynasty, invented the remarkable alphabet now known to southerners as *Hangŭl* ‘Korean writing’, to northerners as *Chosŏngŭl*.

The Korean alphabet did not emerge from Sejong's unaided imagination, as is sometimes implied. Compare the letter forms with TIBETAN (and an old Mongolian script) and similarities begin to appear. Two further strands of inspiration can be identified. First, a careful analysis was made of the sounds of the language, in the Indian (Sanskrit) tradition transmitted via nor-

thern Buddhism. Secondly, the use of an independent character block (for example, 발 *pal* 'foot'), usually a visible combination of more than one letter, is typical of Chinese even more than of Indic scripts: and the character blocks stand well alongside true Chinese characters, as they need to do in written Korean. But the result – a phonemic script in which each sound is represented by a single, always identifiable letter form – was indeed remarkable, and has been rightly hailed as one of the most scientifically designed and efficient scripts in the world.

The alphabetical order of the vowels given in the box is the one that is standard in North Korea. The transcription is a basic form of the most widely accepted standard, called ‘McCune-Reischauer’.

# KRIO

## 350,000 SPEAKERS OF KRIO AS A MOTHER TONGUE

### *Sierra Leone, the Gambia*

One of the ENGLISH CREOLES AND PIDGINS, Krio ('Creole') is the first language of many inhabitants of Freetown and the lingua franca of as many as three million people in Sierra Leone. A variant form, called *Krio*, *Aku* or *Patois*, is spoken in the Gambia. There, however, a pidginised form of WOLOF is in the ascendant.

Was Krio first introduced to Sierra Leone by the freed slaves from the Caribbean who were resettled here in the 18th century? Or did their speech merge with an already existing creole stemming directly from the English West African pidgin? Krio has strong links with the Caribbean creoles. Its grammar in many ways resembles West African languages: some researchers emphasise its similarity to YORUBA (of Nigeria), others to varieties of AKAN; but neither of these languages belongs to Sierra Leone itself.

The early 19th-century Krio speakers of Sierra Leone were evidently a mixed group. Akan was the language spoken when early English trade had concentrated on the Gold Coast (modern Ghana): it thus influenced the pidgin, and all the creoles that stem from it, directly. It contributed to the oldest layer of Krio. The new linguistic influences that came to Sierra Leone with the migration of freed slaves in the early 19th century were themselves complicated: while migrants of older Car-

ibbean background already had an English creole as their mother tongue, many of those of recent African origin had certainly come from Nigeria, to which some, such as the linguist Samuel Crowther (see box at YORUBA), would return.

#### English dialects and the Atlantic creoles

'In the sixteenth century,' Ian Hancock observed, 'the crews of English ships would have included men speaking a gamut of widely differing English dialects. In Sierra Leone Krio there are preserved dialect forms from as far south as Cornwall and as far north as Scotland. Krio words such as *fitrí* (Yorkshire *fittery* "with legs akimbo") and *gánga* (Yorkshire *ganger* "little girl's dress") are still used in Yorkshire today and are quite unintelligible to a modern Devonian or Cornishman. Given also the admixture of non-English speakers in the crews, and the long history of specialized nautical lingo, the form of English heard first by West Africans must have been itself a somewhat flexible compromise'

Hancock in *Pidginization and creolization of languages* ed. Dell Hymes (Cambridge: Cambridge University Press, 1971) p. 290).

#### The serial verbs of Krio

Typical of Krio is the use of English verbs to mark direction, and the development of additional auxiliary verbs:

A bring di kasada **kam** na os

A **bin tek** di buk **go** na skul

Wi **go tray** fo push di trak

I **don was** di klos

I brought the cassava **to** the house

I **took** the book **to** school

We **will try** to push the truck

He **has washed** the clothes

After I. T. Givón in *Readings in creole studies* ed. Ian F. Hancock (Gent: Story-Scientia, 1979) pp. 16–18

# KRU LANGUAGES

**PERHAPS 2,000,000 SPEAKERS**

*Ivory Coast, Liberia, Sierra Leone*

**T**he KRU LANGUAGES are spoken in the forests of southern Liberia and Ivory Coast. They belong to the NIGER-CONGO family.

Although none has as many as a million speakers individually, Kru languages remain important on a regional scale; many speakers now know some English or French. Historically, Kru languages are crucial for their position at the crossroads of African-European interaction.

Oral tradition suggests that speakers of these languages migrated southwards in medieval times from an earlier habitat well inland. What is certain is that Kru languages (often called *Kra*, *Krumen*) were among the first to be encountered by European voyagers on what was then known as the Pepper Coast, centre of the production and export of Guinea pepper or melegueta pepper, once a staple of the African seaborne trade. The oldest Kra wordlist is in a French manuscript of the 1540s, where it is called *langaige de Guynee* 'language of Guinea'.

Bété, of the Ivory Coast, has the largest number of speakers of any Kru language today – yet it is not an official language and seldom appears in print. Bété has borrowed heavily from French, the ruling language of Ivory Coast, in which many speakers are now bilingual. It has also borrowed from English, long a major language of the coastal trade: *bogu* 'book'; *copu* 'glass' from English *cup*. The first ten numerals in Bété are: *bolo*, *so*, *ta*, *mona*, *gbi*, *gbopolo*, *gbiso*, *gbota*, *gbomona*, *kuba*.

Kru people themselves worked on European ships as early as the 16th century, and there are still Klao, Grebo and Bassa-speaking communities in several West African port cities as

a lasting sign of their participation in long distance sea travel.

## Major Kru languages

*Bassa* has 300,000 speakers in Liberia.

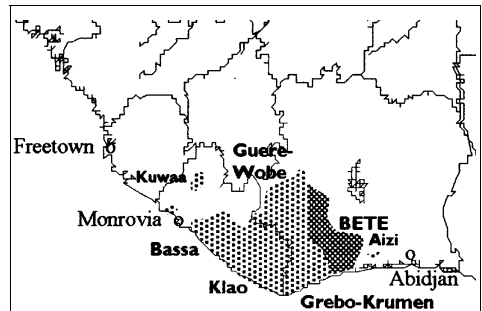
*Bété* dialects are spoken by 700,000 people in central and western Ivory Coast. *Dida*, with about 150,000 speakers, is sometimes regarded as a dialect of Bété.

*Grebo* of eastern Liberia, and *Krumen* or *Kroumen* of western Ivory Coast, form a dialect group with a total of perhaps 250,000 speakers.

*Guéré* (with *Wobe*) of Ivory Coast, and *Krahn* of Liberia, form a similar dialect group, with as many as 450,000 speakers.

*Klao* or *Kru* has its territorial base in eastern Liberia, where there are 150,000 speakers.

The Kru or *Kra* language communities in West African port cities such as Freetown, Lagos and Accra have diverse dialect origins but their lingua franca is generally closest to *Klao* and *Bassa*.



**West Africa: Kru languages**

# KUKI-CHIN AND NAGA LANGUAGES

**PERHAPS 2,000,000 SPEAKERS IN TOTAL**

*India, Burma*

In the Chin Hills of western Burma, and the Indian states of Mizoram and south-eastern Assam, a group of SINO-TIBETAN LANGUAGES is spoken. Their fragmentation – most have only between 10,000 and 50,000 speakers – has three causes: the broken landscape, the absence until very recently of large-scale political units, and the fact that the speakers have evidently been settled over most of their present habitat for many centuries. Thus early dialects gradually differentiated until they reached the point of mutual unintelligibility.

There has, however, been a gradual movement westwards. The name of the Chindwin river, the greatest Irrawaddy tributary, means literally ‘Hole of the Chins’; yet the deep Chindwin valley now lies well to the east of almost all Chin populations.

Indian and Burmese rulers tended, until the British advances in the 19th century, to leave the Chin Hills alone. On the Indian side of the hills the inhabitants were grouped under the name *Kuki*; on the Burmese side the usual name was *Chin* (spelt *khyang* in Burmese, and meaning ‘friend’). In older English sources the forms *Kookie* and *Khyeng* can be found.

The speakers gave to their own communities and languages names that often included the element *-zo*: among them were the *Mizo*, known to others as *Lushei*, who have given their name to an Indian state. *Zo* is said to mean ‘wild’ or ‘independent’, by contrast with *vai* ‘civilised’ or ‘ruled’.

Lushei was almost the only exception in this region to the rule of linguistic fragmentation. Here, approaching the ‘Mizo Hills’ from Assam, explorers and British administrators – who established control in 1890 – found a language that had become a local lingua franca of inter-community travel and trade. Lushei, with about 350,000 speakers, is now the official language of the Indian state of Mizoram, established, after a long struggle, in 1987. It has a Christian majority.

North of Mizoram are a series of tribes known as Old Kuki and New Kuki – because they form two separate ethnic and linguistic groupings and New Kuki is the later arrival. Old Kuki languages are all spoken by very small communities. The major New Kuki languages are Thado and Paite, both closely related to Tiddim, largest of the languages of the northern Chin Hills in Burma. Somewhat further south in Burma is Falam or Hallam, and this is close linguistically to the Old Kuki languages of India.

On both sides of the Chin Hills, the gradual advance of ‘pacification’, followed by ‘administration’, was scarcely complete when the Japanese occupation of Burma foreshadowed the end of British rule. The Burmese side became the Chin Special Division of independent Burma; on the Indian side the state of Mizoram was established, other Kuki peoples being divided among Assam, Manipur and Nagaland.

Although political relations were minimal, local trade routes have long criss-crossed the Chin Hills. Kuki-Chin languages show the in-

fluence both of Burmese and of Assamese (and to some extent of Bengali).

### The Naga languages

North of the Chin languages extend two further groups of related languages, spoken in Nagaland, Manipur and eastern Assam. They are grouped by linguists as 'Naga' and 'Konyak'. Here, too, little political organisation above the village level existed until recently. Naga peoples did pay tribute to the AHOM kingdom of Assam, and in some cases to the king of Manipur. British rule, at least nominal, came in 1881. The Indian state of Nagaland was created in 1963 after a bitter struggle for Naga independence. The multiplicity of local languages has led to the increasing use of *Naga Pidgin* or *Nagamese*, a creolised form of ASSAMESE, which developed at least 150 years ago as a result of local trade patterns. It is now the lingua franca of Nagaland.

#### The introduction of writing: two perspectives

'Why did Naga languages have no traditional script or writing? Speakers explained that a god once gave them writing, on skins, but a dog ate the skins and since then they had not been able to write.'

J.P. Mills, *The Rengma Nagas*  
(London: Macmillan, 1937) p. 286

'With the suppression of headhunting and the establishment of law and order by the British Raj – followed almost immediately by the arrival of the late Rev. F. W. Savidge and myself as Christian missionaries [1892] – a new day dawned upon the Lushai Hills . . . It fell to our lot to reduce the language to writing in such a way that our system could be readily adopted by the people themselves. For this purpose we chose the simple Roman script, with a phonetic form of spelling . . . still used throughout the tribe.'

James Herbert Lorrain (Pu Buanga), *Dictionary of the Lushai language* (Calcutta: Royal Asiatic Society of Bengal, 1940) p. vii

Among these languages there is no indigenous tradition of writing. But Naga languages have a wealth of folk tales and oral poetry; while Lushei is rich in similes. Events that happen 'in turn' are *a tui kang nghâkin* – 'like girls at a drying-up spring'.

### The Kuki-Naga language group

*Southern Kuki-Chin* includes Khumi, Matu, Mün and other smaller languages.

*Central Kuki-Chin* includes Lushei or Lushai or Mizo (350,000 speakers), Haka and Hmar.

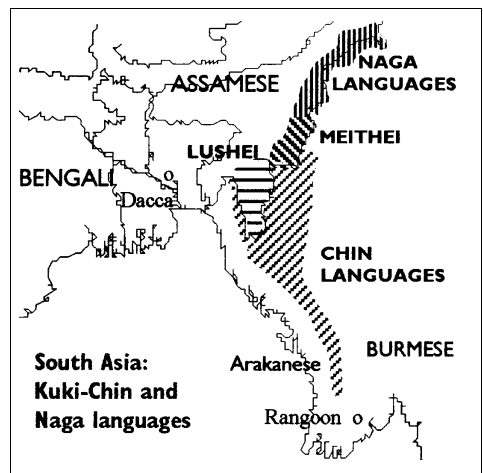
*Old Kuki* includes Falam (100,000 speakers) and many smaller languages.

*Northern Kuki-Chin* includes Thado (150,000 speakers), Tiddim (100,000), Paite and Zome.

The *Naga* languages include Angami, Lotha, Mao and Sema, the largest, with over 60,000 speakers.

The *Konyak* languages include Konyak itself, with nearly 100,000 speakers, Ao and Tangkhul. Throughout their territory the Ao people are actually divided into two subgroups, speaking very different dialects, Chongli and Mon-gsen.

Also part of the Kuki-Naga group are the *Mikir-Meithei* languages, Mikir (200,000 speakers) and MANIPURI, which has a separate entry.



## Numerals in Kuki-Naga languages

	<b>Lushei</b>	<b>Falam</b>	<b>Thado</b>	<b>Angami</b>	<b>Tangkhul</b>	<b>Mikir</b>
1	-khat	-khāt	-khat	po	kha	sī
2	-nhih	-ni	-ni	nā	ni	nī
3	-thum	-thūm	-thūm	sē	thum	thām
4	-lī	-li	li	dā	li	lī
5	-ngā	-ngā	-ngā	ngu	ngā	ngō
6	-ruk	-ruk	-gūp	ru	ruk	rāk
7	-sarih	-rī	-sagi	nā	shini	rāksī
8	-riat	-riet	-gēt	thā	shat	nākep
9	-kuā	-kūok	-kū	kū	ko	sākep
10	shom	sām	-som	kō	tharā	kep

# KUMAUNI

1,200,000 SPEAKERS

*India*

**K**umauni belongs to the Pahari group of INDO-ARYAN LANGUAGES (see map there), like Nepali and Garhwali, its neighbours on either side.

*Kumaūn* was once the small area around Champāwat: the local dialect here is still called *Kumaīyyā*. The land dominated by the ‘fortress of Kumaūn’, first mentioned by the 12th-century poet Chand Baradai, gradually expanded until under British rule the principality stretched from the Tarai foothills to the heights of the Himalaya.

In the mountain passes and the higher valleys the *Bhoṭiā* people speak Tibetan dialects. In the lowlands, recent migrations have led to the use of Hindi as a lingua franca. Between the two, Kumauni remains the everyday language of more than a million people. Beside Hindi as the language of education and government, and the local dialects of which as many as fifteen have been counted, there is an identifiable Standard Kumauni spoken by educated classes over most of the former state, deriving from the dialect of the old capital at Champawat.

There are inscriptions, erected by local rulers, dating from the 17th century. Since 1800 a small body of written poetry in Kumauni has dev-

eloped, but the language is especially important for its oral literature, which includes legends, romances and shorter ritual songs.

जो औरो॑ उपर खाड़ खनछ॑ | ऊ आफी वीमी पड़ैछ॑ ॥  
Jo aurom upar khāṛ khanch, ū āphī vī-mī  
paṛamch.  
One who digs a pit for others will fall into it  
himself.

Numerals in Nepali, Kumauni and Garhwali		
Nepali		Kumauni and Garhwali
ek	1	ek
dui	2	dvī
tin	3	tīn
chār	4	cār
pāñch	5	pām̐c
chha	6	chai
sāt	7	sāt
āṭh	8	āṭh
nau	9	nau
das	10	das



# KUMYK

200,000 SPEAKERS

*Russia*

One of the TURKIC LANGUAGES of the Kipchak group, Kumyk is spoken by the fourth largest ethnic unit of the multilingual Russian republic of Dagestan. There are no historical records of their origins, but the speakers of *Qumuq* or Kumyk appear to have been established here, between the northern Caucasus mountains and the Caspian Sea, at least since the 11th and 12th centuries when the region was conquered for Islam. They can clearly trace their history to the empires of the steppes, like the Tatars, and thus have quite a different origin from that of the AZERI speakers to their south.

In recent centuries, however, Kumyk has been strongly influenced by the culturally dominant Azeri and by the neighbouring Caucasian language DARGWA. Long threatened by Russian expansion, the region was conquered by Russia in 1867. The coastal plains of central Dagestan, once sparsely populated by Kumyk, were resettled with Avar and Dargwa speakers in the 1950s to 1970s – a knock-on effect of the return of mass exiles, such as Chechen, Karachai and Balkar speakers, to their homelands.

The Kumyk capital and cultural centre is Buinaksk (Temir Khan Shura). Oral and traditional poetry was collected in the 19th century. The founder of Kumyk literature, it is usually said, is Yırçı Qazaq, born in 1839, who was exiled to cold Russia as punishment for an illicit love affair and from there wrote a lyrical poetic letter to his beloved Rayhānat, begging her father's forgiveness.

There is now some publishing in Kumyk: the first magazines and newspapers appeared in 1917–18. Arabic script was used until 1928,

the Latin alphabet for the next ten years; the Cyrillic alphabet was introduced in 1938. Russian is increasingly the language of culture, education and politics in modern Dagestan.

Two other small Turkic language communities of the northern Caucasus can be grouped with Kumyk. Both, like Kumyk itself, have been written in three successive alphabets since the Russian Revolution.

## Karachai and Balkar: 'punished peoples'

Karachai and Balkar are usually regarded as a single language, though the speakers see themselves as distinct peoples. They have been settled, side by side, in the northern Caucasus at least since the 14th century, and began to adopt Islam, under Nogai and Crimean Tatar influence, in the 18th. Russia conquered the region in 1827–8. One phonetic feature easily distinguishes the two dialects: the Karachai say *ch* when the Balkar say *ts* (Karachai *küchük*, Balkar *kütsük* 'small').

Speakers of Karachai were among the Muslims of the northern Caucasus who incurred Stalin's distrust. Soon after the German occupation of 1942, the whole population, numbering 75,000, was deported from the Caucasus to the Soviet East in 1944. A quarter of them died on the journey or in the first five years of exile. The survivors and their children returned in 1957, and Karachai speakers now number 160,000. The former republic of Karachaevo-Cherkessia has now been divided, and the Karachai have self-governing status within Russia.

The 43,000 speakers of Balkar were also dispatched eastwards en masse in 1944, and many died. The remnant were allowed to return in 1956 and their descendants now number 85,000. Once sharing the Kabardino-Balkar autonomous region, Balkar speakers now have a self-governing republic within Russia.



# KURDISH

**PERHAPS 14,000,000 SPEAKERS**

*Iraq, Turkey, Iran, Syria, Armenia*

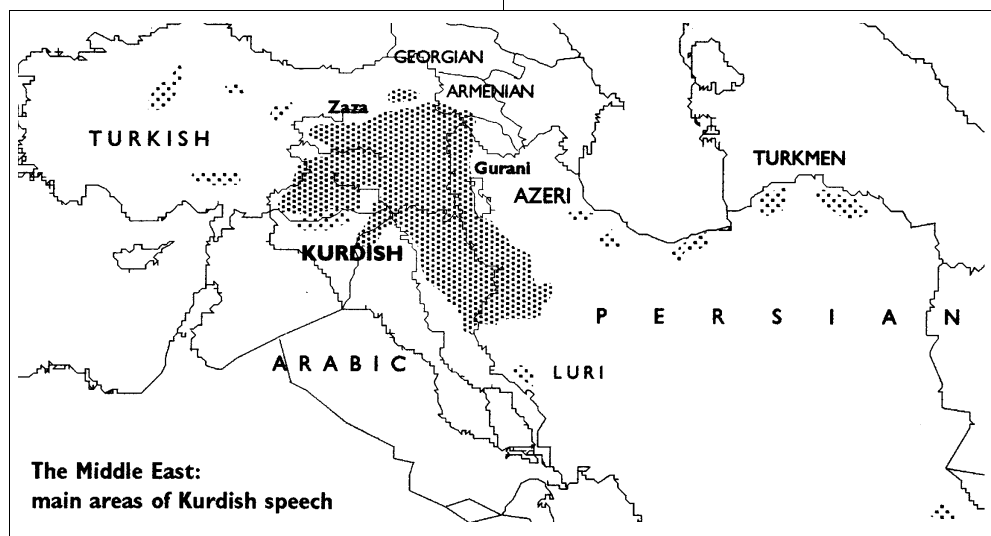
**K**urdish is the second largest of the IRANIAN LANGUAGES in terms of population – but its political position is wholly different from that of Persian, the largest of the group. While Persian is the national language of Iran and Tajikistan, Kurdish language and culture struggle for recognition in all the states in which its speakers live.

Kurds are possibly first named in surviving records of about 400 BC, as the *Kardoukhai* encountered by Greek mercenary soldiers, led by Xenophon, on their march from Mesopotamia to the Black Sea – or possibly four hundred years later, as the *Kyrtioi*, a Median population, listed by the Roman geographer Strabo. At any rate, Kurds occur frequently in historical sources from the time of the Islamic conquest onwards. Redoubtable mountaineers with a nomadic lifestyle, they were independent most of

the time. At the end of the First World War, at the Lausanne Conference, Kurdish territory was parcelled out to Turkey, Iran and Iraq, with a promise of autonomy only for the Iraqi section.

Kurds have been authors in Arabic, Persian and Turkish for about a thousand years, while Kurdish written poetry – in Arabic script – goes back to the 15th century. The classic text is the epic *Memozîn* by Ehmedî Khanî (1650–1706). Kurdish oral literature is remarkably powerful and varied.

In Iraq Kurdish has in general been recognised as a language of education and the press. The struggle for political autonomy has, however, led to bitter warfare recently. In Turkey Kurdish is outlawed in official use and in the press: even tape recordings are not allowed.



**Kurdish, Gurani and Zaza**

The major geographical division within Kurdish is between Northern and Central dialects.

The principal Northern dialect is *Kurmanji*. This is the Kurdish of eastern Turkey, north-eastern Syria and Iranian Azerbaijan. It is also spoken, as a result of recent migrations, in Khorasan, Baluchistan and Afghanistan. It is the usual speech of Kurdish communities in Middle Eastern cities, from Beirut to Istanbul. North-eastern Kurdish includes the *Badinani* of the Mosul region of north-west Iraq. To the same group belongs the dialect spoken in Armenia and independent Azerbaijan. After forced migrations from Soviet Transcaucasia during the Second World War, this dialect is also spoken in Kyrgyzstan and Kazakhstan.

The principal Central dialects are the *Sorani* of Iraqi Kurdistan and the closely related *Kurdi* and *Sine'i* of Iranian Kurdistan.

To the south there are more scattered Kurdish-speaking groups in Luristan, including the Lak tribes.

*Gurani* and *Zaza* are two quite separate Iranian languages, spoken by as many as two million people (there are no reliable figures) in Iran and Turkey. They consider themselves, and are often considered by outsiders, to be Kurds. A considerable number of the 'Turkish workers' living in several countries of western Europe are

Kurds – and a high proportion of these are actually *Zaza* speakers. The *Bajalani*, who speak a Gurani dialect, live in northern Iraq and form an esoteric Islamic sect with a special devotion to the Imam 'Ali, 'black 'Ali' as they call him. Gurani has an established tradition of oral poetry, once performed at princely courts.

Numerals in Kurdish and LURI		
Kurdish		Luri
yek	1	yak
du	2	du
sê	3	se
çar	4	châr
pênc	5	panch
şeş	6	shish
heft	7	haft
heşt	8	hash
neh	9	nuh
deh	10	dah

**Kurdish in writing**

Kurdish as a literary language can be found in three different scripts. The Latin alphabet now used for Kurmanji is similar to that of Turkish. An additional character *ë* is sometimes used.

The Kurdish alphabet: Latin, Arabic and Cyrillic scripts																			
A	B	C	Ç	D	E	Ê	F	G	H	Î	J	K	L	M	N	O	P	Q	R
a	b	c	ç	d	e	ê	f	g	h	î	j	k	l	m	n	o	p	q	r
ز	ی	خ	و	و	و	و	و	و	و	و	و	و	و	و	و	و	و	و	و
А	Б	В	Г	Д	Е	Ё	Ж	З	И	Й	К	Л	М	Н	О	П	Р	С	Т
а	б	в	г	д	е	ё	ж	з	и	й	к	л	м	н	о	п	р	с	т

Until the 20th century Kurmanji was written in the Arabic alphabet. Arabic script is still used for publications in Badinani, Sorani and Sine'i. Arabic script is read from right to left, but in the box the letters are given from left

to right, corresponding to the latin script above.  
A form of the Cyrillic alphabet was assigned to Kurdish speakers in the Soviet Union in 1945. The box shows it in Latin alphabetical order.

# KURUKH

1,200,000 SPEAKERS

*India*

**K**urukh is spoken well to the north-east of the main body of DRAVIDIAN LANGUAGES, on the borders of the Indian states of Orissa and Madhya Pradesh, extending as far as the southern edge of Bihar.

*Kurukh* is the speakers' own name for their language. Outsiders spell it in various ways – *Kurux* and even *Khurūk* will be found – or may use the alternative terms *Orāon*, *Urāon*.

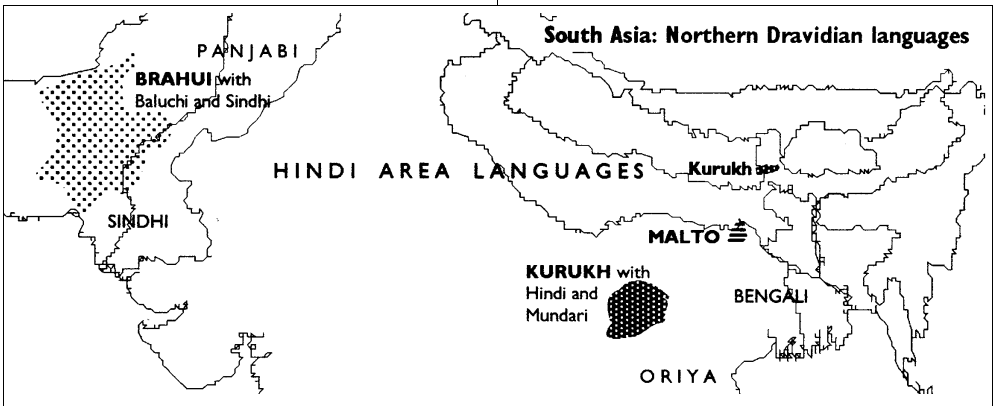
Like other northern and central Dravidian languages it is spoken by hill and jungle dwellers and has no written tradition. Kurukh speakers who settle in the plains turn to other languages of their neighbourhood, such as Oriya, Bengali or Magahi. The number of speakers has, however, remained constant or even increased over the past decades.

The first ten numerals in Kurukh are: *onṭā*, *ēm̐r*, *münd*, *nākh*, *pancē*, *soyyē*, *sattē*, *aṭṭhē*, *naiṃyē*, *dassē*. Only '1' to '4' are Dravidian: the rest are borrowed from Magahi or a related dialect.

## The northern Dravidian languages on the map

*Kurukh* is spoken from the Sambalpur district of eastern Orissa northwards to the Chota Nagpur region. Its speakers have a tradition that they originated in the hills north of Patna. A small number of speakers of a related dialect has in fact been identified in the Nepalese Terai: it is called *Dhangar* or *Nepali Kurukh*. The same tradition speaks of a party who went further east, to the mountains of Rajmahal. These will have been the linguistic ancestors of the 50,000 speakers of *Malto*. It is sometimes called *Rajmahalia* or *Malpaharia* – a name also given to the local creolised Bengali which appears to be replacing Malto on the lips of many.

BRAHUI, completely isolated from the rest of the Dravidian family, is spoken in western Pakistan, and is heavily influenced by the Iranian languages that surround it.



# KYRGYZ

1,500,000 SPEAKERS

*Kyrgyzstan, China*

**S**peakers of one of the TURKIC LANGUAGES, which is now the national language of Kyrgyzstan, the *Kyrgyz* (or Kirghiz or Qırız) burst into history as the opponents, and eventual destroyers, of the old Uighur empire in western Mongolia, which fell in 842 AD.

They did not found an empire of their own in its place. At that time the Kyrgyz were apparently forest dwellers of the upper Yenisei valley: to their north, an Arabic source reported, were ‘frozen desert lands’. It was probably some time later that they began to migrate south-westwards to their present location in the Tien Shan mountains (see map at KAZAKH). Here, after subjection to successive transitory central Asian empires, and after becoming tributary to China in the mid 18th century, they were conquered by Russia in the 19th. Kyrgyzstan was one of the republics of

the Soviet Union. It became independent at the end of 1991.

The first Kyrgyz speaker to visit Europe was probably a slave-girl, θεράπεινα δωριάλωτος ἐκ τῶν λεγομένων Χερχίρ, ‘a prisoner of war from the people called *Kherkhir*’, given to the Greek envoy Zemarchus in 569 AD on his visit to the Türk emperor Ishtemi (Menander the Guardsman, *History*, fragment 10.3; see also box at TURKIC LANGUAGES).

Kyrgyz is the vehicle of magnificent epic poetry.

Written Kyrgyz literature, at first in Arabic script, began to appear only around 1910. With changes in Soviet policy, a Latin alphabet, introduced in 1928, was abandoned in favour of Cyrillic in 1940.

Altın iyerning kashı eken:  
ata jurtnung bashı eken.  
kümüsh iyerning kashı eken:  
tün tüshkön kaling köp Nogay  
jurtnung bashı eken.

A golden saddle has its pommel:  
a people has its chieftain.  
A silver saddle has its pommel:  
the Nogay teeming as shadows  
at nightfall have their chieftain.

*The memorial feast for Kökötöy-khan: a Kirghiz epic poem* ed. and tr.  
A. T. Hatto (London: Oxford University Press, 1977) lines 1–4

## Kyrgyz epic

The *Memorial Feast* for Kökötöy-khan (see box) is ‘the earliest and one of the best of a group of ten oral heroic poems recorded among the northern Kirgiz tribes between 1856 and 1869’ (A. T. Hatto). Variant versions were collected from oral poets about seventy-five years later. It was first published by V. V. Radlov in 1885. Radlov supplied a rough German translation and a study

of the nature of improvisation in oral poetry, making interesting comparisons with the ancient Greek epics attributed to Homer. Later, an edition with a Russian translation was published by the original collector, the Kyrgyz nobleman Chokan Valikhanov, in 1904. These editions were abridged. The full original text had been noted down in Arabic script for Valikhanov. Long believed lost, it was rediscovered in a Leningrad manuscript in 1964.

### Numerals in Kazakh and Kyrgyz

Kazakh (Latin transcription)	Kyrgyz (Cyrillic script)	Kyrgyz (Latin transcription)
bir	бир	bir
eki	эки	eki
üsh	уч	üch
tört	төрт	tört
bes	беш	bes
altı	алты	altı
zheti	жети	zheti
segiz	сегиз	segiz
toghız	тогуз	toguz
on	он	on

from Kurtulus Öztopçu and others, *Dictionary of the Turkic languages* (London: Routledge, 1996)

# LAHU

500,000 SPEAKERS

*China and south-east Asia*

One of the Burmese-Lolo group of SINO-TIBETAN LANGUAGES, Lahu is spoken by a high mountain population in south-western Yunnan, between the Salween and Mekong rivers (see map at BURMESE).

In older sources the *Lahu* are often called *Muhsö*, originally a Tai term, ‘hunter’.

Like their neighbours who speak Wa, the Lahu have been among the chief producers of opium for the world market. They are divided into subgroups with different costumes and somewhat different customs, Black Lahu (*Lahu na*), Yellow Lahu (*Lahu shi*) and others. Each has its distinct dialect.

There is some evidence of a once-autonomous state of the Black Lahu, ruled by thirty-six *fu* or priests, north of Xishuangbanna and east of the Wa country in western Yunnan. If this ever existed, it succumbed to Chinese rule at the

end of the 19th century. It would at least help to explain why a ‘Black Lahu’ dialect, *Meuneu*, remains a kind of standard language and has served as a lingua franca of local trade and communication in the Mekong valley and the mountains to the north of the Golden Triangle.

It is certainly true that religious leaders, *paw-ku*, exert great power among Lahu communities, sometimes claiming apotheosis and inspiring holy war. Baptist missionaries have worked for many years among the Lahu, devising an orthography for the language in Latin script and publishing religious texts.

New Year is the time for courtship among the Lahu and AKHA. Young men are always on the move, visiting neighbouring villages. Boys and girls camp out, close together but in separate groups, around bonfires, and sing alternate verses of traditional love songs.

### Numerals in Loloish languages

	Akha	Black Lahu	Lisu	Yi
1	ti <sup>ˆ</sup> , ti <sup>˘</sup>	te <sup>˘</sup> ma <sub>u</sub>	hti <sup>˘</sup>	t'a <sup>˘</sup>
2	nyi <sup>ˆ</sup> , nyi <sup>˘</sup>	nyi <sup>˘</sup> ma <sub>u</sub>	nyi <sup>˘</sup>	nyi <sup>˘</sup>
3	sm <sup>ˆ</sup> , sm <sup>˘</sup>	sheh:leh <sup>˘</sup>	sa <sub>u</sub>	so
4	oe <sup>˘</sup>	awn_leh <sup>˘</sup>	li <sub>u</sub>	lye
5	nga <sup>˘</sup>	nga <sup>˘</sup> ma <sub>u</sub>	ngwa <sup>˘</sup>	ngɿ
6	k'o <sup>ˆ</sup>	hkaw <sub>u</sub> ma <sub>u</sub>	hchaw <sup>˘</sup>	fu <sub>u</sub>
7	shi <sup>ˆ</sup>	suh <sub>u</sub> ma <sub>u</sub>	shi <sup>˘</sup>	she
8	yeh <sup>ˆ</sup>	hi <sub>u</sub> ma <sup>˘</sup>	h'i <sup>˘</sup>	hi <sup>˘</sup>
9	g'oe <sup>˘</sup>	k'aw <sup>˘</sup> ma <sub>u</sub>	ku <sup>˘</sup>	gu
10	tse <sup>˘</sup>	te <sup>˘</sup> chi:	htsi	tshi

From David Bradley, *Proto-Loloish* (London: Curzon Press, 1979) pp. 338–41, 382, and other sources



# LAMPUNG

**1,500,000 SPEAKERS**

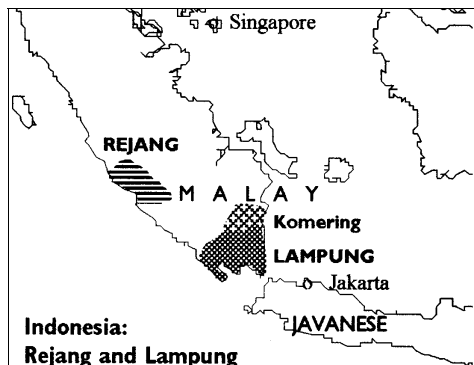
*Indonesia*

**O**ne of the AUSTRONESIAN LANGUAGES, Lampung is the regional language of Lampung province at the southern end of Sumatra. The province is heavily overpopulated as a result of government-directed resettlement from Java: speakers of Lampung are outnumbered, at least five to one, by migrants who speak Javanese or Indonesian.

Contacts with Malay, the trade language of the archipelago, go back hundreds of years. There is much bilingualism in Indonesian, now the official form of Malay.

Lampung has its own script, an offshoot of the Brahmi script of ancient India: this has been taught in schools and some people use it for personal letters.

The first ten numerals in Lampung are: *sai, ghua, talu, pa, lima, onom, pitu, walu, siwa, sapuluh*.



## Rejang and Lampung

There are about 1,000,000 speakers of Rejang and perhaps 1,500,000 of Lampung, related languages of south-western Sumatra. *Komerling*, with a further 700,000 speakers, is a variety of Lampung spoken in the Komerling Ilir and Komerling Ulu districts of South Sumatra province.

# LANNA THAI, KHÜN AND LÜ

**PERHAPS 6,500,000 SPEAKERS**

*Thailand, China, Burma, Laos*

**T**he *Lanna* or *Tai Yuan* language of northern Thailand, with its main centre now at Chiangmai, is very close to the *Lü* of the autonomous region of Xishuangbanna in south-western Yunnan. Between the two territories lies the state of Kengtung, at the eastern extremity of Burma: this was counted in British times as one of the 'Shan States' but its ruling language was *Khün*. Kengtung claimed a tributary state, Muong Sing, which is now at the north-western corner of Laos (for map see **LAO**).

Under its many names, the majority language in these areas is recognisably one; it is written in the same traditional script; and it is the vehicle of the same Buddhist culture. It belongs to the **TAI LANGUAGES**. *Tai Yuan* is the term preferred by some linguists.

*Xishuangbanna* is the Chinese form of the native *Lü* name, *Hsip hsong pan na*, 'Twelve thousand rice fields', for the historic state whose capital is at Jinghong (Shan *Kenghung*, Thai *Chiengrung*, older Chinese name *Ch'eli*).

*Lü* is commonly known in Chinese as *Baiyi* (older transliteration *Pai-i*), as *Xishuangbanna Dai* or simply as *Dai*. In Thailand the term *Lanna Thai* is now most used. British and Burmese have often confused *Khün* with *Shan*.

Local chronicles take the history of the three main states back to the 12th century, when Patseng ruled the Hsip Hsong Pan Na, and the 13th, when Mangrai founded the kingdom of Lanna at Chiangmai and its offshoot at Keng-

tung. The relations between the old Yunnanese kingdom of Nanchao and the Hsip Hsong Pan Na state are unknown, but it, along with neighbouring smaller states, has been tributary to China since the Mongol conquest of Nanchao in 1253. In more recent centuries, these *Tai* states paid occasional tribute to China, to Burma and to Thailand, retaining local autonomy, until colonial frontiers were drawn in the 1890s. Local relations among them have been maintained as much as possible, and their language and culture still give evidence of the multiple influences of their past.

Like its relatives, *Tai Yuan* is a tonal language: the *Lü* variety has six tones. It contains many loanwords from Pali, the language of the Buddhist scriptures. The *Jātaka* tales of the Buddha's former lives, originally written in Pali, have had a strong influence on local traditional literature.

## **'Northern Thai' script**

The traditional script for Lanna Thai, *Khün* and *Lü* is similar to that used for Lao: both derive from the Khmer alphabet. Pali texts, the scriptures of southern Buddhism, can also be written in this script. There was once metal type capable of printing this script at a press in Kengtung: its fate is unknown. In Xishuangbanna a simplified version has recently been introduced, and textbooks and traditional literature have been published. At Chiangmai, offset printing from handwritten calligraphy has been the rule;

but computer typesetting now permits high quality printing of this complex script.

# LAO

**PERHAPS 15,000,000 SPEAKERS**

## *Thailand, Laos*

**L**ao, one of the TAI LANGUAGES (see table of numerals there), is spoken along both banks of the middle Mekong. The majority of speakers, along the right bank and the southern tributaries, are in Thailand and their official and literary language is Thai. The Lao language in Thailand is sometimes called *Isan* or *North-eastern Thai*. The other speakers of this language, those along the more mountainous left bank of the Mekong and its northern tributaries, are the *Lao* or *Laotians*, and they form the largest linguistic group in Laos, whose national language this is.

In the 14th century the powerful state of *Lan Xang* was established; its ruler, at Luang Phrabang, exacted tribute from the other smaller Tai states that already encircled it. Lan Xang declined in the 18th century, and the states of the middle Mekong were fought over by Thailand and Annam. A century later, the French colonial administration in Indochina, asserting the claims of Annam (which they had already annexed), seized the states on the left bank of the river, leaving the right bank to Thailand. The kingdom of Laos, thus created as a French protectorate, became independent in 1949 and the scene of a long civil war.

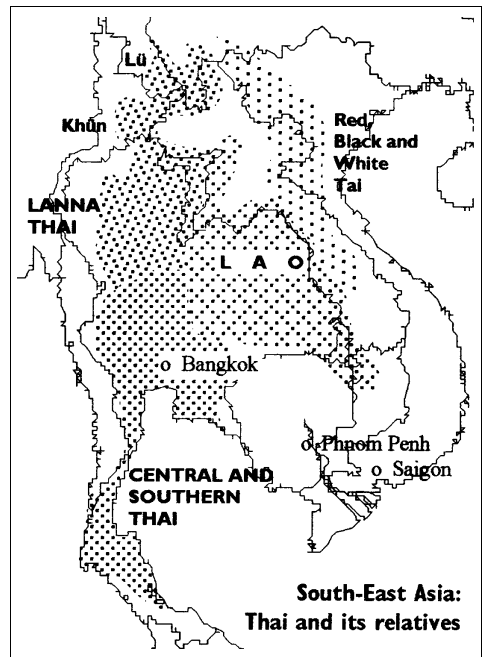
Lao is the language of a strongly Buddhist culture. Traditionally most young men spend some time as monks, and learn to read, if not to translate, the Pali scriptures, which are written in Lao script. The rich older literature of Lao, preserved in palm leaf manuscripts, includes religious and secular poetry as well as chronicles.

All dialects of Lao are tonal: some dialects have five tones, some six. Lao has borrowed cultural and religious vocabulary from Sanskrit,

Pali, Khmer and French. Under French rule many Vietnamese were employed in Laos. Thai, the closely related language of a much more powerful state, is now a strong cultural influence on Lao.

### **Thai, Lanna Thai, Lao and the Tai languages of Vietnam**

THAI or Central Thai is the national language of Thailand. The dialect of Bangkok forms the standard. The dialect of *Khorat* shows some Khmer influence. The *Pak Thai* or Southern Thai dialect, with Malay loanwords, is spoken in Chumphon and Nakhon Si Thammarat.



*Tai Yuan* or LANNA THAI or Northern Thai is the language of north-western Thailand, including the old principalities of Chiangmai, Lamphun, Lampang, Nan and Phrae; slightly varying dialects are spoken in Kengtung, Muong Sing, Jinghong and other, smaller neighbouring states.

*Lao* is the lowland language of Laos, the country once ruled from Luang Phrabang but now from Vientiane. It is also spoken in the north-eastern Thai provinces.

*Red Tai*, *Black Tai* and *White Tai* are all spoken by minorities in Laos and North Vietnam.

### The Lao alphabet

ກຂຄງ ຈສຊບ ດຕຖທ ນບປຝຟຸຟຸ ຢຣລວຫ ທງຫຍພພ ຫຼຫງອຮ  
g k k ng c s s ny d t t n b p p f f m y l l w h ng ny n m l w a h

In the box are shown, above, the thirty-three characters and their Latin equivalents.

Below are the thirty-eight vowel-tone combinations, shown here with the character ອ *a*.

Lao script is a variant of Thai. As with other

scripts of Indian origin, characters are combined into compounds to make up a syllable, including one or more initial consonants, a vowel and a tone mark.

ອະ ອັອ ອາ ອີ ອີ ອີ ອີ ອຸ ອູ ເອະ ເອັອ ເອ ແອະ ແອັອ ແອ ໂອະ ອີອ ໂອ ເອາະ ອອັອ ອໍ  
ອອອ ເອີ ເອີ ເອີ້ວະ ອັຮອ ເອັຮ ອຮອ ເອີອ ເອີອ ອີວະ ອັວອ ອີວ ອວອ ໂອ ໂອ ເອົາ ອໍ່າ

Font: *alice-4.ttf* by Ngakham Southichack

# L A T I N

## Classical language of Europe

**O**ne of the INDO-EUROPEAN LANGUAGES, Latin was the language of the ancient city of Rome. Rome became the capital of an empire that embraced all the Mediterranean countries and most of western Europe. In this way, Latin became a language of world importance, with a literature that is still read today. Latin was, and to some extent still is, the international language of the Catholic Church, the Christian sect out of which all others grew. It has been used as an official international language in many academic and scientific fields. Latin has had an influence on all the modern languages of Europe and the Mediterranean, and is the direct ancestor of French, Spanish and the other ROMANCE LANGUAGES.

The Latin or Roman alphabet was a local variant of Greek script. It has had even wider influence than the Latin language, being used in all continents for writing hundreds of languages. It is still spreading.

Two thousand five hundred years ago, Italic languages, a branch of the INDO-EUROPEAN family, were spoken in central Italy. They include Oscan, Umbrian, Sabine and Latin: inscriptions, in alphabets based on Greek, record the distinctive features of these and other dialects. Latin was the Italic dialect of the district called Latium (modern *Lazio*). Rome, with its seven hills on the south bank of the Tiber, was simply one of the towns of Latium.

The growth of Rome and of Latin may be traced to the city's emergence from the cultural and political influence of Etruria, to the north. Etruscan, a language unrelated to Latin, not yet fully deciphered, and now long extinct, can be seen to have had a heavy influence on Latin in its early stages. The Roman emperor Claudius (ruled AD 43–56) was one of the last professed experts on Etruscan. Latin loanwords from

Etruscan include *caerimonia* 'ceremony', *persona* 'mask', *histrio* 'actor', *baro* 'strong man, lout, baron'; all these, incidentally, were later borrowed into English.

The last of the Italic languages to challenge the position of Latin was Oscan, the official language of the anti-Roman allies in the Social War of 91–88 BC. Latin loanwords from Oscan include *popina* 'cookshop': Oscan had *p* where native Latin had *q*, and this word is cognate with Latin *coquus* 'cook'.

By the 1st century AD Rome ruled not only Italy but all the countries that surround the Mediterranean. Latin was the ruling language of the empire. It gradually extended its range as everyday spoken language, helped by its status in law and administration, by population mobility, by army recruitment and resettlement. The Latin of the empire borrowed a good deal of vocabulary from local languages: from Celtic, for example, came *carrum* 'wagon', *camminus* 'road', *cerevisia* 'beer'.

Greek, which was already the lingua franca of the eastern Mediterranean, retained its position and even increased in importance, for Greek literature, philosophy and science were admired. Well-educated Romans learnt Greek as well as Latin. Greek was also a lingua franca among slaves, and between slaves and owners. Greek loanwords in Latin include *persica* 'peach', *balneum* 'bath' (see box).

Greek had been the first major international language of Christianity, but Latin was soon to claim this role as the Bible and many other texts were translated into Latin and as the bishop of Rome came to be recognised as leader of the church. Christian Latin borrowings from Hebrew and Greek include *sabbatum* 'sabbath', *episcopus* 'bishop'. The Latin Bible in its *versio vulgata*, 'common translation, Vulgate' of about AD 405

influenced the style and vocabulary of later Latin and of all the medieval languages of western Europe.

The earliest surviving literature in Latin consists of Greek plays adapted for the Roman stage, and of a manual on farming by the statesman Cato (c. 150 BC). Later writings, including Vergil's *Aeneid*, the memoirs of Julius Caesar, and the histories of Tacitus, were in a language – *classical Latin* – that had begun to differ strongly from everyday speech. A few texts have colloquial features (Cicero's *Letters*; Petronius' novel *Satyricon*; the Vulgate) but in general literary Latin and spoken Latin diverged from this point onwards, the literary language demanding difficult study, the spoken language of the empire – *vulgar Latin* – gradually differentiating into regional dialects which would become the Romance languages.

Although the western Roman Empire disappeared in 476, it was many centuries before the Romance languages emerged as written languages in their own right. *Medieval Latin*, through all this time, remained the major literary medium of western and central Europe. Many varieties can be distinguished, ranging from the difficult and highly scholarly styles of some Irish and Anglo-Saxon writers to clear and simple tales and sermons intended to be widely understood. Although the classical writings of the 1st centuries BC and AD are generally agreed to be the greatest in literary terms, the quantity of medieval Latin literature is vast.

As literature in the everyday languages of medieval Europe spread, Latin retained its position as the international language of religion and scholarship. Ironically, this position began to crumble at the time of the Renaissance, in the 15th century. New inspiration was found in classical Greek and Roman culture – but it was the modern languages, not Latin, that were developed (with the help of Greek and Latin loanwords) as media for literature of all kinds. Meanwhile the Catholic Church found its own eminence challenged, and translations of the Bible, previously outlawed, were made in many modern languages. *Modern Latin* remained in use in the Catholic Church and among scientists and

scholars: some still used it for literature.

Latin retains some importance in the Catholic Church today. Otherwise it has few residual uses. It is still heard in certain Catholic and academic ceremonies. It is seen in inscriptions and mottoes. A jargon known as *botanical Latin*, developed by Linnaeus and his followers, is demanded in the definitive descriptions of new botanical species. More important, Latin is still studied in schools and universities in Europe and some other parts of the world, particularly for the sake of its literature and as the language of historical source materials for ancient and medieval studies.

Latin is omnipresent in English: in old loanwords (*feverfew* from Latin *febrifuga* 'driving off fever'), in newer loanwords (*re* 'on the subject of', *via* 'by way of'), in loanwords that came by way of medieval French (*chapter* from Latin *capitulum*; *vetch* from Latin *vicia*), in modern scientific terms (*canine* [Latin *canis* 'dog'], *penis*), in fixed phrases, in abbreviations of phrases (e.g. [*exempli gratia*] 'for the sake of example'; *i.e.* [*id est*] 'that is'). A similar range of Latin loans is to be found in many other European languages.

Written classical Latin had 20 consonants and 5 vowels, but only 23 letters, because two letters can function either as vowels or as consonants. Non-standard spellings in inscriptions, graffiti and manuscripts help to show how the sound pattern of Latin gradually shifted as the Romance languages emerged.

The first ten numerals in Latin are: *unus, duo, tres, quattuor, quinque, sex, septem, octo, novem, decem*.

### Latin and Sabine at the origins of Rome

'Latin was not even the language of all the seven hills of Rome: there were Sabine villages on the Quirinal, and the two languages clearly interacted at a very early stage.

'One of the special characteristics of Sabine was the development of *d* to *l*. Some Latin words show *l* where we might expect *d*: e.g. *lingua* supplanting Old Latin *dingua*, cf. English *tongue*; *lacrima* supplanting Old Latin *dacruma*, cf. Greek *dákrý*, English *tear*. Such Latin words

are therefore very early Sabinisms . . . Had it not been for these, we might now be discussing *dinguistics!*'

Robert A. Hall, Jr, *External history of the Romance languages* (New York: American Elsevier, 1974) pp. 59, 76, abridged

'to say') developed the implication 'what is said by a god or prophet' and so came to mean 'destiny', like the English word *fate* which is borrowed from it and like its direct descendant, Spanish *hado*. In the Romance languages a feminine form of this word gained the more concrete meaning of a spirit that brings good or bad luck, a 'fairy': Italian *fata*, Spanish *hada*, French *fee*. A derivative of this French word, *féerie* 'fairyland', is now used in English with the meaning of the parent word: it is *fairy*.

**Abstract and concrete**

In the course of two thousand years some Latin words have changed their meaning quite drastically.

**Balneum:** Latin *balneum* 'bath' was a loanword from Greek *balaneion* for a luxury that was to become typical of Roman civilisation. In several Romance languages the word still has this meaning, Spanish *baño*, Italian *bagno*, French *bain*, Romanian *baie*. In Italian, however, new senses arose. The cellars of the castle of Livorno were called *bagno* because they were below sea level, but they were used as dungeons for Turkish slaves: hence *bagno* in older English, and *bagne* in French, 'dungeon, workhouse'. English *bagno* was also used for 'brothel', because in medieval Italy and other countries the bath-houses were seen to be serving this purpose. Finally in Romanian place names *baie* means not 'bath' but 'mineral source' or 'mine', a natural shift of meaning due to the mineral-rich natural springs of Transylvania.

**Fatum:** Latin *fatum*, originally 'what is said' (*fari*

**The Latin alphabet**

For Latin twenty-three letters were used, A B C D E F G H I K L M N O P Q R S T V X Y Z.

The characters *k x y z* were wanted only in Greek loanwords. The distinction between short and long vowels was not marked in the script. The letters *i* and *v* could serve either as consonants or vowels. Other languages have adjusted this alphabet slightly. In many, *j* is added as the consonant variant of *i*, and *u* appears as the vowel variant of *v*. The *w* was introduced as a 'double-v' to make a distinction that had been unnecessary in Latin between the sounds we know as *v* and *w*.

These are the basic typographical styles (font groups, as they might be called now) that have been used in printing the Latin alphabet:

**Roman**

ABCDEFGHIJKLMNOPQRSTUVWXYZ  
abcdefghijklmnopqrstuvwxyz

**Italic**

ABCDEFGHIJKLMNOPQRSTUVWXYZ  
abcdefghijklmnopqrstuvwxyz

**Black-letter**

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z  
a b c d e f g h i j k l m n o p q r s t u v w x y z



# LATVIAN

**1,500,000 SPEAKERS**

## *Latvia*

**L**atvian and Lithuanian are the two surviving **BALTIC LANGUAGES** (see map and table of numerals there). Latvian is the official language of the Republic of Latvia, which became independent from the Soviet Union in 1991.

The name of *Latvia* derives from the speakers' own name for themselves, *Latvis*. The language was once better known in English as *Letnish*; in French it is still called *Letton*.

Latvian speakers, expanding northwards from neighbouring Lithuania, were in this region by the 10th century, and probably rather longer if, as linguists believe, Latvian began to separate from Lithuanian around AD 600.

Latvia was under the rule of the Teutonic knights, largely German-speaking, and then of German landowners and bishops, from 1158 to 1562: during much of this time it was divided between two states, Courland and Livonia. Riga, now capital of Latvia, was in its origin a Hanseatic port and its everyday language was Low German. Poland ruled Courland from

1562; Livonia was ruled by Sweden from 1629. But both territories eventually fell to Russia, which eventually governed the whole of Latvia until 1918 and seized it again in 1940. Germans occupied Latvia in 1941–4. This complex history led to a division in Latvia between Orthodox, Catholic and Lutheran Christianity: the majority of the population is Lutheran.

Hundreds of thousands of Latvians died in the Second World War and its aftermath. After 1944 the ethnic situation was complicated by mass deportations of Latvians to Siberia and Kazakhstan, and mass immigration of Russians. The Russian minority amounted to 40 per cent of the population on independence.

A Lutheran catechism was published in Latvian in 1586. In the 19th century Latvian speakers became increasingly aware of their national and linguistic distinctness and their folklore.

Latvian reflects its history in its numerous loanwords, from Swedish, German, Polish and particularly Russian.

# LEZGHIAN

**400,000 SPEAKERS**

*Russia, Azerbaijan*

**O**ne of the Lezghian or Samurian group of North East CAUCASIAN LANGUAGES (see map there), Lezghian has been one of the official languages of Dagestan since soon after the Russian Revolution. Its status in Azerbaijan is far more precarious.

Islam was introduced in the 15th century to this part of the Caucasus. The territory was annexed by Russia between 1812 and 1865.

Arabic was the language of literature and education for this Muslim mountain people; Azeri was becoming the lingua franca of everyday life. In turn, Lezghian served as a lingua franca for the smaller number of speakers of Agul, Rutul and Tabasaran. On the occasions when Lezghian was written, Arabic script was used.

With the development of Soviet nationality policy, Lezghian was recognised as a literary language both in Dagestan and Azerbaijan, and a Latin orthography was introduced in 1928 – followed, as usual, by a Cyrillic alphabet in 1938. The number of Lezghian speakers grew steadily through Soviet times, although in 1939 Lezghian lost its position as an official language in Azerbaijan, and in the 1950s Russian was made the educational and official language for

speakers of Agul, a position that Lezghian had formerly held.

Since Azerbaijan became independent in 1991 many Lezghian speakers from northern Azerbaijan have fled northwards to Dagestan. Their lands have been resettled with Azeri refugees from Armenia and with Meskhetian Turks, exiled from Georgia to Uzbekistan in Stalin's time, now fleeing ethnic hostility there. Lezghian speakers remain dissatisfied with the division of their territory between Russia and Azerbaijan.

The three main dialects of Lezghian are Kürin, Akhty and Kuba, the latter spoken in Azerbaijan. The literary language is based on Kürin.

There are a few thousand Lezghian speakers in villages near Balıkesir and İzmir in Turkey. The Muslim migrants settled there in 1865, after the Russian conquest of their homeland, and are said to cling tenaciously to their language and culture with its traditions of hospitality.

The first ten numerals in Lezghian are: *sad*, *qwed*, *pud*, *qud*, *wad*, *rughüd*, *erid*, *muzhüd*, *k'ud*, *ts'ud*.

# LI

800,000 SPEAKERS

*China*

**L**i is a language of Hainan Island, and is one of the KADAI LANGUAGES (known in China as Zhuang-Dong languages). It is thus distantly related to the T'ai group.

*Li* is the Chinese form of the speakers' own name for themselves, which varies according to dialect: *Hlai*, *Dli*, *B'li*, *Le*, *Lai*, *Loi*, *Day*, *K'lai* are all found.

Traditionally, trade in Li villages was in the hands of Chinese shopkeepers and pedlars. Li, particularly the southern dialect spoken by the Ha Li, has been noticeably influenced by Chinese: a large number of speakers are now bilingual. This is not to say that Li speakers were easily brought under Chinese rule. The 'Wild Li of Hainan' were famous for their long resistance to conquest – and, fairly or unfairly, for cannibalism too. Fifty rebellions can be found in the history books, extending over the 2,100 years since the first Chinese colonies were established on Hainan.

The last rebellion was the most decisive. In 1943 Wang Guoxing, the leader of Li resistance to the Nationalist Chinese government, made contact with the Communist guerrilla forces on the island. Li speakers were thus among the very few Chinese minorities to ally themselves with the Communists, and the Li Column was soon celebrated in Chinese legend for its contribution to the struggle. The former 'Wild Li' are now much more highly regarded. The Hainan Li and Miao Autonomous Region was created in 1952.

Li is a tonal language, and in essence monosyllabic, like its T'ai relatives. The proportion of vocabulary shared between Li and the T'ai languages – the vocabulary from which their histor-

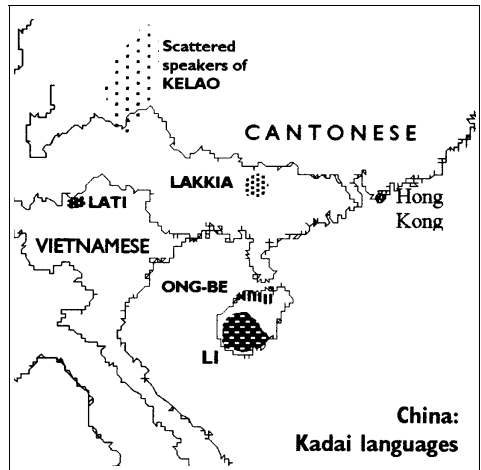
ical relationship is demonstrated – is actually quite small, indicating a very long separation.

Li had no traditional writing system: if Li speakers became literate, they wrote Chinese. The language shows some influence from the YAO dialect (inaccurately called 'Miao') that has been spoken on Hainan since the 16th century.

## Li and related languages

There is little detailed information as yet on dialect divisions of Li. *Southern Li* is in broad terms the dialect of the Ha, who live in the southern part of Hainan near the coast. *Northern Li*, or *Loi* to some authors, has numerous tribal and dialect divisions, largely to be found inland. Five Li dialect groups are generally recognised.

Other Kadai languages, related to Li but at some distance in time, are spoken by small and remote communities of the southern Chinese



mainland. They include *Kelao* or *Gelo*, *Lakkia* or *Laqua*, and *Lati*. None has as many as ten thousand speakers; *Lati* has only a few hundred, in the Vietnamese hills just south of the Chinese border.

*Ong-be* or *Bê*, a language of obscure origins,

possibly *Kadai* but with numerous loanwords from Chinese, is spoken by a fairly large population, perhaps as many as half a million, on the northern coasts of Hainan. Speakers of *Ong-be* are very much Chinese in their culture, and are not recognised as an official minority.

# LINGALA AND BANGALA

**PERHAPS 12,000,000 SPEAKERS, MAINLY AS SECOND LANGUAGE**

*Congo (Kinshasa), Congo (Brazzaville),  
Central African Republic*

**O**ne of the BANTU LANGUAGES, Lingala developed in the 19th century from dialects spoken by the Ngala people, inhabitants of the banks of the middle Zaire near Mbandaka and Nouvelle Anvers. It appears closest to the now-minor languages Boloki and Bobangi.

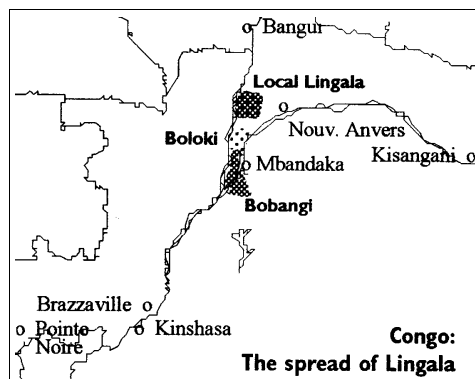
The Ngala, originally fierce opponents of Stanley's advance, eventually became allies of the European colonists. Their trade language, Mangala, had already spread upriver and downriver before European soldiers and traders arrived in large numbers. When they did, they took up Mangala as a medium of communication over an even wider region. As new towns and trading posts developed along the Zaire and its tributaries, a creolised Mangala, called Lingala, became the preferred language among migrants of varying mother tongues who had to find a living there. Much army recruitment also took place among Ngala speakers near Nouvelle Anvers, and Lingala became the language of the army. Soon new recruits were required to know 600 words of Lingala.

Lingala – language of local government, the army and the new towns – appeared so important in official eyes that in 1918 the Belgian government tried to make it the official language of the whole of Belgian Congo.

As development continued to encourage migration and intermarriage, Lingala rapidly became a mother tongue in towns and cities. It is

now the predominant language of the middle Congo, upriver almost to Kisangani and downriver all the way to Kinshasa, and is spoken by many inhabitants of the big city. Even the speakers of Kongo dialects – the majority in Kinshasa – need Lingala as a second language. Originally a language of trade and employment, Lingala is now a major mother tongue of modern Congo (Kinshasa). It is also spoken by some in Brazzaville (though KITUBA is the majority lingua franca there) and along the line of the railway to Pointe Noire on the Atlantic coast.

Its use as a language of trade, of education and religion, and of the home has led to the rapid development of varieties of Lingala: in Kinshasa, for example, 'street Lingala' or 'river Lingala' is quite different from the *Langue scolaire* or 'school Lingala'.



Originating as an inter-Bantu lingua franca, Lingala remained difficult for others – hence the emergence from it of Bangala. This, with 3,500,000 speakers in Congo (Kinshasa) and the Central African Republic, is a pidgin form of Lingala which arose in contact between Africans and Europeans. It is now a language in its own right, characterised by much-simplified grammar and word formation. Use in the army has spread Bangala in the Uele area and in the north-east of Congo (Kinshasa) among a mainly non-Bantu-speaking population.

**Lingala and Bobangi**

In the 19th century *Lingala* was still merely the language of one of several trading communities on the middle Congo. Its growth since then has been astonishing.

*Bobangi* or Bangi, a related Bantu language, was spoken by the traders of the river Ubangi, who also ranged the Zaire as far as Malebo Pool (Stanley Pool) where Kinshasa and Brazzaville now stand. It now has about 70,000 speakers in Congo (Kinshasa) and Congo (Brazzaville).

Numerals in Kongo, Lingala and Bobangi			
	Kongo	Lingala	Bobangi
1	-mosi	m-ókó	-oko
2	-ole	mi-balé	-bale
3	-tatu	mi-sáto	-sato
4	-ya	mi-nei	-nei
5	-tanu	mi-táno	-tano
6	-sambanu	motóbá	motoba
7	nsambwadi	nsambo	ncambo
8	nana	mwambi	mwambi
9	vua	libwá	libwa
10	kumi	zómi	zomu

# LISU

650,000 SPEAKERS

*China, Burma and south-east Asia*

One of the SINO-TIBETAN LANGUAGES, Lisu is spoken in the upper valleys of the Salween and Mekong rivers, and by numerous migrant communities in the mountains of south-east Asia.

Speakers call themselves *Lisu*; Chinese, Lanna, Lahu and other neighbours tend to use a form such as *Lishaw*. An older Chinese term is *Yeh-jen* 'wild men'. This may be the origin of Kachin and Burmese *Yawyin*, *Yawyang*, a name often found in older English writings on the Lisu.

The first published wordlist of Lisu appeared in 1871, in a report by a British explorer of the Burma–Yunnan route.

The 'wild Lisu' or Black Lisu of the north are traditionally fiercely independent, living in easily defended villages and – at least until recently – exacting tribute from traders who passed below them. They practised slash-and-burn agriculture, hunted, and collected wild honey. Above Lisu villages that adhere to the traditional religion there will be a spirit grove of sacred trees. Lisu priests were said to have a secret religious language, and, according to Erik Seidenfaden, a hieroglyphic script in which sacred texts were recorded: if this exists, no details of it have been published.

There have been massive conversions of Lisu speakers to Christianity: the mission of the Church of Christ has been particularly successful among them. The dialect of west central Yunnan, very similar to that of the northern Shan State, acts as a standard language, with an orthography in a missionary-devised extension of Latin script in which both Christian and secular

texts have been published. For a table of numerals see LAHU.

Northern dialects show a characteristic sound change: initial *py-* *by-* *my-* become *c-* *j-* *ny-*.

'Chinese Lisu', spoken by the people who used to be called 'tame Lisu' in central Yunnan and in some communities in northern Thailand, is a creolised variant of the language, resulting from extensive intermarriage with Yunnanese Chinese speakers.

## The Lisu alphabet

BPpDTtGKkJCcZFfMNLsRAXWYgVh  
 AVEEIOUŋtD , . , : ;  
 b p ph d t th g k kh j c ch dz ts tsh m n l s r ng sh  
 v y h nh kh a æ e ö i o u ü ghı gha ghe ˊ ˋ ˊ ˋ ˊ ˋ

In standard Lisu only the upper-case Latin letters are used; additional sounds are symbolised by upside-down letters. The six tones of Lisu are marked by punctuation symbols. The vowel *a*, when it follows a consonant, is unmarked (just as in Indic scripts). This unusual writing system was devised by the highly successful missionary J. O. Fraser, whose *Handbook of the Lisu (Yawyin) language* appeared in Rangoon in 1922.

A second writing system for Lisu is a syllabary of several hundred characters, apparently devised by a Lisu peasant, Wang Renbo, around 1925. It looks a little like Chinese script, but is not really based on it or any other, because Wang Renbo was illiterate. Although unsystematic and therefore hard to learn, Wang Renbo's script continued to be used in the neighbourhood of his home town, Weixi.

# LITHUANIAN

**3,500,000 SPEAKERS**

## *Lithuania*

**L**ithuanian is one of the BALTIC LANGUAGES (see map and table of numerals there). It has the distinction of being the language that has changed least over the several thousand years since PROTO-INDO-EUROPEAN began to split up into the dialects and languages now spoken in a wide swathe across Europe and southern Asia. Lithuanian is the official language of Lithuania, which was the first constituent republic of the Soviet Union to make good its escape from the collapsing superpower.

The *Litva* were already here, at the south-eastern corner of the Baltic, in the 10th century. Their earlier history is really unknown, though archaeologists consider it likely that their ancestors had occupied the region for a long time: certainly a Baltic language influenced an early form of FINNISH and Estonian over two thousand years ago, suggesting that the two language groups were adjacent at that time.

In the 14th century the Grand Duchy of Lithuania had become the largest state in Europe, stretching all the way from the Baltic to the borders of the Khanate of Crimea. The country was still pagan, but on the marriage of Queen Jadwiga of Poland to Grand Duke Jagiello of Lithuania (in 1385) it became officially Catholic, almost the last European nation to adopt Christianity. It now formed part of a double Polish-Lithuanian state in which Polish became the ruling language (until very recently Polish was the majority language of the Lithuanian capital, Vilnius). When this state was broken up, in the 18th century, Lithuania was annexed by Russia and Russian was the new language of prestige. Lithuanian sank into the status of a peasant tongue. After rebellions in 1830, printing and

teaching in Lithuanian were outlawed and many schools and Roman Catholic monasteries were closed. The German occupation of 1915–18 was equally oppressive.

Independent from 1918, Lithuania was once more seized by the Soviet Union in 1940, and suffered a second German occupation in 1941–4. The ethnic situation was complicated by mass deportations of Lithuanians to Siberia and Kazakhstan, and mass immigration of Russians. There was meanwhile considerable emigration of Lithuanian speakers to western Europe and America – in 1830 to 1910 these were mostly unskilled workers; in the 1940s they were refugees from Soviet occupation. Many rapidly assimilated to their new linguistic environment; others were ready to return as Lithuania declared its independence once more on 11 March 1990 and Russia ceased to dispute the position in September 1991.

The first Lithuanian printed text, a catechism, appeared in 1547. The first dictionary was printed in the 17th century. A more extensive publishing trade developed in the 19th century – and at this period there was already a vigorous press serving Lithuanian communities in the United States. As the nationalist movement flourished under Russian repression, Jonas Mačiulis (1862–1932), ‘prophet of the Lithuanian renaissance’, emerged as perhaps the greatest 19th-century poet.

Lithuanian vocabulary has striking points of similarity with the most ancient known Indo-European languages: *ašis*, Latin *axis* ‘axle’; *avis*, Latin *ovis* ‘sheep’; *katras*, Sanskrit *kataras* ‘which’. Lithuanian is a tonal language, as ancient Greek and Sanskrit were and as proto-



Indo-European probably was. Nouns have seven cases. There are three numbers, singular, dual and plural. Verb forms, however, are much simpler and more regular than the usually reconstructed proto-Indo-European forms.

In modern literary Lithuanian there has been a concerted attempt to supersede the loanwords from German, Polish, Ukrainian and Belorussian that are found in early texts.

# LOMWE

**2,000,000 SPEAKERS**

*Mozambique, Malawi*

**L**omwe, one of the BANTU LANGUAGES, is spoken in inland Mozambique and in parts of neighbouring Malawi (for map see MAKUA).

The speakers of Lomwe are sometimes called *Manguro* and *Alolo* in older narratives. *Alolo* is a YAO name for them.

Their older history is unknown. They are traditionally farmers, and the rich, fertile Lomwe valleys of Mozambique are heavily populated. There has been long interaction between

Lomwe and Nyanja speakers, and apparently much migration in both directions. The Miha-vani, a trading group from the west of the Lomwe-speaking country in Mozambique, are said to be a 'mixture' of Lomwe and Nyanja. After a famine in 1900, however, many Lomwe speakers settled in Nyasaland (now Malawi). More went there later as migrant labourers.

There are now about a million Lomwe speakers in each country. In Malawi they are called *Nguru* or *Anguru*.

# LOZI

450,000 SPEAKERS

*Zambia, Zimbabwe, Angola*

**T**hreatened by the Zulus under their great ruler Chaka, in 1823 thirty thousand Sotho people of the Kololo group, led by the twenty-year-old Sebitwane, marched away from their homeland on the north banks of the Vaal in modern South Africa. Their long march westwards and northwards, involving a great deal of fighting, ended when they conquered the Luyi-speaking Lozi, whose lands were in the upper Zambezi valley in modern Zambia. The Kololo became the ruling caste, and their own dialect of Sotho – one of the BANTU LANGUAGES – was their language of empire (for map see SOTHO). It was already the court language, and a lingua franca of the Upper Zambezi, when David Livingstone visited Sebitwane's capital in 1853.

The Lozi rose against the Kololo in 1864, and ruled from then onwards. The language of the Kololo had already spread widely. The result of the political reversal was not, as might have been expected, a resurgence of the local Luyi language. The Lozi victor, Lewanika (ruled 1878–1916), expanded the boundaries of his kingdom. Lewanika's policies encouraged the spread of the language of the former conquerors even further. Until 1941 Luvale and Southern Lunda speakers of northern Barotseland were ruled by the Lozi under British suzerainty.

Originally known as *siKololo*, this linguistic offshoot of Sotho is now called *Lozi* after its speakers, or *Rози* or (in their own language) *siLozi*. The first missionaries who worked with this language spelt its name *Rotse*, and that of the tribe *baRotse*. That is how the Lozi territory in Northern Rhodesia, nowadays Zambia, came to be called *Barotseland*.

Lozi speakers are traditionally gardeners of the margins of the Central Barotse Plain through which the Zambezi flows. They cultivate at least eight different kinds of gardens, distinguished by soil, fertilisation and irrigation. Fishing and cattle-herding are also commonly practised.

Lozi is widely used as a lingua franca of south-western Zambia, understood by many men who are speakers of other regional languages. The number of women who speak it is smaller, since women travel less. Lozi is the main language of Livingstone and one of the eight official languages of Zambia, used beside English in local administration and education. The first ten numerals in Lozi are: *ngwi*, *-peli*, *-talu*, *-ne*, *ketalizoho*, *ketalizoho ka alimungwi*, *ketalizoho ka zepeli*, *ketalizoho ka zetalu*, *ketalizoho ka totune*, *lishumi*. *Ketalizoho* means literally 'finish the hand'.

There are still some speakers of the old language of the valley, Luyi, a dialect of Luyana, which is related to LUVALE to the north. Luyi survives in religious ceremonies and in traditional songs.

## Why did Lozi survive?

'The Sotho speakers, a minority, brought with them only a few women, and most of those were rendered barren by the endemic malaria of the Zambezi flood plains. One would expect the language of these conquerors to disappear without trace, especially after the return of the old Luyi dynasty. But Lozi is alive and well, still a typical Sotho dialect, mutually intelligible with Southern Sotho to this day.

‘A factor which may have made the difference here is the early and continuous presence of missionaries. They used a Sotho Bible in their work: to this day Lozi is the only liturgical language in Barotseland, the only literary language in the region, and in fact for decades the

only African language in Zambia committed to print.’

I. T. Givón in *Readings in creole studies* ed. Ian F. Hancock (Gent: Story-Scientia, 1979) pp. 11–12, modified

**Noun classes in Lozi**

As in other Bantu languages, nouns in Lozi belong to classes marked by prefixes:

Singular			Plural		
Class	Example		Class	Example	
I	mu-tu	person	IX	ba-tu	people
II	mu-lomo	mouth	X	mi-lomo	mouths
III	li-lundu	mountain	XI	ma-lundu	mountains
VI	lu-limi	tongue		ma-limi	tongues
IV	si-shimani	big boy	XIV	bi-shimani	big boys
	si-katulo	shoe	XII	li-katulo	shoes
V	kuhu	chicken		li-kuhu	chickens
IV	ka-twa	trap	XIII	tu-twa	traps
VIII	bu-ta	bow			

From Yashutoshi Yukawa, ‘A tonological study of Lozi verbs’ in *Studies in Zambian languages* (Tokyo: ILCAA, 1987) pp. 73–128

# LUBA

7,800,000 SPEAKERS

Congo (Kinshasa)

**T**he speakers of Luba, one of the BANTU LANGUAGES, inhabited the plateaus of southern Congo (Kinshasa) from the upper Sankuru to the upper Lualaba valleys.

In 1869 Swahili-speaking adventurers from Zanzibar founded the stronghold of Nyangwe on the Lualaba as a centre for trade and slaving raids. The ‘Congo Free State’ and Belgian colonial governments inherited the Zanzibar trading networks here. Roads, railways, river harbours and mines were developed: the baLuba people, formerly a source of slaves, still provided labourers and, eventually, office and transport workers. Wherever the new arteries went, Luba speakers settled and worked. Luba speakers multiplied in local administration, teaching and the clergy. Thus Luba is not only a local language but has become a lingua franca whose territory covers the two Kasai provinces: in other words, much of southern Congo (Kinshasa), from the Lubilash valley to the Lualaba valley. There are Luba-speaking communities in the big cities of Kinshasa, Kisangani and Lubumbashi, too. In Kinshasa, however, LINGALA is the usual lingua franca. Luba is now one of the national languages of Congo (Kinshasa).

The lingua franca is clearly based on western Luba, the dialect adopted in education and the church. Christian missionaries encouraged this by using western Luba in teaching and preach-

ing. Luba includes loanwords from Portuguese (see box), from English (*sisibete* ‘bedsheet’; *mbulankeci* ‘blanket’), from Zaire Swahili and nowadays increasingly from French, the ruling language of Congo (Kinshasa).

The first ten numerals in Luba are: *u-mwe*, *i-bidi*, *i-satu*, *i-naayi*, *i-taanu*, *i-sambombo*, *muanda muteke*, *muanda mukulu*, *citeema*, *diikumi*.

## The geographical spread of Luba

The major dialects of Luba are the central (*kiLuba*, Luba-Shaba or Luba-Katanga) and the western (*ciLuba*, Luba-Kasai or Luba-Lulua).

*Luba* belongs to a larger dialect continuum. This also includes *Bangubangu* (120,000 speakers), *Hemba* or Luba-Hemba or Eastern Luba, *Kanyok* (200,000 speakers), *Sanga* or Luba-Sanga (450,000 speakers), *Songe* or Luba-Songi or North-east Luba (nearly 1,000,000 speakers) and *Kaonde* (220,000 speakers) in Zambia, where it is one of the eight official languages of the country.

*Kuba* or *Luna* or Northern Luba also belongs to the group. The Kuba kingdom was an independent power in the 18th and 19th centuries. Hence this was once an influential language, and features of Kuba can be found in the variety of Luba that is now used as lingua franca.

### Noun classes in Luba

Like other Bantu languages, Luba nouns belong to classes marked by prefixes:

Singular			Plural		
Class	Example		Class	Example	
I	mu-untu	person	IX	ba-antu	people
II	mu-keta	arrow	X	mi-keta	arrows

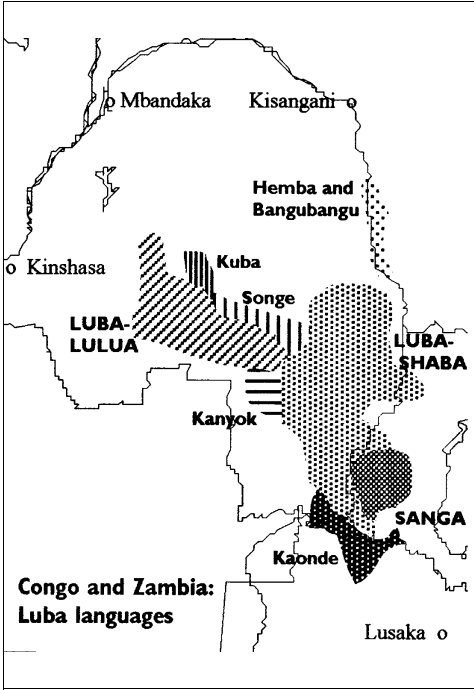
III	di-laandi	snail	XI	ma-laandi	snails
VIII	bu-ta	bow		ma-ta	bows
IV	ci-laamba	clothing	XII	bi-laamba	clothes
V	n-koonko	finger ring	XIII	n-koonko	finger rings
VI	lu-kasu	hoe		n-kasu	hoes
VII	ka-minyiminyi	scorpion	XIV	tu-minyiminyi	scorpions

From Yashutoshi Yukawa, 'A tonological study of Luba verbs' in *Studies in Cameroon and Zairean languages* (Tokyo: ILCAA, 1992) pp. 303–62

**Loanwords and noun classes in Luba**

'The Luba word for "box" or "chest" is *mushete*. Its meaning aroused my suspicion [as a potential loanword] from the start, but I was unable to pin down its origin. When I found it in KIMBUNDU as well, I was certain that it must be a loanword and that its origin must be Portuguese. The Portuguese word for this article is *caixete* and this would become something like *\*kashete* in a Bantu language. But the Bantu speakers seem to have rejected this form of the word since the first syllable *ka-* has the shape of the prefix of class 12 [of Guthrie (see box at BANTU LANGUAGES); VII in the table above], which denotes only small things; for big things and for things made of wood, the *mu-* prefix [II above] is used and has therefore been substituted.'

Jan Knappert, 'Contribution from the study of loanwords to the cultural history of Africa' in D. Dalby and others, *Language and history in Africa* (London: Cass, 1970) pp. 78–88



# LUNDA

550,000 SPEAKERS

*Congo (Kinshasa), Zambia, Angola*

One of the BANTU LANGUAGES, Lunda is spoken in the wooded plateau north and west of the headwaters of the Zambezi river. Once the language of the Lunda empire of Mwata Yamvo (ruled from what is now Kapanga district of the Shaba region of Congo (Kinshasa)), Lunda has since the European conquest and partition been spoken by almost equal numbers in Congo and Zambia and a smaller population in the Moxico province of Angola. It is one of the eight official languages of Zambia.

Genealogies preserved in oral tradition take Lunda history back to the 16th century, when a King Mwaku ruled a Lunda population in Congo (Kinshasa). The Southern Lunda people (and also the CHOKWE) are said to derive from migrations southwards in the 17th century. The Northern Lunda of Mwata Yamvo meanwhile came to be ruled by a LUBA dynasty: the southerners, who customarily inherit in the female line, claim to preserve the traditions of the early Lunda more fully than the northerners, who adopted male inheritance from their Luba overlords. In the late 19th century Chokwe speakers, expanding northwards once more, conquered Mwata Yamvo. Even then the Southern Lunda were not quite cut off from the old empire of the north, to which they still occasionally paid tribute. The Chokwe expansion covered much of 'Lunda' province of Angola, a region once subject to Mwata Yamvo: there are few Lunda speakers in this province now.

Northern Lunda is naturally influenced by Luba and Chokwe; Southern Lunda shows strong Portuguese influence, even in Zambia, because of continuing migration across frontiers. The first six numerals in Southern Lunda are: *-mu*, *-yedi*, *-satu*, *-wana*, *-taanu*, *itaanu naciimu*.

## Two genres of Lunda oral literature

The Southern Lunda fall into about twelve named matriclans, which are dispersed and not localised units. As most of the clan names are also common among Luena [see LUVALE], Chokwe and Luchazi, they were presumably known among the Lunda prior to their emigration. Possessors of the same clan name were formerly forbidden to marry . . .

'Common possession of the clan name implies the obligation to give assistance and hospitality. When a stranger from another village with the same clan name arrives, he quotes his clan name and the clan "formula". The formula consists usually of names of some of the clan progenitors and references to their exploits, and often contains archaic terms.

'*Senseka* is a mutual assistance agreement which holds, not between clans, but between the Ndembu and Akosa groups of Lunda. A member of one, visiting the other, calls them *asense nindi*, while they call him *asense netu*. They have an obligation to offer him hospitality, including a hut, food, and a mat to sleep on, while the *asense netu* has theoretically a right to do almost anything, from committing adultery to mere reviling, without incurring penalties. In practice he will often make outrageous jokes, e.g. with the women, forbidden among the members of the groups concerned.'

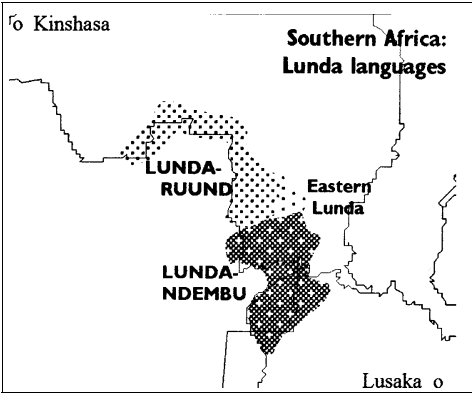
Merran McCulloch, *The Southern Lunda and related peoples* (London: International African Institute, 1951) citing work by C. M. N. White, F. H. Melland and J.C. Sakatengo

**Northern, Eastern and Southern Lunda**

The three groups speak rather different dialects. Northern and Southern Lunda are said to be no longer mutually intelligible.

The 250,000 speakers in Congo (Kinshasa) are divided between Northern, Eastern (*Luwunda*) and Southern Lunda. In Zambia and Angola all but a few speak Southern Lunda.

*Northern Lunda* is sometimes called *Luunda* or *Ruund*: among Bible translators it was traditionally *Lunda of Kambove*, where the first translators worked. The name *Ndembu* or *kiNdembu*, belonging to a single population group in north-western Zambia, has often been used as an alternative for *Southern Lunda* as a whole. This dialect was



called by translators *Lunda of Kalunda*, after a missionary centre in eastern Angola.



# LUO LANGUAGES

**PERHAPS 5,800,000 SPEAKERS**

*Uganda, Kenya, Sudan, Ethiopia, Tanzania*

**A** subdivision of the Western Nilotic group of NILO-SAHARAN LANGUAGES, Luo in its various related forms is spoken by a scattered series of peoples extending from western Ethiopia to the northern extremity of Tanzania. Traditionally, most speakers of Luo languages are cattle herders.

These languages differ from one another largely as a result of their linguistic and social environment – most of them are surrounded by other, quite unrelated languages. Anywa, for example, has borrowed vocabulary from Amharic, English and Oromo. Luo has interacted closely with Luyia, Karamojong (see **TURKANA**), Swahili and English. Shilluk shows Arabic influence: it is said that Shilluk speech once extended further north towards Khartoum, and has retreated before the advance of Arabic.

There is a press in Lango-Acholi and a small-circulation press in Alur. Anywa is an official language in Ethiopia, and the New Testament has appeared in Anywa in a variant of Ethiopic script.

Luo distinguishes between dental and alveolar consonants (dental *t d* pronounced as in French; alveolar *ṭ ḍ* pronounced as in English). The language has a sing-song sound, with two pitch levels. Among other things, the tone pattern serves to distinguish between the complete and incomplete tenses of verbs. Anywa also has high and low tone, with a similar distinction in verb tenses: *àn ònák kàcì* 'I shall be hungry'; *àn ònàk kàcì* 'I am hungry'. Anywa vowels may be either clear or 'breathy'; in Luo a similar distinction of vowel register is between clear and 'hollow' (on vowel register see box at **KHMER**).

The first ten numerals in Anywa are: *aciel*,

*ariio, adak, angween, abiic, abiciel, abirio, abara, abingween, apaar*. As can be seen, the numeral base is '5': '6' is '5 + 1', and so on.

Shilluk is rich in metaphor. *Adúk góngó lùyì*, 'the grey one going under a pool', is a riddling term for a loaf which is ready to be baked under ashes.

'Anywa is spoken with the tongue forward in the mouth. One of the Nilotic customs of this tribe is to remove the lower incisor and canine teeth, and the characteristic position of the tongue at rest is lying against the lower lip in the space vacated by these teeth. This tongue position is difficult for some foreigners to master, and we seem to them to be talking "in our throats".'

Marie Lusted, 'Anywa' in *The non-Semitic languages of Ethiopia*  
ed. M. L. Bender (East Lansing:  
African Studies Center, Michigan  
State University, 1976) pp. 495–512

## Names and peoples

*Jo Luo*, 'Luo people', is the local name of the speakers of Luo, on the Kenyan and Tanzanian coast of Lake Victoria. They are known as *aba-Nyoro*, people of Nyoro, by their neighbours who speak the Bantu language **LUYIA**. English speakers sometimes called them *Nilotic Kavirondo* in contrast with the Luyia or *Bantu Kavirondo*. *Jo Luo* is also the local name of the speakers of Lwo, near Wau in Sudan: from this phrase derives *Jur*, their name in Arabic and in some English writings.

*Jo Anywa* is the name that speakers of another

of these languages, in Ethiopia and Sudan, use for themselves. 'They have been much oppressed by the neighboring tribes throughout the years; which has no doubt increased their own solidarity, and their name *anywa*, from *nywak* 'to share', reflects their practice of sharing food and all their belongings with the other members of the group' (M. Lusted, 1976). In Amharic and Oromo they are called *Yambo*, in Nuer and Dinka *Ber*. *Anuak* is an alternative English form.

*Shilluk* comes from an Arabic adaptation of *Colo*, the name that the Shilluk give to themselves.

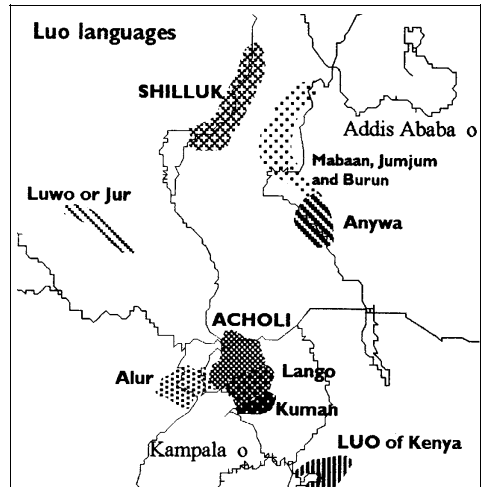
### Luo, Shilluk and their relatives

*Mabaan*, *Jumjum* and *Burun*, with possibly 75,000 speakers, are to be found in the plains of Dar Fung, in Sudan, between the two branches of the Nile.

*Shilluk* of Sudan has roughly 175,000 speakers, largely surrounded by territories where DINKA is spoken.

*Luwō* or *Jur* of Sudan has 50,000 speakers.

*Anywa* or *Anuak* of Ethiopia and Sudan has perhaps 66,000 speakers (an estimate by M. L. Bender). There are four dialects, *Adongo*, *Ciro*, *Lul* and *Opēno*, all close enough to be mutually



intelligible without difficulty. Some speakers consider *Adongo* to be 'good Anywa', the best dialect.

*Acholi*, with 700,000 speakers, *Lango*, with 800,000, *Kuman*, with 150,000, and *Alur* form the central group of Luo languages. The first three are all spoken in northern Uganda; Acholi extends into Sudan. Alur is spoken in Congo (Kinshasa) and Uganda by roughly 800,000 people. These four are sometimes regarded as a single language, *Lwo*.

*Luo* (sometimes distinguished as 'Luo of Kenya') has about 3,000,000 speakers in Kenya and 200,000 in Tanzania.

# LURI

**PERHAPS 4,000,000 SPEAKERS**

*Southern Iran*

**L**uri is one of the IRANIAN LANGUAGES, spoken by a mainly nomadic people in south-western Iran.

The mountain peoples of Luristan have a long history of fierce independence. Alexander the Great was among the many military leaders who have found themselves in danger in these mountains. 19th- and early 20th-century travellers tell of adventures and dangers in Luristan. Surprisingly little up-to-date information is to be found. The government of modern Iran is gradually imposing subjection and conformity on Luristan, and Luri is now heavily influenced by Persian, the national language. It is also full of Arabic loanwords. Mongolian loans include *tushmāl* 'chief of a clan'; *kūrān* 'encampment'.

'Down to the beginning of the 20th century, our knowledge of the Lur dialects was confined to 88 words collected by Rich, to four Bakhtiari verses in Layard [*Description of Khuzistan*] and to some thirty words collected by Houtum-Schindler. As late as 1901 we find the thesis stated that Luri is closely related to Kurdish and may even be described as one of its dialects. The merit of having first established the important fact that Kurdish and Luri are quite separate is due to O. Mann. Although there are Kurd tribes in Luristan, the true Lurs speak dialects which belong to the south-western Iranian group, like Persian and the dialects of Fars' (V. Minorsky, 'Lur' in *Encyclopaedia of Islam*, 2nd edn, vol. 5 (Leiden: Brill, 1986) pp. 821-6).

Luri is an unwritten language, but there is a rich oral literature. Some has been published, notably from the Bakhtiari dialect. There are fairy stories, love songs, wedding songs (*wāsi-*

*nak*), lullabies (*lālā'ī*) and epic poetry celebrating semi-legendary heroes.

For a table of numerals see KURDISH.

## The Lurs in 1932

'The handsomest people in Baghdad are the Lurs of Pusht-i-Kuh. They stride about among the sallow-faced city Shi'as in sturdy nakedness, a sash round the waist keeping their rags together, a thick felt padded affair on their backs to carry loads, and their native felt cap surrounded by a wisp of turban. They . . . sleep in the shade on the pavement, careless of the traffic around them, and speaking their own language among themselves: and you will think them the veriest beggars, until one day you happen to see them shaved and washed and in their holiday clothes, and hear that they belong to this tribe or that tribe in the mountainous region that touches Iraq's eastern border, and find that they are as proud, and have as much influence in their own lonely districts, as any member of a county family in his. They own three hostels, or "manzils", in Baghdad . . .

'Once a year the Lurs of Pusht-i-Kuh who work in the Baghdad custom house give a theatrical performance and show to a small audience the life and traditions of their province. There are bandits in white, with faces [covered] all except the eyes . . . there are songs on the high sobbing note like yodelling of the Alps.'

Freya Stark, *The valleys of the Assassins*  
(London, John Murray, 1934)

**Languages of southern Iran**

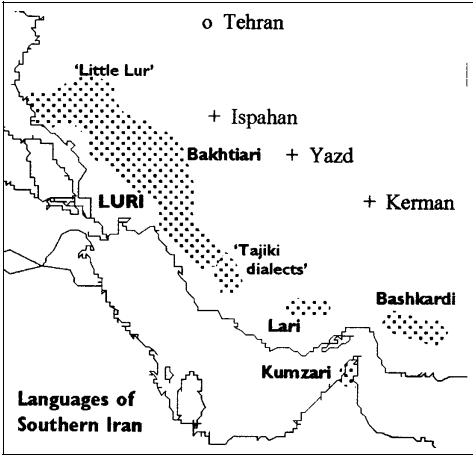
*Luri* or *Lori* is spoken in the regions of Fars, Khuzestan and Ispahan. The *Great Lur* dialects are those of Fars, the *Little Lur* those of Khurramabad.

*Bakhtiari*, the best-known Great Lur dialect, is spoken by as many as 650,000 seasonal nomads who occupy Bakhtiari and Ispahan in the summer and Khuzestan in the winter.

In the province of Fars several other dialects are spoken in enclaves surrounded by Persian. They have been called – confusingly – *Tajiki* dialects.

The old languages of the southern Persian cities are lost, supplanted by Persian, except for the dialects of the Jews of Ispahan, Yazd and Kerman and those of the Zoroastrians of Yazd and Kerman.

To the south-east, *Larestani* or *Lari* and the dialects of Bashakerd (*Bashkardi*) are also wholly



distinct from Persian. There is no clear information on the number of speakers of these dialects.

*Kumzari* is an Iranian dialect spoken at the tip of the Musandam Peninsula in northern Oman.

**Bakhtiari couplet**

Khīnum chi āu gulōu zi gar tukiste,	My blood like rosewater flowed from the mountains;
heme jūnum tash girih, dabem pukiste.	Fire seized my soul; my powder-flask is broken.
After I. M. Oranskij, <i>Les langues iraniennes</i> (Paris: Klincksieck, 1977) pp. 144–6	

# LUVALE

**600,000 SPEAKERS**

*Angola, Zambia, Congo (Kinshasa)*

**O**ne of the BANTU LANGUAGES, Luvale is spoken around the headwaters of the River Zambezi (see map at CHOKWE). As a result of colonial partition, Luvale speakers are split between the North-western province of Zambia and Moxico province of Angola, with a smaller number further north in Shaba province, Congo (Kinshasa).

*Luvale* and *Lovale* are the names used in Zambia for this language; its own speakers call themselves *vaLuvale* and their language *chiLuvale*. Balovale Province is named after them. In Angola the language is called *Luena* or *siLuena* and the people *vaLuena*.

The Luvale language and people, like the closely related Southern Lunda and Chokwe, whose territories lie to the east and north-east, have a traditional origin in the LUNDA empire of Mwata Yamvo. As this empire spread its rule southwards, in the 16th and 17th centuries, it is said that Lunda invaders mixed with longer-established inhabitants, speakers of Mbwele.

But Mbwele, Luvale and Chokwe are more closely related to one another than they are to Lunda, so the invaders may in truth have had little linguistic influence.

Many Luvale speakers keep cattle and grow cassava. They are known outside their own territory as traders in these products – and as brothel-keepers. Others live by fishing and hunting. Those initiated into these skills are given a second name, a ‘hunter’s name’, to be used when they are on an expedition.

Luvale is used as a second language by some speakers of Southern Lunda and CHOKWE. Luvale is also a close relative of the Luyi language once spoken by the LOZI, and still used by a small number of speakers immediately to the south on the Upper Zambezi. It is one of the eight official languages of Zambia.

The first ten numerals in Luvale are: *-mwe*, *-vali*, *-tatu*, *-wana*, *-tanu*, *-tanu na -mwe*, *-tanu na -vali*, *-tanu na -tatu*, *-tanu na -wana*, *likumi*. The correct noun class prefix, singular or plural, is inserted at each hyphen.

# LUXEMBURGISH

**300,000 SPEAKERS**

## *Luxembourg*

**T**he Grand Duchy of Luxembourg is one of the smallest independent states in Europe. It has a long history, going back at least to the 10th century, when Luxembourg was a feudal province of the 'Holy Roman Empire' of Germany. In later centuries the state belonged to Burgundy, to Germany, to the Spanish Netherlands, and to newly independent Belgium. In 1839 Luxembourg was divided from north to south: the western half, mostly French-speaking, became the Belgian province of Luxembourg, while the eastern became fully independent.

Luxembourg is now a trilingual state. French is the language of government; German is the language of business and of the press. But nearly all the people of Luxembourg speak another language at home: this is *Letzeburgisch*, 'Luxem-

burgish', a local dialect of German which has – quite recently – been recognised as the third official language. Children study all three at school, and Luxemburgish is heard alongside German and French on radio and television.

Scholars have traced the Luxemburgish dialect, or something very close to it, a thousand years back in occasional written documents. But there is only a limited literature in modern Luxemburgish, mostly 'folk' poetry. After the Second World War there was a local movement to drop Standard German entirely and to make Luxemburgish the main official language. A new spelling system was invented: *Letzeburgisch* was to become *Lezebuurjesh*. Nobody took any notice.

For a table of numerals see GERMAN.

# LUYIA

**3,750,000 SPEAKERS**

*Kenya, Uganda*

**L**uyia or Luhya forms a generally accepted grouping of BANTU LANGUAGES of western Kenya. They are the dialects of the region once called 'Bantu Kavirondo' and now buLuyia; their speakers may be called abaLuyia, and form the third largest language community in Kenya.

The early history of these languages is unknown. Oral tradition tells of settlement in this region only in the 15th century by peoples who had migrated over great distances, yet buLuyia is actually close to a very early centre of Bantu dispersion. There has been interaction with speakers of the neighbouring Nilo-Saharan languages LUO and KALENJIN, most noticeable in the Nyore, Logooli, Isukha, Idakho and Tiriki dialects: among these communities some clans are historically Kalenjin. It was shortly after 1900 that Luyia-speaking territory came under British rule.

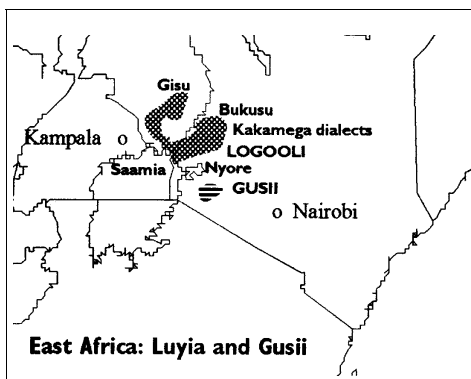
Luyia dialects are rich in folk-tales and legends, and narrative skill was traditionally much admired. Songs are incorporated at fixed points in the oral narratives.

In the early 20th century Christian missionaries began to write religious texts and Bible translations in several of these dialects. The Friends' African Mission was established at Kaimosi in 1902, and its home dialects were Logooli and Tiriki. The Church Missionary Society began work in 1912 and developed a unified dialect based on Wanga, Marama, Kisa and Tsotso, with elements of Samia and Khayo: this new literary form of Luyia was named luHanga. The Church of God worked in Nyore. The Roman Catholic missionaries used Wanga, Isukha, Bukusu and other dialects. The 'Lumasaba' or Gisu dialect was used by missionaries in Uganda, who began

work with this 'cannibal tribe' (as local rumour described it) in 1899. The first grammar appeared in 1904.

The newly written dialects were so close to one another that in the 1930s some missionaries, particularly of the Church Missionary Society, pressed for the adoption of a unified literary Luyia which could be used by all (or at least all in Kenya). But the dialects corresponded closely with local administrative divisions and had firm support among speakers: although mutually intelligible, they had distinctive local features which were already enshrined in writing. Much work was put in on Standard Luyia from 1942 for about twenty years, notably by a missionary linguist, Deaconess L. L. Appleby. However, two major dialect districts, Bukusu and Logooli-Tiriki, never accepted the proposed standard; elsewhere enthusiasm gradually waned, and schools never adopted Standard Luyia.

GANDA is the lingua franca for Gisu speakers in Uganda: their everyday speech includes Ganda loanwords as well as many from Swahili and



English. Swahili is the language of religious education in the many Islamic schools of buGisu.

Based on P. A. N. Itebete,  
'Language standardization in western Kenya'  
in *Language in Kenya* ed. W. H. Whiteley  
(Nairobi: Oxford University Press, 1974)  
xpp. 87–114 and other sources

**The Luyia dialects and Gusii**

*Logooli* or Maragoli is the southern dialect, and one that is often regarded as a separate language. Close to it, and mutually almost indistinguishable, are the three eastern or Kakamega dialects, *Isukha* (or Besukha), *Idakho* (or Bedakho) and *Tiriki*. Together these four have 500,000 speakers.

*Nyore* and *Saamia*, with their separate written traditions, are occasionally counted apart from the other central Luyia dialects.

The remaining central dialects are Nyala, Khayo, Wanga, Tachoni, Kabras, Tsotso, Marama and Kisa. The central dialects have 3,750,000 speakers.

The northern dialects of Luyia are *Bukusu* or *Vugusu* of Kenya (once called baKitosh) and *Gisu* or Kisu or Masaba of Uganda. Together these dialects have about 1,000,000 speakers. Although *luMasaaba* is often used as a name

for the Gisu language, the term *baMasaaba* really means 'people of Mount Elgon' – which is called *Masaaba* locally – and thus includes the speakers of Sebei, a Nilo-Saharan language related to KALENJIN.

GUSII, another Bantu language of Kenya, is sometimes considered very closely related to Luyia and particularly to Logooli.

**Numerals in Logooli and Isukha**

Logooli		Isukha
-ala	1	-ala
-bili	2	-bili
-bhaga	3	-bhaka
-ne	4	-ne
-tano	5	-rano
sita	6	sita
saba	7	saba
munane	8	munane
tisa	9	tisa
kumi	10	khumi

As in many other Bantu languages, a noun class prefix is added to the numerals '1' to '5'. Numerals above '10' are borrowed from Swahili.

R. A. Kanyoro, *Unity in diversity: a linguistic survey of the Abaluyia of western Kenya*  
(Vienna, 1983) p. 118

**Three Luyia written languages**

The written language in most of the dialect areas was that of the dialect in which the first materials had been written by the mission which was operating in that area – and not necessarily that of the area itself. In a dialect area which was under the influence of two or more missions, then two or more written forms of the language co-existed. In Bwisukha the Catholics, Friends' African Mission and Church of God operated:

The Son of God took on a human body

**Friends' African Mission Logooli form**

Mwana wa Nyasaye yavugula muvili gu mundu

**Catholic Isukha form**

Mwana wa Nasayi yabukula mubili ku mundu

**Church of God Nyore form**

Omwana wa Nyasaye yabukula omubili kw'omundu

Example from P. A. N. Itebete (1974)



# MABA

**PERHAPS 300,000 SPEAKERS**

*Chad*

**M**aba is one of a small group of NILO-SAHARAN LANGUAGES only distantly related to the rest. All are spoken in the highlands of eastern Chad and south-western Sudan (see map at KANURI). It is called by its speakers *Bura Mabang*.

Historically Maba is the most important of its group because it was the language of the Islamic state of Wadai, founded by a certain Abd al-Karim in the 17th century. From their old capital of Wara and their later capital Abéché, the Wadai

rulers encouraged trade and Islamic culture. Maba, language of the ruling class, and Arabic, language of religion, were the two lingua francas of this remote empire. Later, when the Empire of Bornu reduced Wadai to vassal status, KANURI became a third language of power there.

Maba is still widely known in south-eastern Chad and in the neighbouring borderlands of Sudan.

The first ten numerals in Maba are: *tek, bar, kungāl, asāl, tor, settāl, mindrī, ūya, adoī, atuk*.

# MACEDONIAN

**2,250,000 SPEAKERS**

*Macedonia, Bulgaria, Greece*

**M**acedonian was the last of the SLAVONIC LANGUAGES to gain official recognition – in 1944, when Yugoslavia was reconstituted as a federal republic. Macedonian dialects form part of the South Slavonic continuum that extends from Bulgarian to Slovene. Standard Macedonian is now the language of the independent Republic of Macedonia. It is written in the Cyrillic alphabet.

Macedonia (*Makedoniya*) takes its name from the ancient kingdom ruled in the 4th century BC by Philip II and his son Alexander the Great. The early core of this kingdom fell within the borders of modern Greece, but later conquests took in much of the valley of the River Axios or Vardar, thus partly coinciding with the modern Republic of Macedonia. Slavonic speakers settled here in early medieval times. Until the 20th century their language was usually called Bulgarian.

OLD SLAVONIC, the language of the religious texts through which Christianity spread among Slavonic speakers, was actually based on the Slavonic dialect spoken in the 9th century in Macedonia, in the country near Saloniki. Thus modern Macedonian may be regarded as today's nearest relative of the classical language of the Slavs. Ironically, it was the last of the modern Slavonic languages to develop a literary form.

In the valley of the Vardar, until the beginning of the 20th century, there were three languages of culture. Turkish ruled, for this was part of the Turkish Empire till 1912. Church Slavonic was used in religion. Greek was the language of Christian education. Also spoken were Albanian, Aromunian, Romani and – the language of the

majority – Macedonian. There was a great corpus of oral poetry in Macedonian, and in the 19th century some poetry was printed, but a movement to foster the local language in education and culture was slow to develop.

In 1912 the territory where Macedonian was spoken was split between Yugoslavia, Bulgaria and Greece. The national languages, Serbian, Bulgarian and Greek, were imposed in education and administration.

In spite of temporary shifts in policy Macedonian remains out of favour in Bulgaria and Greece. The minority in Greece numbers almost 200,000. In Bulgaria, speakers are officially regarded as Bulgarian and no census figures are available. However, the great majority of Macedonian speakers, in Macedonia itself, have since 1944 been able to regard their language as a fully fledged medium of national communication.

The Western dialects of Macedonian, spoken over most of the Republic of Macedonia itself (for map see SERBIAN), are the basis of the standard language. They have the unusual feature of not one but three definite articles – all suffixed to the noun, as in other Balkan languages – to indicate distance from the speaker: *vol* 'ox', *volot* 'the ox', *volov* 'the ox here', *volon* 'the ox there'. For a table of numerals see SLOVENE.

## Ancient Macedonian

Little is known of the language of ancient Macedonia. The royal family spoke Greek – the elite language of the whole region – and patronised Greek culture. Their official genealogy traced their descent from Greek heroes. *Alexandros* and *Philippos* are actually Greek names. But plenty of

Macedonian rulers and military figures had non-Greek names, and Greek never spread widely among the ordinary people of Macedonia. Alexander's Macedonian troops got their commands in 'Macedonian', and some words of this language survive in Greek historical texts and glossaries.

Macedonia was annexed by the Roman Republic in the 2nd century BC. Ancient Macedonian was eventually supplanted by Latin, and it has no connection with modern Macedonian.

The political ideology of modern Greece sets store by its ancient heritage. The view accepted by most Greek historians is that Philip and his family 'were Greek' (after all, they spoke Greek) and that, in any case, ancient Macedonian was a form of Greek. In fact the recognised ancient Greek dialects, such as Doric, Ionic and Aeolic, were already so different from one another that, looking only at their forms and setting aside their cultural interconnections, modern linguists might well have defined them as separate languages. Macedonian was very different again from any of these. Many scholars outside Greece prefer to say that it was Indo-European (like other languages of the ancient Balkans) but not Greek.

# MADURESE

**9,000,000 SPEAKERS**

*Indonesia*

**M**adurese is one of the AUSTRONESIAN LANGUAGES, spoken on the island of Madura, north-east of Java, and on nearby smaller islands.

It is usually said to be very close linguistically to Javanese. Linguists have disputed the precise relationships among the languages of the Greater Sunda Islands. What is in no doubt is the long history of mutual interaction between Javanese and Madurese. They share many aspects of culture – including the traditional script or *ak-sara*, ultimately of Indian origin, and the separate ‘formal’ and ‘informal’ registers of normal speech. Within the formal register, speakers mark, by the choice of a whole range of words, the relative social status of themselves and their

hearers. The words of the formal register are almost all borrowed from Javanese. The same is true in Balinese, incidentally, and it indicates that the social levels so carefully distinguished were learnt, by speakers of both languages, from speakers of Javanese.

Apart from these loanwords Madurese has borrowed quite extensively from Malay, as well as from Arabic and Dutch.

Madurese is now written in the Latin alphabet, as used for Indonesian, more often than in the old script. The first ten numerals in Madurese are: *settong*, *dhuwa*, *tello*, *empa*, *lema*, *ennem*, *petto*, *ballo*, *sanga*, *sapolo*. For a script table and a map see JAVANESE.

# MAGAHII

**PERHAPS 20,000,000 SPEAKERS**

*India*

**T**he three INDO-ARYAN LANGUAGES of Bihar, Bhojpurī, Maithilī and Magahī, are often thought of as dialects of HINDI (see map there). No statistics are available: the census reports this language as ‘Hindi’. Yet in many ways it is closer to Bengali, its eastern neighbour. For a table of numerals see MAITHILĪ.

Magahī is spoken in the southern half of Bihar, and extends across the border of West Bengal and Orissa, where the Eastern Magahī dialect (*Pañcparganiyā* ‘of the five districts’) shows Bengali and Mundari influence. If a standard Magahī can be recognised, it is that of the cities of Patna and Gaya.

There is no real tradition of written literature

in modern Magahī. Educated speakers will tend to incorporate Hindi words and turns of phrase in their speech, and to use standard Hindi on formal occasions.

For a table of numerals see MAITHILĪ.

*Magadha* has been the name of this region for over two thousand years, and gave its name to *Magadhī*, one of the Prakrit dialects of ancient India. This is regarded as the linguistic ancestor of six modern languages – Bengali, Assamese, Oriya and the three languages of Bihar. A typical sound change of Magadhī was the shift of *dh* to *h*, of which the modern language name itself provides an example.

# MAGINDANAO

**1,000,000 SPEAKERS**

*Philippines*

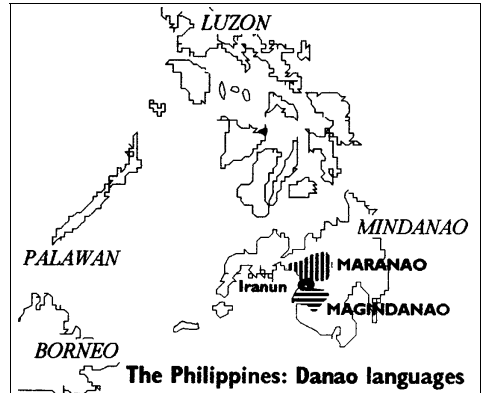
**T**he Danao languages – Magindanao, Iranun and Maranao – are spoken in south-western Mindanao, in the Philippines. All three are languages of Muslim peoples; together, their speakers make up the largest Muslim community in the Philippines.

The history of these peoples is traced in the 19th-century *tarsilas*, manuscripts containing genealogies that go back five hundred years to the arrival of Islamic proselytisers in Mindanao. Even then, it seems, there were three adjacent peoples here, linked by trade relations through the speakers of Iranun, who were perhaps always the intermediaries, and whose dialect shows the influence of both the others.

The first ten numerals in Magindanao are: *sa, dua, talu, pat, lima, nm, pitu, walu, siaw, sa-pulu*. In Maranao ‘1’ is *isa* and ‘10’ is *sa-wati*. In Iranun ‘3’ is *tulu*.

## The Danao languages

*Magindanao*, with 1,000,000 speakers, is the language of the Pulangi river basin, often inundated



– hence their name, which appears to mean ‘liable to flood’. It was originally the name of a town founded by the Iranun at the site of modern Cotabato City.

*Iranun* or *Ilanun* is spoken on the coast of Illana Bay and to some distance inland. There is an émigré community of Iranun speakers in Sabah, Malaysia.

*Maranao* has 600,000 speakers around Lake Lanao. The name means ‘lake people’.

# MAITHILI

**PERHAPS 22,000,000 SPEAKERS**

*India, Nepal*

**L**ike its neighbours Bhojpuri and Magahi, Maithili is one of the INDO-ARYAN LANGUAGES, and is sometimes regarded as a dialect of HINDI (see map there). Many speakers call their language Hindi, and Hindi is the usual language of education and literature.

Maithili is spoken in the northern half of the Indian state of Bihar and across the border into Nepal, where it is the second most widely spoken language of the country. Maithili was first recognised as a separate language in G. A. Grierson's *An introduction to the Maithilī language of North Bihār* (1881–2).

The name is a recent invention derived from *Mithilā*, a semi-legendary city and kingdom of classical Sanskrit literature. The term *Bihari* has been used (for example, in the *Linguistic survey of India*) to cover all three languages, Maithili, Magahi and Bhojpuri. There is, however, no particular reason to treat them as one.

Indo-Aryan speakers are thought to have been already dominant in Mithilā by the 6th century BC – the time of the Buddha, who lived and preached here. In succession to SANSKRIT and Magadhī Prakrit, *Avahattha*, a mixed language containing Apabhramsa and Maithili elements, was once used for literature. The oldest literary text in an identifiable Maithili is the 14th-century *Var ṇa-ratnākara*, ‘a sort of

lexicon of vernacular and Sanskrit terms, a repository of literary similes and conventions dealing with the various things in the world and ideas which are usually treated in poetry’ (S. K. Chatterji).

Maithili was once written in *mithilaṣar* script, resembling that of medieval Bengal; it is still in use in some religious contexts. The language is nowadays normally written in Devanagari script, like Hindi. Like Bengali to the east, Maithili has lost the gender distinction in nouns (though local grammars, under Hindi and Sanskrit influence, still claim to distinguish masculine and feminine genders). Its vocabulary shows the influence of the Austroasiatic language SANTALI, which is spoken in the southern hills.

## Numerals in Bhojpuri, Maithili and Magahi

	<b>Bhojpuri</b>	<b>Maithili</b>	<b>Magahi</b>
1	ēk	ekə	ek
2	dūi	dui	dū
3	tīni	tini	tīn
4	cāri	cari	cār
5	pāñc	pañcə	pāñc
6	chav	chəv	chau
7	sāt	satə	sāt
8	āṭh	aṭhə	āṭh
9	nav	nəv	nau
10	das	dəsə	das

# MAKASAR

1,600,000 SPEAKERS

*Indonesia*

**M**akasar, one of the AUSTRONESIAN LANGUAGES, is the language of the old Sultanate of Gowa, at the south-western tip of Sulawesi (Celebes). The city of Ujungpandang, near old Makassar, is still an important stopping place on the sea routes among the islands of Indonesia.

Makasar traders have been long distance travellers. Makasar words have been identified in AUSTRALIAN LANGUAGES, evidence of contacts established before Europeans began to explore Australia. Ujungpadang remains a multiethnic market centre, used by numerous speakers of Chinese, Malay and the many local languages of Sulawesi. Makasar remains the lingua franca of the city. Fishing is a source of wealth here: in the countryside rice is grown in the irrigated fields of the coastal strip.

Makasar is one of a group of related dialects, all spoken in south-western Sulawesi. Its neighbour BUGIS (see map and table of numerals there) stands somewhat apart linguistically; Mandar, to the north-west, belongs to a third subgroup. In

all three languages there is a traditional literature of epic tales and chronicles, recorded in Lontara script, retelling the adventurous history of the small kingdoms of Sulawesi.

## Lontara script

Lontara script, originating like so many others from the Brahmi alphabet of ancient India, has been the traditional form of writing for Buginese, Makasarese and Mandar. In Bugis it was once used for laws, treaties, trading contracts and maps, but is now confined to purely ceremonial uses – it can be seen in use at marriage ceremonies, for example. In Makasar, however, the script is still used by young and old, for such things as personal letters and students' notes, though officially the Latin alphabet, as used for Indonesian, is favoured.

*Lontar* is the Malay name of the palmyra palm, whose leaves are traditionally used for manuscripts in India, south-east Asia and Indonesia.

### The Lontara script

ka ga nga nka pa ba ma mpa ta da na nra ya ra la wa sa a ha



# MAKONDE

1,100,000 SPEAKERS

*Tanzania, Mozambique*

**M**akonde is spoken in south-eastern Tanzania, along the Ruvuma river, and also on the northern edge of Mozambique (for map see MAKUA). Its speakers are traditionally farmers, growing maize and sorghum on the dry but fertile plateaus of the region. They also sold rubber, gum-copal and ivory at the market of Mikindani, an important slaving port. They suffered repeatedly from slave raids and invasions in the 19th century, until the Sultan of Zanzibar established peace in 1876.

The Makonde of Tanganyika joined in the unsuccessful Majimaji rebellion against the invading Germans in 1905. Makonde is now the second largest regional language of Tanzania. The Makonde of Mozambique were among the staunchest fighters for the liberation movement of the 1960s led by Frelimo.

Speakers call their language *chiMákonde*. The Portuguese spelling is *Maconde*.

Like some other BANTU LANGUAGES of the Tanzanian coast, Makonde has word stress instead of the tones that are such a prominent feature in most languages of the family. It shows Swahili influence – though its speakers, unlike some of their neighbours, have not adopted Islam.

In turn, *kiMwani*, a local form of Swahili spoken on the coast of northern Mozambique, is much influenced by Makonde.

## The secret songs

'The girls' initiation is called *ciputu*. The girls of a village are initiated together; they are taken to the bush by a chief instructress, each girl assisted by a sponsor [a sister, aunt or friend]. Each mother, when she comes to take her daughter home, receives a burning brand from a fire which has previously been carried round the village in procession. With this she lights new fire in her hut. After the initiation each girl is semi-secluded for a month. At the end of the time there is a reunion in the *ciputu* house, the girls are bathed and taught secret songs; beer is brewed for the sponsors, and when the sponsors are finally paid they reveal the meaning of the secret songs.'

Mary Tew, *Peoples of the Lake Nyasa region* (London: Oxford University Press, 1950) pp. 28–9

## Two Makonde riddles

*Kunchuma nchagwa wangu ku Nambela, ndila kupita alila* – 'I bought a slave in Tanganyika: he walked along crying': the bells on a dancer's costume.

*Chindjidi nkanywa mene kulova* – 'What goes into the mouth does not get chewed': tobacco smoke.

M. Viegas Guerreiro, *Os Macondes de Moçambique*, vol. 4 (Lisbon: Junta de Investigações do Ultramar, 1966) p. 328

# MAKUA

**3,500,000 SPEAKERS**

*Mozambique, Malawi, Tanzania*

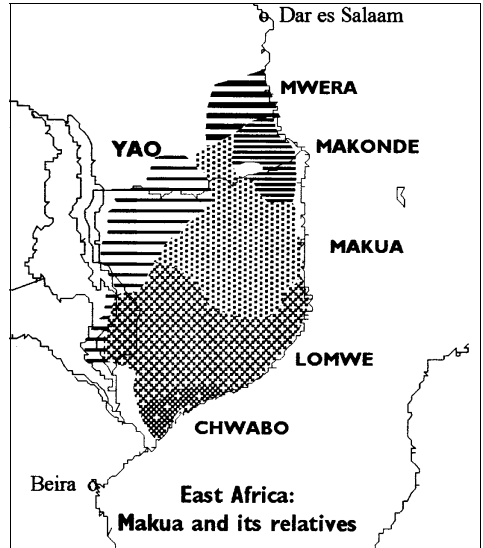
**M**akua or *eMakua* is a major language of northern Mozambique, and serves as a lingua franca for speakers of minority languages in its neighbourhood. It is one of the BANTU LANGUAGES.

Traditionally a farming people, Makua speakers controlled trade routes between the southern end of Lake Nyasa and the Indian Ocean. They dealt with Swahili and Gujarati traders at the coast: their stock was groundnuts, sesame seeds, castor-oil seeds and maize, and they were prominent in the slave trade. Those living near the coast adopted Islam under Swahili influence.

## Makua and its relatives

One subgroup of the Bantu languages of Mozambique consists of Makua with LOMWE (2,000,000 speakers, half in Mozambique, half in Malawi) and *Chwabo* or Cuabo (650,000).

A neighbouring group, not quite so closely related, consists of MAKONDE (1,100,000 speakers,



Tanzania and Mozambique) with *Mwera* (350,000, Tanzania) and YAO (1,200,000 in Malawi, Tanzania and Mozambique).

# MALAGASY

10,000,000 SPEAKERS

## *Madagascar*

**M**alagasy is the national language of Madagascar. The geographically unexpected link between it and the AUSTRONESIAN LANGUAGES was noted as long ago as the 18th century.

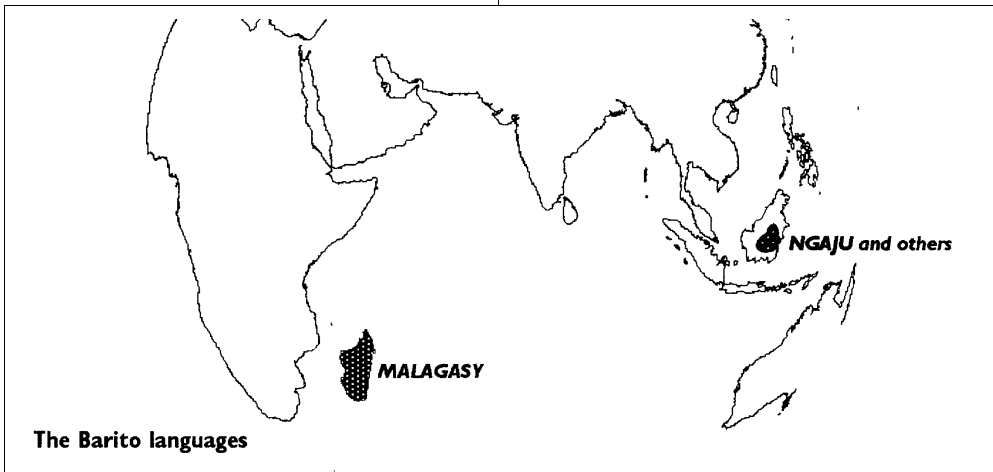
Linguistic comparison has now shown beyond a doubt that Malagasy is a language of the otherwise obscure South-east Barito group: its linguistic relatives are thus to be found in a region of southern Kalimantan (Borneo) in Indonesia. It is not clear when or why migrants from this area colonised the great island of Madagascar, but this is what must have happened. Otto Christian Dahl, who demonstrated the link in 1951, suggested that Madagascar was colonised by traders from the Hindu kingdom of southern Borneo around the 4th century AD.

Certainly Malagasy now shows the influence of Malay, Javanese and Sanskrit, evidence of early involvement in the developed culture of the Indonesian islands. Easily traceable, too, are

loanwords from Swahili and Arabic – the trading languages of the East African coast – Ngazija of Grande Comore, French of the former colonial rulers of Madagascar, and English of the pirates whose headquarters it was in the 18th century. Loanwords include *dité* ‘tea’ from French *du thé* ‘some tea’; *mompera* ‘priest’ from French *mon père* ‘my father’, used in addressing Catholic priests.

Written literature in Malagasy goes back to the 15th century, when the Arabic *ajami* script was used for astrological and magical texts. In 1823 King Ramada I decreed that the Latin alphabet should be used, and Christian missionaries set to work on translations from the Bible.

Malagasy has a complex system of personal and demonstrative pronouns, varying with the distance between the speaker and the person or thing denoted. The first ten numerals in Malagasy are: *iray, roa, telo, efatra, dimy, enina, fito, valo, sivy, folo*.



**Place names of Madagascar**

Malagasy forms long compound words easily. Nouns referring to a place are prefixed *an-*. These two facts explain the striking form of many names on the map of Madagascar. The capital city, *Tananarive* in French, is more correctly called *Antananarivo*, 'the village of a thousand'.

**Malagasy and the Barito languages**

*Merina* 'the high people', dialect of the central highlands, is the basis of standard Malagasy. There has been controversy over the grouping of other dialects, but *Sakalava*, of the west coast, is also important.

The Barito languages of Kalimantan include several small language communities – and *Ngaju* or Land Dayak, which has a quarter of a million speakers: see also box at IBAN.

**Ibonia: epic hero of Madagascar**

Hoy Iboniamasoboniamanoro:  
 'Izaho no taranaky ny omby mahery.  
 Raha tezitra aho  
 sokina an-tampon-doha,  
 mitraka aho vaky ny lanitra,  
 miondrika aho vaky ny tany,  
 hitsahiko ny hazo mibaraingo,  
 hitsahiko ny tany mitriatra,  
 hitsahiko ny tanitra manao vara-maina.'

He of the clear and captivating glance said:  
 'I am the descendant of powerful bulls.  
 When I am angry  
 The top of my head bristles,  
 When I raise my head the sky bursts open,  
 When I bend the earth bursts open,  
 When I trample on trees they twist,  
 When I trample on the earth it splits,  
 When I trample on the sky it thunders.'

The Merina people of inland Madagascar have rich historical and poetic traditions. The 'Royal histories of the Merina', *Tantaran'ny Andriana*, collected by François Callet in 1864, are said to be the most comprehensive oral history ever collected in any part of Africa.

Many tales and poems cluster around the figure of Ibonia. The epic of Ibonia, collected and published by the Norwegian missionary Lars Dahle in 1877, is the undisputed classic of Malagasy literature. The text quoted here is from a shorter folk poem, *hainteny*, which alludes to the epic hero.

*Ibonia* tr. Lee Haring (Lewisburg: Bucknell University Press, 1994);  
 text from Jean Paulhan, *Les hain-teny merinas* (Paris: Geuthner, 1913)  
 pp. 268–9; translation from Leonard Fox, *Hainteny: the traditional  
 poetry of Madagascar* (Lewisburg: Bucknell University Press, 1990) p. 215

# MALAY, INDONESIAN AND MALAYSIAN

**PERHAPS 35,000,000 SPEAKERS**

*Indonesia, Malaysia, Singapore, Brunei, Thailand*

**F**or at least a thousand years Malay has been the principal lingua franca of 'insular south-east Asia', of the great Malay Archipelago which is now politically divided between Indonesia and Malaysia.

The lands on either side of the Straits of Malacca have been called the Malay country for all this time. The origin of the name is uncertain, though it may possibly come from a Dravidian word by way of Sanskrit *Malaya* 'mountain'. Malay is locally known as *bahasa Melayu* 'Malay language' (*bahasa* deriving from Sanskrit *bhāṣaḥ* 'language'); its speakers are *orang Melayu* 'Malay people'.

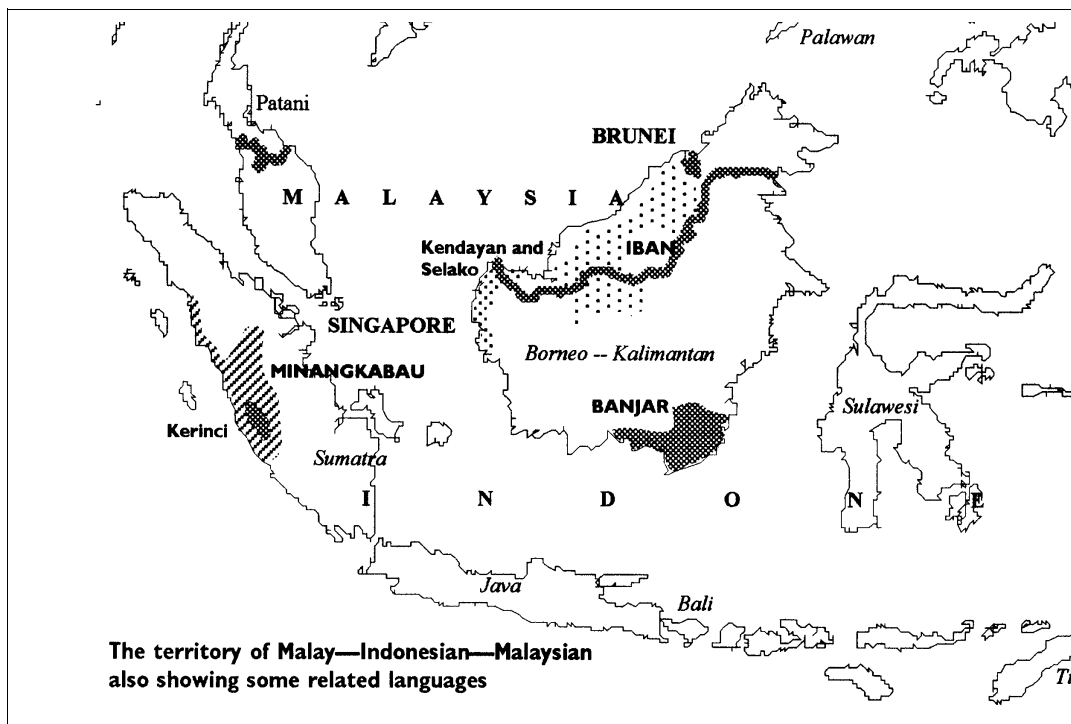
For the designated national language of independent Indonesia the old name, redolent of the British-controlled Malay peninsula, was thought inappropriate. So this form of Malay is called *bahasa Indonesia* 'Indonesian language', a term introduced around 1930. In due course independent Malaysia followed suit with *bahasa Malaysia* 'Malaysian language'. *Malay* is still a term that we cannot do without – for Singapore and Brunei, and for the many regional and non-standard varieties of the language.

Malay is one of several related AUSTRONESIAN LANGUAGES that originate in the arc between western Sumatra and eastern Borneo: they are now believed to have come here, three thousand years ago or more, as the result of gradual, prehistoric migration from the Philippines, and ultimately from Taiwan.

Two other members of the group are MINANG-KABAU and IBAN. These two, also, have spread as vehicles of trade and water travel – but Malay itself, from its early homeland on either side of the Straits of Malacca, has spread much further.

Old Malay is first recorded in the early inscriptions of the kingdom of Śrīvijaya. They are found near Palembang in southern Sumatra and on the nearby island of Bangka, and are dated to 683–6. Later inscriptions come also from the Malay peninsula, just across the straits. Singapore's history might have begun with a 12th-century inscription – but the British, frightened that it would excite anti-colonial feeling, destroyed it before it had been deciphered. By the 12th century 'Malaya' – perhaps the same kingdom, under a different name – was well known to Arabic and Chinese travellers, from whom even Marco Polo heard of it. Its language had certainly already become a lingua franca of trade in the archipelago. In a different form, 'Classical Malay', it was to be a language of literature, rich in historical texts.

In the Malay peninsula the great trading city of Malacca, founded in 1403, was briefly dominant throughout the region. It retained some importance even when ruled by European powers, successively Portuguese, Dutch and British. In Java the old capital of Jakarta fell to the Dutch in 1619 and for more than three centuries was renamed *Batavia* after a Latin name sometimes used for the Netherlands. It was in the early 19th century that the British made inroads into Dutch influence in the archi-



pelago. The position of both colonial powers was undermined by Japanese occupation in the early 1940s. Indonesia became independent in 1945 and Malaysia in 1957. Thus the long term result of British–Dutch competition is the modern divide between Malaysia, formerly a group of British territories and protectorates, and Indonesia, formerly the Dutch East Indies.

Malay, in slightly different forms, became the national language of both states.

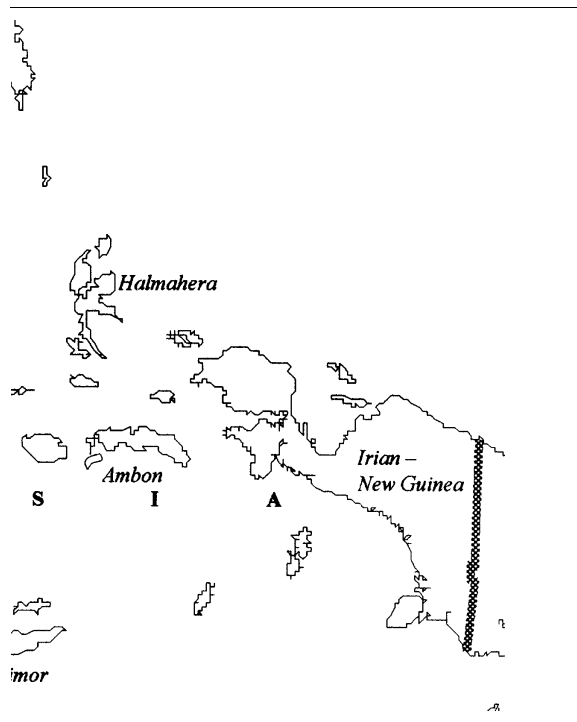
In Malaysia it is the majority language, though there is also a large Chinese-speaking population and some significant indigenous linguistic minorities. In Indonesia the number of regional and local languages is very large indeed. Javanese is actually the largest linguistic community, but Malay, or rather *bahasa Indonesia*, is the only language with the potential to draw the country together, a potential that has been energetically tapped by the resolutely nationalistic policies of Indonesian governments.

The Śrīvijaya inscriptions are in an Indian script. With the arrival of Islam, Arabic (*jawi*) script was being used for Malay by the 14th

century. Under British and Dutch influence Latin script has now become standard.

### Malay at the crossroads

Malay is a spectrum of many dialects and registers. There are the formal, official languages of Indonesia, Malaysia and Brunei; the varieties used in the press, broadcasting, schools and religion in these three countries and in Singapore; the everyday colloquial of great cities including Kuala Lumpur and Jakarta (whose local speech is sometimes called *Batawi*); the lingua franca used in towns and markets by people with different mother tongues, *Bazaar Malay*; the local speech of scattered trading ports and colonies, most of them bilingual, from *Ambon Malay* to *Sri Lankan Creole Malay* (which alone has 50,000 speakers); major regional varieties including *Banjar* of eastern Kalimantan and *Patani Malay* of southern Thailand; local varieties spoken by peoples far from the mainstream of Malay culture, including the ‘aboriginal’ Malay of inland districts of the Malay peninsula. Tin-



miners of Malaya traditionally used a 'secret language', a special form of Malay with arcane vocabulary, to avoid offending the spirits that guard the ore. *Baba Malay* is a term sometimes used for the pidginised Malay, strongly influenced by Min (see CHINESE), that is spoken by communities of southern Chinese origin in Malacca and other Malaysian cities.

Malay is a language of state education throughout Indonesia and Malaysia. It is also the traditional language of religious education for some Muslim peoples of the archipelago and for the CHAM Muslims of Cambodia.

Malay is still quite recognisably a member of the Malayic group of Austronesian languages: but in some ways it is now rather different from the rest. A lingua franca needs to be easy to grasp, and Malay has a more approachable structure than its relatives. It has undergone extremely varied external influences, with loanwords from many local languages of the archipelago (notably Javanese), from Sanskrit, Chinese, Arabic, Portuguese, Dutch and English. *Ketam* 'crab' comes from an Austroasiatic

language, cf. Khmer *ktam*, Mon *gatam*. *Pasar* 'market' comes from Persian *bāzār*, which has been borrowed into many languages of the world. A 'fair' is picturesquely called *pasar malam*, 'night market', in Malay. The black market is *pasar gelap*, 'dark market'.

Local forms of Malay have other loanwords. In Larantuka Malay *nyora* 'woman' comes from Portuguese *senhora*; *om* 'uncle' comes from Dutch *oom*.

As the one essential language for trade and administration in the archipelago, Malay has contributed many loanwords to English. They include *compound*, which with the meaning of 'yard, enclosure' comes from Malay *kampung* 'enclosure, quarter of a city'; *gong*; *rattan*; *sago*; *orang utan*, literally 'people of the forest'.

### The Malayic languages

Standard MINANGKABAU is based on the dialect of Padang. Four dialect groups are usually distinguished: Tanah Datar, Limapuluh Kota, Agam, Pasisir.

*Kerinci* has 300,000 speakers in inland Sumatra, on the slopes of mount Kerinci.

*Banjar*, the major eastern Kalimantan dialect of Malay, sometimes regarded as a separate language, has around 2,000,000 speakers centred on the city of Banjarmasin. It shows strong Javanese influence.

The speakers of IBAN and related 'Ibanic' or 'Malayic Dayak' languages (including *Kendayan* and *Selako*) are to be found along the rivers of western Sarawak and western Kalimantan.

*Malay* itself began its spread, perhaps two thousand years ago, from the lowlands of central eastern Sumatra, southern Malaya and south-western Borneo. On the mainland it extends to the southern Thai province of Patani. By sea, Malay gradually reached as far as Sri Lanka, the Mergui archipelago (where it is spoken by the so-called Sea Gypsies) and the northern coast of New Guinea.

Local varieties of Malay have grown up at several points in the archipelago. Each has its history, often several centuries long, and its own blend of loanwords which may come from Por-

tuguese, Dutch, Chinese, local languages, and now from standard Indonesian as well. *Larantuka Malay*, *Ambon Malay*, *Kupang Malay*, are among better-known examples. Speakers are now usually bilingual in Indonesian, which they use for more formal and prestigious purposes.

**Languages of Timor**

In a multilingual archipelago, Timor is one of the most multilingual of islands.

The Austronesian languages of Timor belong to the Timor-Flores group. They include *Timor*

(650,000 speakers), *Tetun* (300,000), *Galoli* (50,000) and *Mambai* (80,000). *Rotinese* or Roti (120,000) is spoken both on Roti itself and on the nearby western tip of Timor. *Lamaholot* (300,000 speakers) is a language of Solor, Lomblen, Pantar and Alor islands.

The Timor-Alor-Pantar family of languages is quite independent of Austronesian: it may have distant relatives in north-western Irian Jaya. It includes *Makasai* (70,000 speakers), *Bunak* (50,000) and *Kolana* (50,000), the majority language of Alor island.

Numerals in Malay					
	Minangkabau	Malay	Jakarta Malay	Banjar	Iban
1	cie'	satu	atu	asa	satu
2	duo	dua	duɛ	dua	dua
3	tigo	tiga	tigɛ	tiga	tiga
4	ampe'	empat	əmpat	ampat	ampat
5	limo	lima	limɛ	lima	lima'
6	anam	enam	ənəm	anam	nam
7	tujuəh	tujuh	tuju'	pitu	tujoh
8	(sa)lapan	(de)lapan	dələpan	walu	lapan
9	sambilan	sembilan	səmbilan	sanga	semilan
10	sa-puluəh	sepuluh	sə-pulu	sa-puluh	sa-puloh
From K. Alexander Adelaar, <i>Proto Malayic: the reconstruction of its phonology and parts of its lexicon and morphology</i> (Canberra: Australian National University, 1992) and other sources					



# MALAYALAM

22,000,000 SPEAKERS

*India*

**M**alayalam is one of the DRAVIDIAN LANGUAGES of India, the state language of the mountainous south-western coastal state of Kerala (for map see TAMIL).

*Malei* means 'hill country' in Tamil: Kerala was 'Male, where pepper comes from' to a Greek geographer of the 6th century. The name *Malabar*, evidently derived from this native word, was applied to the region in Persian and then in European geographical texts. It is now old-fashioned, but the related form *Malayalam* 'mountain region' has remained as a name for the language.

Until around AD 1000 Malayalam was not a separately identifiable language. Early inscriptions from Kerala are in Old Tamil, which was spoken both in Kerala and in the state now called Tamilnadu.

From that date onwards, Tamil and Malayalam have grown apart – most obviously in the matter of loans from Sanskrit. Later Tamil, unusual among the cultivated languages of India, has not been receptive to Sanskrit loans. Malayalam is so full of Sanskrit words, and even of Sanskrit forms and syntax, that the two are inextricably interwoven. The acknowledged masterpieces of Malayalam literature are the 16th-century versions of the Sanskrit *Mahābhārata* and *Rāmāyaṇa* by Tuñcattu Eḷuttacān. But even in spoken Malayalam Sanskrit words are to

be heard in practically every sentence.

Numerals in Malayalam and Tamil			
Malayalam			Tamil
oru	1		oṇṇu
raṇṭu	2		ireṇṭu
mūnnu	3		mūṇu
nālu	4		nālu
añju	5		añcu
āru	6		āru
ēru	7		ēlu
ettu	8		eṭṭu
ompatu	9		ompatu
pattu	10		pattu

## Malayalam script

Malayalam script achieved its present form in the 17th century. Although showing a clear underlying resemblance to the Tamil-Grantha alphabet it looks very different in print, heavy and blockish where Tamil is light and angular. Some letters are used mainly in Sanskrit loanwords.

Malayalam is spoken by sizeable communities in Singapore and Malaysia. Here the Arabic script is in regular use: in Kerala, too, among Muslims, Malayalam is occasionally printed in Arabic script.

## Malayalam consonants

കഖഗഹഘന ചടഇയയര ടഥഡന തഥദധ ന പ ഫ ബ ഭ മ യ റ ല ള ഴ റ്റ ള്ള ഷ സ ഹ

k kh g gh ṇ c ch j jh ṇ̣ ṭ ṭh ḍ ḍh ṇ̣ t th d dh n p ph b bh m y r l ḷ ṣ ṣh

# MALTESE

**350,000 SPEAKERS**

*Malta*

**O**ne of the SEMITIC LANGUAGES, Maltese originates in the Arabic conquest of the island of Malta in AD 870. It is thus a direct offshoot of ARABIC, but is now very distinct from its parent because of Malta's later history.

Some now prefer to replace the English form *Maltese* with the local *Malti*.

European political influence and Christianity arrived with the Norman conquest in 1090. For more than three centuries, from 1530 to 1798, Malta was the stronghold of the former crusading order, the Knights Hospitaller of St John. The island was taken by Britain in 1800 but regained independence in 1964–79. Through all this period the local tongue survived and flourished, though it was not recognised as the national

language either by the Knights (who used Latin and Italian) or at first by the British (who worked hard to replace Italian with English). Malta, staunchly Catholic, is now officially bilingual in Maltese and English, while Italian retains a powerful influence.

Maltese is written in the Latin alphabet with four extra letters, *ċ ġ ż ħ*. It has many Italian loanwords: so many that some used to claim it as a dialect of Italian. English loanwords are fewer. Two years of French rule, 1798–1800, bequeathed two essential greetings to Maltese: *bongu* and *bonswa* from *bonjour* and *bonsoir* 'good morning, good evening'.

The first ten numerals in Maltese are: *wieħed*, *tnejn*, *tlieta*, *erbgha*, *ħamsa*, *sitta*, *sebgħa*, *tmienja*, *disgħa*, *għaxra*.

# MANDEKAN

**5,000,000 SPEAKERS**

*Mali, Senegal, Guinea,  
Ivory Coast, Burkina Faso, Sierra Leone*

**M**andekan, one of the Mande group of NIGER-CONGO LANGUAGES, is, historically and culturally, among the most important languages of the world. Its extent matches, and sometimes goes beyond, the borders of the medieval Islamic Empire of Mali. Five hundred years ago, as the old empire faded away, Mandekan began to split into dialects, each of which has sometimes been counted as a separate language.

This article deals with them together: for if political disintegration made the dialects grow apart, Mandekan songs and epics, transmitted by itinerant poets, slowed down the process of change, kept archaic forms of language in current use, and aided mutual understanding across this vast area.

*Mandekan* is now the preferred term among linguists for the group of dialects sometimes called Manding or Mandingo. Local versions of this same term include *Malinke*, *Manenka*, *Mandinka*. The variety called *Bambara* is known to its own speakers as *Bamana-koma*, *Bamananke* being their name for themselves. *Dyula* is also spelt *Jula*.

Mandingo – under these various names – is the national language of Mali, a major language in four other countries, and the vehicle of a great oral literature. Why has almost nobody outside Africa heard of it? It has too many names; its territory has long been fragmented, by the succession of native empires and the now-fossilised colonial frontiers; and the history and literature of Africa is still far too little known.

Several thousand years ago, dialects ancestral to the Mande and Atlantic groups of Niger-

Congo languages must have arrived here in inland West Africa. The two did not necessarily arrive at the same time. The Atlantic group is represented notably by FULANI. ‘Proto-Mande’ perhaps began to divide into the future Mande language group around 2000 BC. The modern Mande languages include MENDE, SUSU, SONINKE and Mandekan.

Much later, with political developments culminating in the Empire of Mali, Mandekan itself spread widely across the region, and began to split into dialects, two of which are now represented by Bambara and Dyula.

Most recently, Bambara and Dyula themselves, originating as regional dialects of Mandekan, have spread with trade and politics.

## Language of the Mali Empire

A thousand years ago the Empire of Mali was founded. Its most powerful ruler, Kankan Musa (Mansa Musa: reigned 1312–37), held sway from the upper Senegal across to the middle Niger, and from the Sahara to the rainforest. Like many of his subjects, Kankan Musa made the long pilgrimage to Mecca, and his capital was a centre of Arabic learning and science.

Arabic was the language of high culture, religion and diplomacy, but Mandekan was used throughout the empire in administration and trade. Indeed it travelled further, on trade routes southwards into the forests and northwards across the Sahara. Some Mandekan words are recorded by Arabic authors who had travelled in the Western Sudan in the 12th to 14th centuries.

When Europeans began to visit the West African coast in the 15th century they found that Mandekan was the most useful lingua franca between the River Gambia and the Rio Grande.

The empire fell before the rising power of the SONGHAY under Askia Mohammed. With the end of political unity, the widely spread dialects of Mandekan were no longer linked to a central standard, and since the 16th century they have diverged more and more.

### Malinke

Malinke and Mandinka are the most common terms for the regional and local dialects into which Mandekan has divided. Viewed on a map, they still recall the shape of the empire. They form a wide arc that begins from the Atlantic coast of the Gambia, southern Senegal (Casamance), Guiné and northern Guinea. It stretches from there eastwards across south-western Mali (the most populous part of the country) and northern Ivory Coast to north-western Burkina Faso, where the Black Volta roughly marks an eastern boundary.

Most speakers of these dialects are Muslims. Many are businessmen, and Muslim communities of 'Malinke' or 'Maninka' speakers have grown up in various West African cities such as Monrovia.

### Bambara

Bambara, a dialect of the older Mandekan, developed as a new standard language in the lands between the upper Niger and its tributary the Bani. Here, in the early 17th century, after the Empire of Mali had disappeared and its eastern successor, the SONGHAY Empire, had declined in turn, the non-Islamic Bamana-nke established an independent state. It developed into the two kingdoms of Bamana Segou (whose capital was at Segou) and Bamana Masa-si, 'the Bamana of the King's clan' (whose capital was at Kaarta). In the 19th century, not long before the French invasion, both kingdoms fell to the FULANI Empire.

With the increase of these kingdoms' power and trade, Bambara became an important second language over a wide region – and it still is, in spite of the political ups and downs of the last century. Some of the local languages are closely related Mandekan (Malinke) dialects; some are quite different, even entirely unrelated, such as Songhay.

Bambara is now an official language, almost the national language, of Mali. Under French rule Bambara speakers were widely employed in the civil service and the army: thus Bambara became an important language in the administration of Mauritania and Niger (as well as Mali, Senegal, Guinea, Ivory Coast and Upper Volta) and was the lingua franca of local troops.

### Dyula

Dyula, a second offshoot of the older Mandekan, spread in quite a different way. The Malinke dialects of western Burkina Faso and northern Ivory Coast – especially those of the region of Kong – have, over the centuries, gradually extended their influence in an entirely peaceful way. Alongside their influence, their language spread as well.

Relatively well educated, the Muslim speakers of the Dyula dialect gradually constituted a local elite. They were called on to arbitrate in disputes. Their names became fashionable. In its own region, Dyula turned into a prestige language, one that was used by chieftains and nobility, in lawsuits and other formal contexts. It is still a major local language in Burkina Faso and northern Ivory Coast.

The Dyula were, first and foremost, long distance traders between the upper Niger valley and the southern coast. Thus, far beyond their home region, and in an altered form that is often called Commercial Dyula, their dialect developed into a language of trade and the marketplace – the language of traders whose mother tongue is often one or another dialect of Mandekan. Linguists are not yet committing themselves as to how radically this kind of Dyula differs from the other kind.

This variety of Dyula is widely known as a second language throughout Ivory Coast and into western Ghana, especially in towns and cities and along main roads and railways. Further west, another and more creolised variety of Mandekan known as *Kangbe*, the 'clear language', is used as a lingua franca where Mandekan speakers regularly trade.

## Language and literature

Mandekan oral literature is noted especially for its epics. The cycle of Sunjata and the historical legends of the Bambara kingdom of Segou with its 17th-century ruler Da Monzon are retold by professional poets, *griots*, members of a hereditary and much-respected profession. Traditionally, each *griot* family was attached to a noble family and performed at its ceremonies, but *griots* also travelled and transmitted news: their repertoire was extensive and varied, though generally of epic form interwoven with songs. *Griots* are nowadays as likely to be heard on the radio as performing to a live audience. Hunters' songs are a recognised separate genre, including both historical and modern poetry performed by a different kind of specialist, the *dônso-jeli*, 'hunters' *griot*.

Several versions of the Mandekan epics have been collected, published and translated into English and French.

Mandekan has borrowed heavily from Arabic, language of the Islamic religion and culture that are shared by practically all speakers of Mandekan. Arabic *tabib* 'doctor' is the origin of Bambara *tubabu* 'European'. More recently loanwords of French origin have multiplied: *monperi* 'Catholic priest' from the French way of addressing priests, *mon père* 'my father'; *leglisi* 'church' from French *l'église* 'the church'.

The dialects of Mandekan differ in their tones and sound patterns, in their vocabulary – including the most basic items, such as articles and prepositions – but not in their grammatical structure.

A Mandekan dialect such as Bambara has

three tones, low, high and rising, but the tones are not marked in writing (*ê* and *ô* are open vowels), and they are not as crucial to the meaning of every word as they are in Chinese or in some of the languages of the West African coast. For all that, it is best to get the tones right: the words for 'yes' and 'no' are written identically – *ònhòn* – and differ only in that 'no' has a low tone on the second syllable.

### Mande Tan and Mande Fu

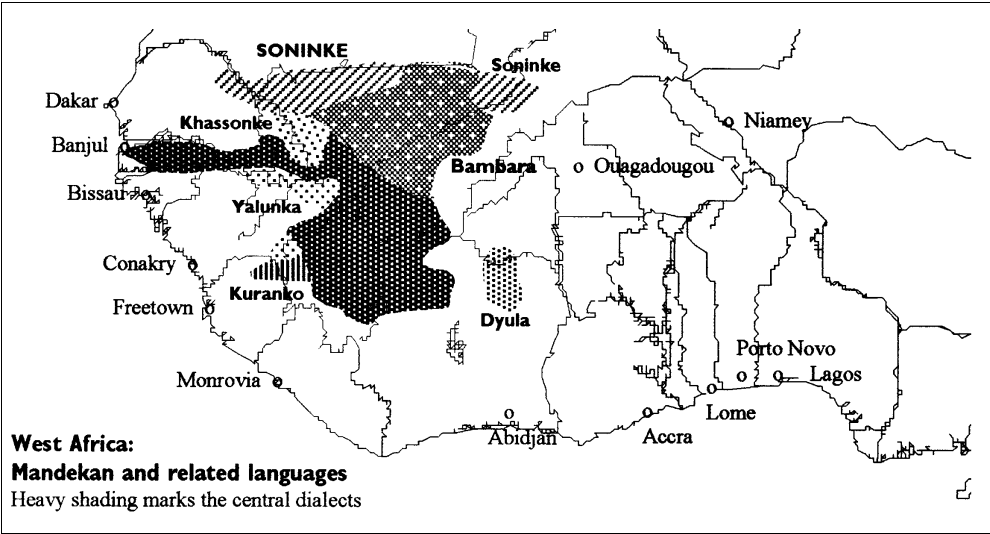
The first ten numerals in Bambara are: *kelen, fila, saba, naani, duuru, wòoro, wolonwula* or *folon-fila, segin, kònòntòn, tan*. This identifies Bambara (and Mandekan as a whole) as a member of the Mande *Tan* or northern Mande subgroup, distinguished from Mande *Fu* or southern Mande languages (for example *МЕНДЕ*) by the word used for 'ten'. The classification was introduced by the French linguist Maurice Delafosse in 1901.

## Mandekan speech and the Empire of Mali

*Mandekan dialects* can be classed into Western, Northern and Southern. The last includes two dialects which have become important international languages, Bambara and Dyula. *Bambara* centres in central Mali; it is also spoken in Ivory Coast, east central Senegal, Gambia, and Kénédougou province of Burkina Faso. *Dyula* centres in north-eastern Ivory Coast and western Burkina Faso.

Aberrant Mandekan dialects, which apparently separated from the main group relatively early, include *Khassonke* of far western Mali; *Kuranko* or Koranko in northern Sierra Leone and Guinea.

SONINKE is among the languages spoken by early opponents of the Empire of Mali. It also is a member of the Mande group, and still has about a million speakers on the borders of Mali, Mauritania and Senegal.



**The Mandekan epic**

The greatest achievement of Mandekan literature is certainly the epic cycle surrounding the hero Sunjata (or Sundiata or Son-Jara), a ruler of the 13th century: his miraculous birth and boyhood, his exile, his return and victory over the susu. In these societies epic poetry has been the responsibility of hereditary *jelilu*, ‘*griots*’ (this is the French term for West African bards, now used internationally).

Epics of Sunjata have been popular wherever Mandekan dialects are spoken, from Mauretania to Ivory Coast. The epic language differs from everyday speech, blending the features of several dialects. This simple war-song – not epic verse, but sometimes incorporated in epic – may be said to be Sunjata’s eulogy:

Minw bè sènè kè  
i ka sènè kè;  
minw bè jago kè  
i ka jago kè;  
minw bè kèlè kè  
i ka kèlè kè.  
Jata ye kèlè kè!

Those who want to go to the farm  
Can go to the farm;  
Those who want to go to trade  
Can go to trade;  
Those who want to go to war  
Can go to war.  
Jata went to war!

Maninka version by Ban Sumana Sisòkò, Mali’s foremost *griot*. From *The epic of Son-Jara: a West African tradition* ed. John William Johnson (Bloomington: Indiana University Press, 1986) p. 19

# MANIPURI

**1,100,000 SPEAKERS**

*India*

**M**anipuri is (with English) one of the two official languages of the state of Manipur, in the far north-east of India: it is the mother tongue of 60 per cent of the inhabitants of the state. It is closely related to the KUKI-CHIN AND NAGA LANGUAGES (see map there) and is thus one of the SINO-TIBETAN language family.

It is better known to linguists by the local name of the people and their language, *Meithei*. The middle consonant of this is a lateral click – the different-looking form in Assamese, *Mekle*, is another way of writing the same word.

Manipur is the only extensive, relatively level river plain in the area where Kuki-Naga languages are spoken. It is no coincidence that a single kingdom grew up here, and that a single language is spoken, without significant dialect divisions, by a relatively large population group. This is still in origin a strongly rural culture. The name of Imphāl, capital of Manipur, means ‘house collection’: it grew up less as a town than as a group of villages surrounding the royal

enclosure, and as the largest of the many markets of the state.

The Hindu religion of India was introduced to Manipur in the 18th century. A Burmese invasion in 1762 was followed by a long period of instability. British rule came in 1890; within independent India, Manipur gained the position of a state in 1972. Imphāl, once the royal capital, now the state capital, is best known for the British–Japanese battle there in 1944.

As Vaishnavites, Manipuri speakers do not hunt, eat meat or drink alcohol. Their staple is rice, grown in the irrigated fields of the valley.

With Vaishnavism came Indian, specifically Assamese, influence on the Manipuri language, and a continuing cultural interchange between Manipur and Assam. The classic Manipuri tale of Khamba and Thoibi became an Assamese classic as *Khamba Thoibir sadhukatha*, translated by Rajanikanta Bordoloi (1869–1939), author and anthropologist.

The first ten numerals in Manipuri are: *a-mā*, *a-ni*, *a-hūm*, *ma-ri*, *ma-ngā*, *ta-rūk*, *ta-rēt*, *ni-pān*, *mā-pan*, *ta-rā*.

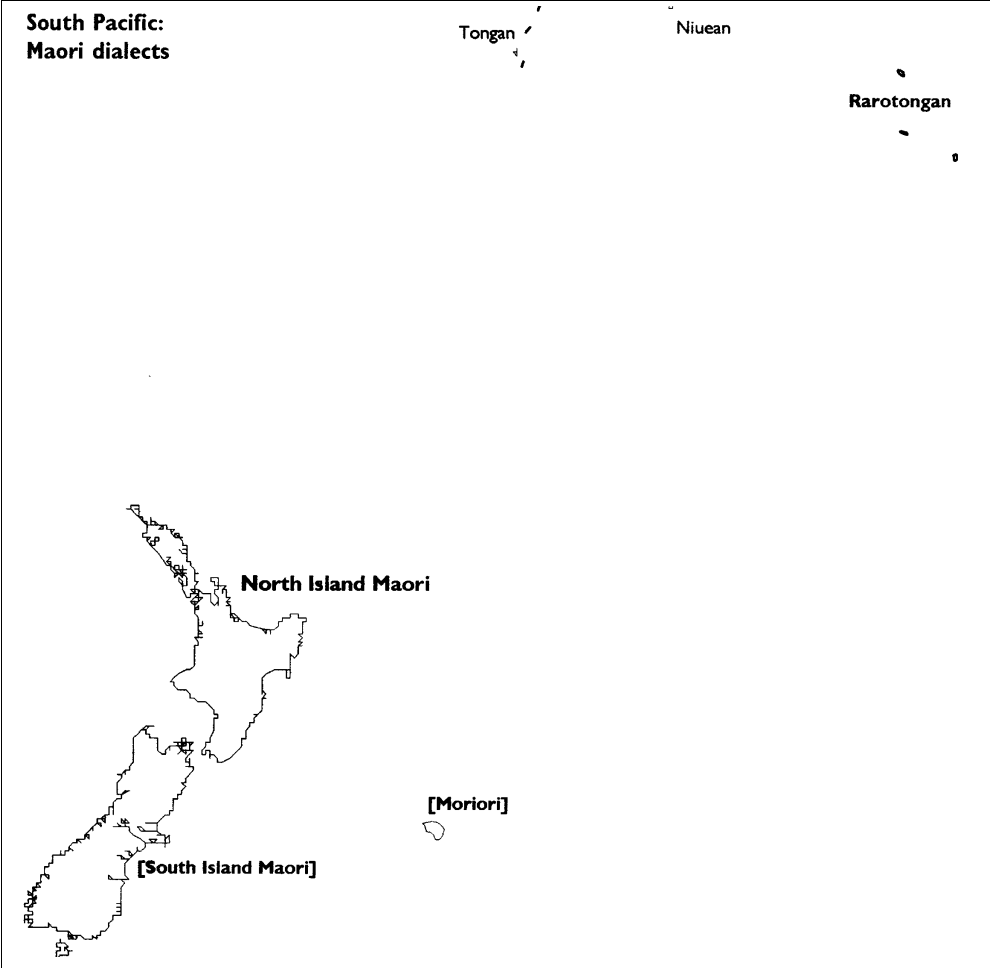
# MAORI

100,000 SPEAKERS

*New Zealand*

One of the AUSTRONESIAN LANGUAGES, Maori is spoken where it is as the result of one or more migrations up to 1,100 years ago, from eastern Polynesia to the previously uninhabited islands now called New Zealand.

When Europeans first came to New Zealand, Maori was spoken, in several dialects, in both North and South Islands. According to oral tradition there was continual travel and contact among speakers of the different dialects, extend-





ing to both islands, and this certainly helps to explain why dialect differences were rather minor considering the length of time over which they had been able to develop.

The first printed book in Maori was a reader, *E korao no New Zealand* 'Talk from New Zealand' by Thomas Kendall, published in 1815. More than a thousand publications had appeared by 1900, and there is a complete Bible translation in Maori. Much traditional oral literature has been collected and published. The New Zealand government, however, has (at least until recently) worked to suppress Maori in favour of English.

South Island Maori is now extinct; Maoris from North Island conquered South Island in the early 19th century, killing many speakers and absorbing the remainder into their own tribes. Remaining North Island Maori speakers number about 100,000 – all of them bilingual in

English – though three times as many people identify themselves as Maoris. Numerous New Zealand place names are of Maori origin.

Maori has ten consonants, *p, t, k, m, n, ng, wh, r, h, w* and five vowels, *a, e, i, o, u*: *wh* is variously pronounced *f, ph* or *wh*. The first ten numerals (similar to those of SAMOAN) are: *kotahi, rua, toru, whaa, rima, ono, whitu, waru, iwa, tekau*.

### Maori dialects

*North Island Maori* is divided into western and eastern dialects. *South Island Maori* is extinct, as is *Moriori* of the Chatham Islands, which appears to have been closest to the eastern North Island dialect of Maori. *Rarotongan* or *Cook Islands Maori* (40,000 speakers, more than half of them settled in New Zealand) is close enough to Maori for some mutual intelligibility, suggesting a historical link between the two populations.

# MARATHI

50,000,000 SPEAKERS

*India*

**M**arathi, one of the major INDO-ARYAN LANGUAGES in terms of the number of speakers, is one of those that are less known abroad, simply because of the smaller amount of emigration from the Indian state of Maharashtra.

The region has been called *Mahārāṣṭra* from ancient times. The adjective *Marāṭhī*, now used as the name for the language, is itself as old as the 2nd century AD (in the Sanskrit form *Mahārāṭhī*).

Words identifiable as Marathi occur in texts and inscriptions from this region from the 5th century onwards, more especially in those that are in the Prakrit dialect *Māhārāṣṭrī*, which is often considered the ancestor of Marathi – though some linguists doubt the direct link between them.

Inscriptions and texts entirely in Marathi begin in the 11th century but Marathi literature flowers in the 13th, with the remarkable Marathi verse exposition of the Sanskrit *Bhagavadgītā* by the young poet Jñāneśvarī (1271–96). The language as recorded to 1350 is counted as Old Marathi: Middle Marathi, in which there is a major corpus of historical prose, is dated from 1350 to 1800.

It may be that Marathi coexisted with KANNADA as a spoken language for an extended period during the first millennium AD, and that heavy influence from both Kannada and Telugu led to its rather rapid differentiation from Sanskrit, as compared with the more conservative Hindi.

Some linguists have gone as far as to argue that Marathi is historically a creole. Against this, it is itself conservative in some ways, retaining the

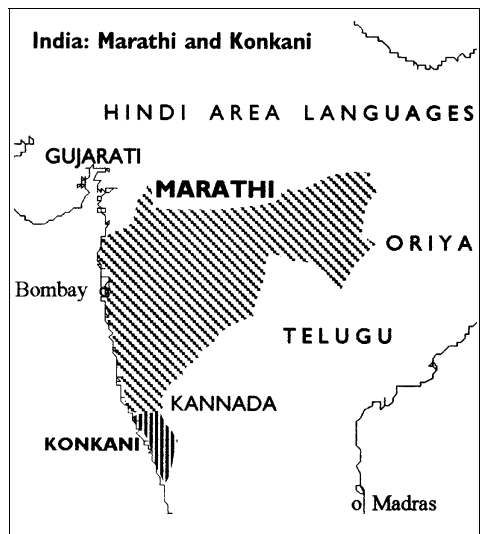
three genders of Sanskrit, masculine, feminine and neuter, and the stem alternations known as *guṇa* and *vr̥ddhi* (for example, *paḍ* ‘to fall’, *paḍ* ‘to make fall’).

From the 14th to the 17th centuries Persian was the language of government in Maharashtra and exerted influence in its turn. There are now many English loanwords.

## Marathi and Konkani

Standard *Marathi* is based on the *Deśī* or ‘country, hinterland’ dialect, specifically on that of the great city of Pune (older English form *Poona*). The *Koṅkaṇī* dialect is spoken along the western coastal strip, Konkani. The main eastern dialects are *Varhāḍī* and *Nāgpurī*.

*Halbi* is often considered a language distinct from Marathi, Oriya and Chhattisgarhi (which all



border on it). It has over half a million speakers in Bastar at the southern extremity of Madhya Pradesh, and some features link it to GONDİ, the Dravidian language of the jungle tribes of that area. The numerals '1' and '2' are *goṭok*, *duiṭhān*. The numerals '3' and above are like Marathi.

*Konkanī* – the language – is spoken from Ratnagiri southwards, notably at Goa, Mangalore and Cochin and in several other scattered coastal enclaves between Goa and Trivandrum.

### Marathi in writing

The cursive मोडी *Modi* script was invented for Marathi in the 17th century, but it is now less commonly found even in handwriting: since 1800 it has been officially replaced by a Devanagari alphabet, almost as used for Hindi but

with a few different character shapes. It is usually called बालबोध *Balbodh* 'that can be understood by a child'.

#### Numerals in Marathi

एक	१	ek
दोन	२	don
तीन	३	tīn
चार	४	cār
पांच	५	pāñc
सहा	६	sahā
सात	७	sāt
आठ	८	āṭh
नऊ	९	na
दहा	१०	dahā

#### The Devanagari alphabet for Marathi

अ आ इ ई उ ऊ ऋ ए ऐ ओ औ क ख ग घ ङ च छ ज झ ञ ट ठ ड ढ ण त थ द ध न प फ ब भ म य र ल व श ष स ह  
a ā ī ū ō ̐ ē ai o au k kh g gh ṅ c ch j jh ṇ ṭ ṭh ḍ ḍh ṇ t th d dh n p ph b bh m y r l v ś ṣ h

# MARI

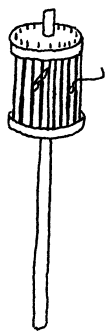
600,000 SPEAKERS

*Russia*

**M**ari is one of the URALIC LANGUAGES, spoken in the middle Volga valley in Russia, in the self-governing republic of Mari El. It is rather more distantly related to Finnish than neighbouring Mordvin is.

## To capture the queen

The Mari are renowned beekeepers. This sketch from the notebooks of the 19th-century Mari ethnographer Timofei Yevseyev shows the reed cage in which the queen bee was traditionally enclosed so that the swarm would settle around her. As she was trapped, the farmer would say this spell:



Müksh-awa, surtetəm saj onyzhə,  
yeshetəm pashash kəchkərə  
i pashalan tunəktə!

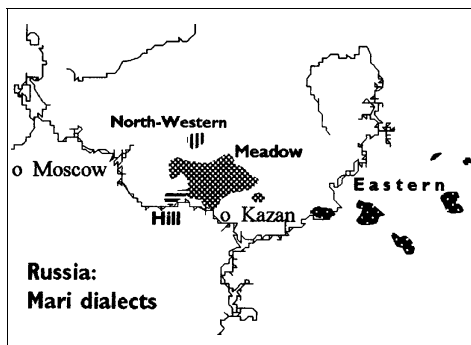
Queen bee, build your palace,  
put your family to work and teach them  
to make honey!

T. J. Jewsewjew, 'Bienenzucht bei  
den Tscheremissen' ed. J. Erdödi in  
*Journal de la Société Finno-Ougrienne*  
vol. 73 (1974) pp. 168–204

*Mari*, the people's own name for themselves, means 'human being' and seems to be a loan from proto-Indo-Iranian. The older term *Cheremis* (German *Tscheremissisch*) is still used by linguists: it was originally the Chuvash name for their Mari neighbours.

Present-day Mari country is the remnant of an area that once reached much further to the west. At some time, maybe three thousand years ago, early Mari, early Mordvin and proto-Finnic must have been spoken in adjacent regions, probably stretching all the way from the Volga to the Baltic Sea. The expansion of Russian broke up this hypothetical Finno-Ugric dialect continuum.

Since as long ago as the 8th century, Mari and CHUVASH speakers have lived side by side in their present territories, and there are many Chuvash loanwords in Mari. If the Mari had once been independent, they were so no longer after the Tatar conquest in the 13th century. Tatar domination was replaced by Russian rule in the 16th century.



### Hill Mari and Meadow Mari

About half the speakers of Mari live in the Mari El Republic. Hill Mari, the western dialect, and Meadow Mari of the central and eastern districts, have developed into separate literary

languages, both written in Cyrillic script. The Cheboksary reservoir covers a considerable area of former Hill Mari speech. There are also rather different Eastern and North-western dialects, spoken outside the borders of the republic.

# MARSHALLESE

30,000 SPEAKERS

*Marshall Islands*

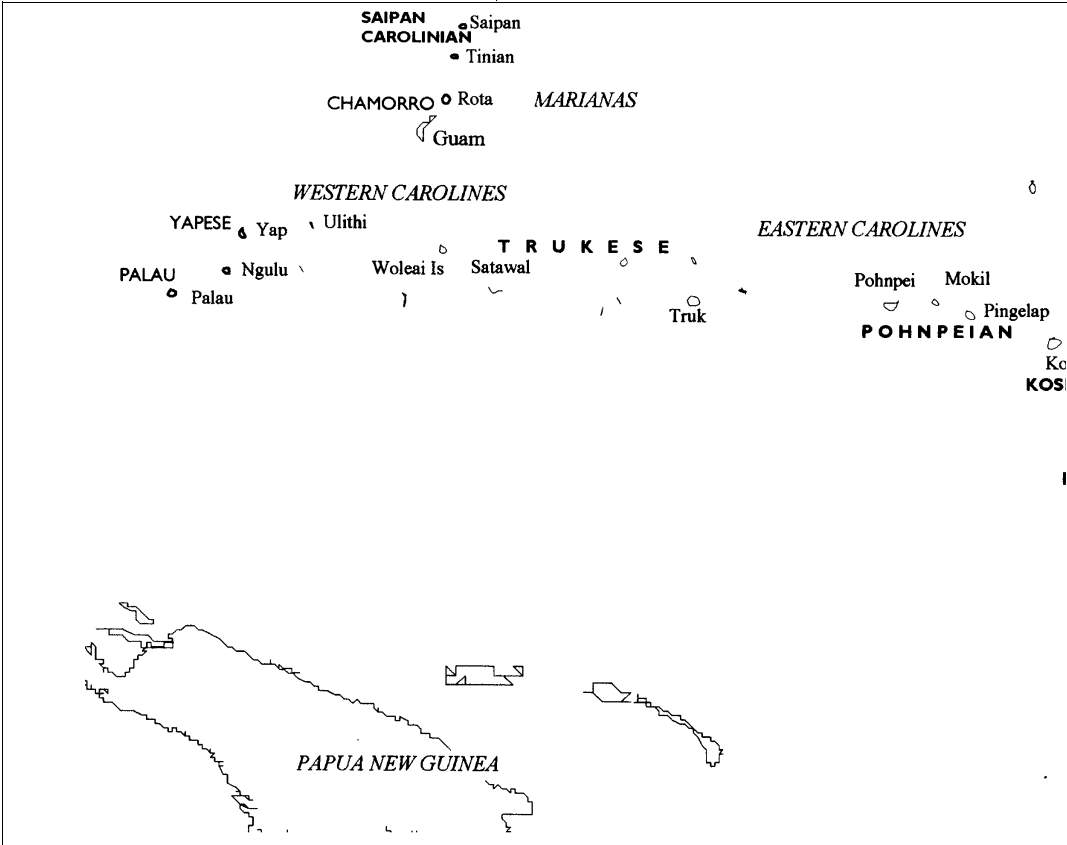
**M**arshallese, one of the Micronesian group of AUSTRONESIAN LANGUAGES, is spoken on the two island chains of the Republic of the Marshall Islands.

Already of interest to American missionaries, the Marshall Islands were briefly a German protectorate and the focus of German trade towards the end of the 19th century. Japanese administration followed after the First World

War, and the United States took over in 1945. The Marshall Islands became independent in 1986.

English and Japanese have had the strongest effect on the language.

Linguists have differed astonishingly over the vowels of Marshallese. Byron Bender, in a 1968 paper, argued that there were three vowel phonemes (*a e i*), other vowel sounds being condi-



tioned by surrounding consonants. The usual spelling recognises nine. Sheldon Harrison, in 1995, distinguished twelve, and added a distinctive feature of length, giving a total of twenty-four.

Marshallese has lost the complex series of numeral classifiers that are found in the other Micronesian languages (see box at KIRIBATI).

## Micronesian languages

KIRIBATI is the language of the Kiribati (Gilbert) island chain, of neighbouring Banaba (or Bwaanaba or Ocean Island), which has a slightly different dialect, and of Niu. It is also spoken in the Line Islands, and there are several migrant communities in the western Pacific. *Northern*, *Central*, *Southern* and *Bwaanaba* dialects are distinguished.

*Kosraean* or Kusaie is the language of 5,000

inhabitants of Kosrae State (Federated States of Micronesia); many inhabitants also speak Pona-pean.

*Nauruan*, with 4,000 speakers, is the traditional language of Nauru. This island, German at the end of the 19th century, was under Australian administration from 1920 until 1968. As a language of trade and later of government, English has been the strongest influence throughout, and there are signs that Nauruan is being abandoned in favour of English.

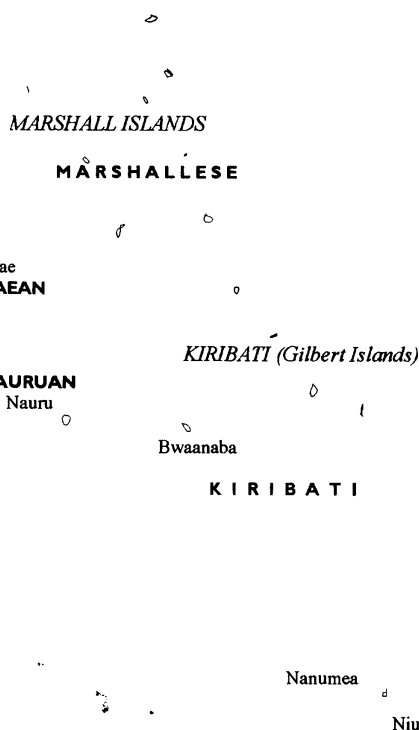
*Pohnpeian* or Ponapean has about 22,000 speakers. It is the major language of Pohnpei State (Federated States of Micronesia). *Ngatikese*, *Pingelapese* and *Mokilese* are counted as dialects of Pohnpeian or as closely related languages. Pohnpeian has a 'high language' with a partly separate vocabulary, used in speaking about people of high rank. Pohnpeian spelling uses *-h* to mark a long vowel, rather like German: *dohl* 'mountain'. German missionaries designed the orthography.

*Truk* or Ruk or Trukese (22,000 speakers) is the major member of a dialect chain with a total of 40,000 speakers. These languages are spoken in the Caroline Islands. Once nominally Spanish – but plied by American missionaries – they were sold to Germany in 1899, passed to Japan after the First World War and to the United States after the Second World War. They now form states of the Federated States of Micronesia. The languages have numerous English and Japanese loanwords. *Saipan Carolinian*, one of this group, is spoken in the Northern Marianas.

*Marshallese* has the largest number of speakers (30,000) of any Micronesian language. The western chain of the Marshall Islands is known locally as *Rālik*, 'western islands', the eastern chain as *Ratak*, 'eastern islands'. This is the major dialect division of Marshallese.

The dialects differ in vocabulary – and also in the vowels of words whose first two consonants are identical: Rl. *ipping*, Rt. *piping* 'good at jumping'; Rl. *ellor*, Rt. *lelor* 'shadow'; Rl. *ittil*, Rt. *tūtil* 'burn'; Rl. *ekkot*, Rt. *kōkot* 'strong'.

### Western Pacific: Micronesian languages



See KIRIBATI for a table of numerals in Pohnpeian, Truk and Marshallese.  
Examples from Sheldon P. Harrison in

*Comparative Austronesian dictionary* ed. Darrell P. Tryon (Berlin: Mouton De Gruyter, 1995- )  
pt 1 pp. 879–918

Numerals in Kiribati, Marshallese, Pohnpeian and Truk				
	Kiribati	Marshallese	Pohnpeian	Truk
1	teuana	cuon	ēu	eet
2	uoua	ɽuo	riau	érúúw
3	teniua	cilu	silū	éen
4	aua	emæn	pāieu	fáán
5	nimaua	ɭaləm	limau	niim
6	onoua	cilcino	weneu	woon
7	itua	cilcilimcuon	isu	fúús
8	waniua	ɽuwalitɔk	walū	waan
9	ruaiua	ɽuwatimcuon	tuwau	ttiw
10	tebwina	cɔngɔul	eisek	engoon

In Marshallese, ‘6’ originates as ‘3 + 3’ and ‘7’ as ‘3 + 3 + 1’.

A series of prefixes is added to the numerals in counting various classes of objects: *féfóch* ‘four long objects’, *faché* ‘four flat objects’, *wonossak* ‘six pieces of copra’, *wonowo* ‘six bunches of bananas’. These examples are from Truk. Marshallese, alone among the Micronesian languages, does not have the numeral classifiers.

Tryon

*Comparative Austronesian dictionary* ed. Darrell P. Tryon (Berlin: Mouton De Gruyter, 1995–); George L. Campbell, *Compendium of the world’s languages* (London: Routledge, 1991)



# MASAI

**750,000 SPEAKERS**

*Kenya and Tanzania*

**O**ne of the Eastern Nilotic group of NILO-SAHARAN LANGUAGES (see map at TURKANA), Masai is a major language of Kenya and Tanzania, spoken by a traditionally pastoral and war-like people who until recently ruled an extensive territory in the East African hinterland.

*Masai* or *Maasai* is the speakers' own name for themselves. They trace their origin to a migration from the north, led by *Maasinta*, who taught his people how to climb a mountain that lay in their way, and how to brand their cattle.

Masai-speaking country lies across many of the trade routes from the East African coast. The Swahili slave trade route from the Uganda region and Lake Victoria was one of these. The Masai did not take part in the trade but exacted tribute from trade caravans. Thus the language has interacted with Swahili. It has also been influenced by its neighbours: for example, the Kenya dialects of Masai have loanwords from KIKUYU. There has been a long tradition of regular trade in this region in a four-day market system. Women were the traders, and travelled between peoples even when they were (as often) at war.

Europeans began to visit Masai country in 1848. It was divided between British and German colonial governments at the end of the 19th century. German Tanganyika became a British protectorate after the First World War but Masai speakers remained divided between two administrations.

The first ten numerals in Masai are: *obo*, *aare*, *okuni*, *oonguan*, *imiet*, *ile*, *oopishana*, *isiet*, *ooudo*, *tomon*.

'The name of a dead child, woman, or warrior is not spoken again, and if the name is an ordinary word, that word is no longer used by the family. Thus, if a Masai man called *Ol-onana*, "the gentle one", died, on his death the word *nana* would be replaced in his family by some such word as *polpol* "smooth".'

G. W. B. Huntingford,  
*The Southern Nilo-Hamites*  
(London: International  
African Institute, 1953) p. 116

The *ol-* in this name is a masculine gender prefix, a form typical of Masai: *ol-arányàni* 'male singer', *enk-arányàni* 'female singer'; *ol-álém* 'sword', *enk-álém* 'knife'.

# MAYAN LANGUAGES

**T**he Mayan languages are a long-recognised linguistic family of south-eastern Mexico and Guatemala, one which some would now classify in the wider family of AMERIND LANGUAGES. There are about thirty surviving languages, with perhaps three million speakers in total. Spanish, as national and international language, is now progressively undermining the vitality of all the Mayan languages.

The relationship among the Mayan languages was noted by Lorenzo Hervás y Panduro in *Catalogo delle lingue conosciute e notizia della loro affinità e diversità*, published in Italy in 1784.

Historically these languages are of great importance, for they belonged to the Maya culture that flourished in the Yucatán peninsula in the 4th to 9th centuries, well before the Spanish conquest. Chontal, Chol, Tzeltal, Tzotzil and Yucatec were the main languages of classic Maya civilisation – and they all influenced one another, forming what is sometimes called the ‘Lowland Maya linguistic area’.

But the history of Mayan languages is much older than that. Linguists have made solid progress in the reconstruction of ‘proto-Maya’, ancestral to all the modern languages. Proto-Maya was, it is thought, spoken in the south-eastern, highland parts of modern Maya territory, at least four thousand years ago, and then began to split into dialects and to spread by migration and conquest. Archaeology suggests that Mayans (presumably speaking early forms of Chol, Chontal, Tzeltal and Tzotzil) were to be found in the lowland regions around Tabasco from, at the latest, 800 BC.

Some of the vocabulary of proto-Maya can be recovered: maize (Yucatec *kan*), with some of the technical terminology of maize-growing; marrow, sweet potato, chilli, avocado, maguey, cotton and cacao, among other crops; terms for cooking and weaving, and for religious rites.

The influence of the prehistoric Olmec culture, of central Mexico, was strong in Ma-

yan territory. Olmec itself is not thought to have been a Mayan language, but a member of the Mixe-Zoque group. The hieroglyphic writing system found on lowland Mayan sculptures, and now partly readable (see box at YUCATEC), is clearly a development of the still-undeciphered hieroglyphs used by the Olmecs.

## The major Mayan languages

*Cakchiquel* has 350,000 speakers in Guatemala, where it is used in primary education. Few modern printed texts exist, except school textbooks, yet the 16th-century *Memorial of Tecpán-Atitlán*, a mythological and historical compilation in Cakchiquel, is an important source for later pre-Columbian central America and for the Spanish conquest. The Cakchiquel capital, Iximché, became the first capital of Guatemala.

*Chol* has 90,000 speakers in Chiapas province, Mexico.

*Chontal*, with only 40,000 speakers today, was the language of Tabasco.

*Huastec* has perhaps 70,000 speakers in total. The two main dialects are those of *Veracruz* and *San Luís Potosí* in Mexico.



*Kekchí* or *Quecchí* has nearly 300,000 speakers in Guatemala and Belize.

*Mam* is a group of dialects of western Guatemala, with a total of at least 250,000 speakers.

QUICHÉ is the major Mayan language of modern Guatemala, with at least 750,000 speakers.

*Tzotzil* and *Tzeltal*, neighbouring languages in Chiapas province, Mexico, have 150,000 speakers in total.

YUCATEC or 'Maya', with 500,000 speakers or

more, is the modern Mayan language of Yucatán. Yucatec is one of a group of four closely related languages. The other three, *Mopan*, *Itzá* and *Lacandón*, are spoken by tiny population groups in central Guatemala and Belize. Yucatec itself clearly had a different fate: it moved north into the peninsula of Yucatán, perhaps two thousand years ago – then, as the vehicle of a great civilisation, it spread outwards again, almost swamping its three relatives.

### Numerals in Mayan languages

	<b>Cakchiquel</b>	<b>Quiché</b>	<b>Tzotzil</b>	<b>Yucatec</b>
1	jun	hun	jun	jun
2	ca'i'	cab	chib	ca
3	oxi'	oxib	'oxib	ox
4	caji'	cahib	cahib	can
5	vo'o'	oob	ho'ob	ho
6	vaki'	vacuc	vachib	vac
7	vuku'	vucub	hucub	vuc
8	vakxaki'	vahxac	vaxachib	vaxac
9	beleje	beleheb	valuneb	bolon
10	lajuj	lahuh	lajuneb	lajun

# MBUNDU

**4,800,000 SPEAKERS OF THE TWO LANGUAGES**

## *Angola*

**T**wo important BANTU LANGUAGES of Angola share this name. The first was the language of the *oviMbundu*, a trading people who in the 19th century were second only to the speakers of SWAHILI in the length of their journeys across southern Africa.

The name is in origin identical with *Bantu*, 'people'. To distinguish it from its neighbour this language has been called *Benguella Mbundu*, *South Mbundu*, and *Umbundu*. OviMbundu traders are called *Mambari* in older historical sources.

South Mbundu is the westernmost of the languages that are historically linked with the expansion of the LUNDA empire in the 16th century. The westbound conquerors were called *Jagas* by contemporary Portuguese sources. They ruled the Benguela highlands of west central Angola. For the next three centuries they dominated the trade that linked coastal Benguela with the upper valleys of the Kasai, Lualaba and Zambezi – the regions now known as Shaba and Zambia. OviMbundu kingdoms such as Bailundu had trading agreements, sometimes amounting to monopolies, with inland states such as the Lunda empire of Mwata Yamvo and the LOZI kingdom ruled by Lewanika. At first slaves were the most valuable commodity in which they dealt: by the end of the 19th century greater profits were coming from ivory, beeswax and rubber.

The South Mbundu language is historically the speech of the Benguela highlands. With trade and slavery it gradually spread along the trade routes of all southern Angola. It came to be particularly widely used in the coastal cities of the south, Lobito, Catambela, Benguela (older

Portuguese spelling *Benguella*) and Moçâmedes, where people from many parts of Angola, with different mother tongues, had come to live and work.

South Mbundu is close to Lunda, CHOKWE and LUVALE. However, it is clearly influenced by a different Bantu language which was presumably already spoken on the highlands at the time of the conquests – and this was related to the Bantu languages of sparsely populated Namibia, such as Herero (see map). This influence is seen in the sound pattern of Mbundu, notably its initial vowels. These turn many noun class markers into two-syllable prefixes: *imbo* 'village'; *ovaimbo* 'villages'.

Early in the twentieth century the Benguela Railway was built, eastwards from Lobito and Benguela to the mining lands of Shaba and Rhodesia (modern Zambia). Its route ran through Ovimbundu territory, and it supplanted the Mbundu trading caravans. Ovimbundu prosperity declined rapidly – yet, with rail transport, Mbundu has spread even further as a common language for inland peoples living near the railway.

## **Mbundu of the north**

*Loanda Mbundu* (or North Mbundu or Kimbundu or Quimbundu or Ndongo) is in origin the language of the Luanda district, and is related to KONGO. The Portuguese seized the harbour of Luanda (older spelling *Loanda*) from the declining Kingdom of Kongo, in the late 16th century. They made it a metropolis of the slave trade and the capital of their realm of Angola.

*Gentio de Angola*, ‘Gentile of Angola’, is one of the very earliest printed books in an African language. It is a series of lessons in Christianity, with Portuguese and Kimbundu texts on facing pages. There are some notes on the grammar and pronunciation of the language. This rare work was published in Lisbon in 1642 or 1643: in the surviving copy in the British Museum the last digit of the date cannot be read.

Luanda Mbundu is still a lingua franca of the trade routes in parts of northern Angola and south-western Congo (Kinshasa).

*K’ilu lyuti olongila vyakahandangala, p’osi esove lakamwe* – ‘There are many little birds at the top of the tree, but the ground beneath is clean’. The answer to this South Mbundu riddle? Stars.

José Francisco Valente, *Gramática Umbundu: a lingua do centro de Angola*. Lisboa: Junta de Investigações do Ultramar, 1964

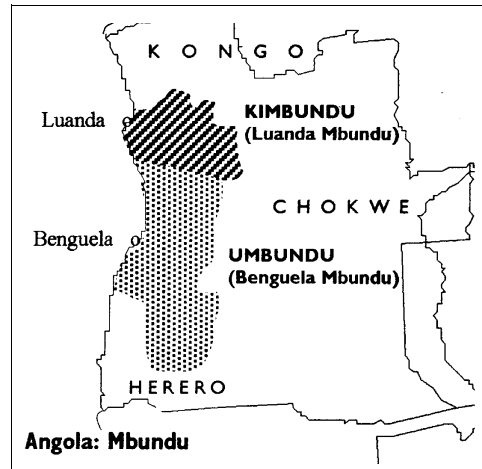
## Benguela Mbundu and Luanda Mbundu

South Mbundu or *Benguela Mbundu*, language of the oviMbundu, has about 3,000,000 speakers. In spite of its extensive territory, which was never politically united, it has only minor dialect divergences, an indication of the ubiquity of travel and trade.

South Mbundu resembles Lunda but shows substrate influence from a relative of the Bantu languages of Namibia. The best known of these are *Herero* (75,000 speakers), *Kwanyama* (150,000) and *Ndonga* (250,000).

North Mbundu or *Luanda Mbundu*, once the dialect of a southern region of the Kingdom of Kongo, has 1,800,000 speakers in Luanda Province, Angola. A close relative is Kongo.

Numerals in South Mbundu and North Mbundu		
South Mbundu		North Mbundu
mosi	1	moxi, úmue
vali	2	txiari
tatu	3	tatu
kwāla	4	kuāna
tālo	5	tānu
epandu	6	samanu
epandu vali	7	sambari
echelāla	8	nake
echea	9	txiela
ekwĩ	10	kuĩ



# MENDE

1,000,000 SPEAKERS

*Sierra Leone, Liberia*

**M**ende, one of the Mande group of NIGER-CONGO LANGUAGES, is spoken by a people of inland Sierra Leone. Although Mande languages separated from the rest of the Niger-Congo family many thousands of years ago, their history is obscure and it is quite unknown how long Mende has been spoken in its present territory. The Mende were in recent centuries a slave-trading and slave-owning people: this will have helped to spread their language (slavery was abolished in 1928).

In the 19th century Mende speakers for the first time penetrated as far south as the Atlantic coast, which was at that time largely the domain of Bullom (see map at TEMNE). The first Christian school that taught in Mende – a landmark among African languages – was established at Charlotte in 1831 by the Quaker missionary Hannah Kilham.

The name of *Mende* is native. It is the same in origin as that of several related languages and dialects – MANDEKAN, Mandingo, Malinke – and was the obvious choice by linguists, in the form *Mande*, to denote the whole group with Niger-Congo.

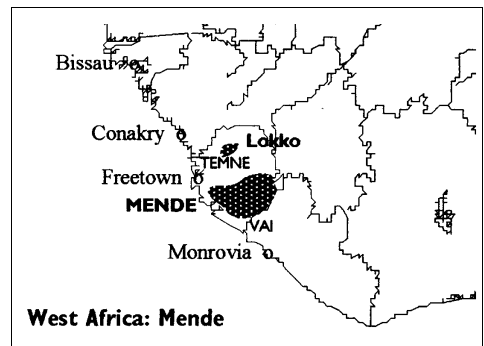
The southward expansion of the Mende came at the right time to open the hinterland of Sierra Leone to British trade and eventually to the British protectorate, proclaimed in 1896. Mende speakers, unaccustomed to any centralised government, were at first outraged by such innovations as taxation, but their rebellion of 1898 was suppressed and they began to work more peacefully alongside the colonial administration. Thus the Mende language spread further. It is now the

major language of southern Sierra Leone, used as a lingua franca by many whose first language is Bullom, Banta and Kissi, among others.

The first ten numerals in Mende are: *yera, fele, sawa, nani, lolu, woita, wofera, wayakpa, tau, pu*. Mende is thus one of the Mande Fu languages (see box at MANDEKAN). It is a tonal language, with four tones which are an essential part of the vocabulary and the grammar: *pú* 'put', *pù* 'England', *pû* 'cave', *pǔ* 'ten'. In Mende linguists have identified five separate noun 'plural' categories, each with distinct forms: indefinite plural, definite plural, plural of a person and companions, plural of masses, plural of agents.

## Mende and its neighbours

*Mende*, once called Kosso or Mande, the language of inland southern Sierra Leone, has a northern dialect in an enclave which became separated perhaps a century ago, *Lokko* or Landogo. Port Lokko, in Temne-speaking central Sierra Leone, takes its name from the Lokko slaves who were formerly exported from here.



# MERU

**1,250,000 SPEAKERS**

*Kenya*

**M**eru is one of the BANTU LANGUAGES: its territory lies in central Kenya, to the north-east of its close relative KIKUYU (see map there), but almost completely surrounded by Oromo-speaking territory. Meru society, like that of the Kikuyu, was traditionally warlike. Among more peaceable occupations, speakers of Meru are well known as beekeepers.

*Thagicu* is the name of the region from which, long ago, speakers of Kikuyu and related languages may have dispersed, and Thagicu is now the usual name for the language group to which both belong. Local legend tells of the migration

of Meru speakers from the coast around 1700; of their old name, *Ngaa*; and their new name, *Meru*, said to mean 'peaceful place'.

Meru serves as a second language for some speakers of smaller neighbouring tongues, Chuka and Tharaka.

Another, quite different, community is also called Meru by outsiders. The Meru of Mount Meru in Tanzania call themselves *Varwa* and speak a Chagga dialect, *Kirwa* or *kiMeru*. Their recurrent conflicts with their German and British rulers helped to discredit the Tanganyikan 'protectorate' and bring about independence in 1961.

# MIAO

**PERHAPS 5,500,000 SPEAKERS**

*China and south-east Asian countries*

**M**iao is a group of closely related dialects which, together with YAO (see table of numerals there), make up the Miao-Yao language family. This was once thought to be distantly related to the Sino-Tibetan languages (including Chinese) but the evidence was weak. It has now been argued by Paul Benedict that Miao-Yao is a member of the wider family of AUSTRO-TAI LANGUAGES.

*Miao* is the official Chinese term for this large linguistic minority; in Vietnamese it is *Meo*, or in full *Man Meo* – *Man* being the Vietnamese and Laotian term for Yao speakers, who were the first of this linguistic family to become known in Vietnam. Miao speakers themselves do not like the term, calling themselves *Hmong*, a name that many linguists also use.

Speakers of Miao were probably known to the Chinese at least two thousand years ago (though Chinese records often include other peoples under the name *Miao*). Their centre, as far back as records go, has been the middle Yangtze valley and what is now the province of Guizhou, where half the five million Miao speakers of China are to be found now. Miao legends tell of an ancient migration from the cold north: historians generally do not believe it.

The history of the language for the last few centuries, so far as it can be reconstructed, has been of retreat under the numerical – and political – pressure of Chinese. Miao is essentially spoken by hill peoples with a record of resistance and rebellion – serious uprisings took place in 1698, 1735, 1795 and 1854. Resulting Chinese campaigns to quell revolt may well have cata-

lysed the gradual Miao migration southwards of recent centuries.

By about 1870 all remaining independent Miao political units had disappeared. But even recent reports tell of traditional epic poetry, recited by skilled performers, narrating Miao heroic deeds in rebellions against the Chinese. As late as 1941–3 the Chinese government was forbidding the use of the language and the wearing of traditional Miao costume.

Once perhaps spoken over a fairly unified territory, Miao now makes up a complex series of speech enclaves – in fact none of the minority languages of China is more widely scattered. Everywhere it is surrounded by Chinese and Tai languages, spoken by wet rice growers in the plains and valleys, while Miao speech remains typical of the hills. Miao speakers, individual families and whole villages, are highly mobile. A migration south-westwards into Vietnam, Laos and Thailand was certainly under way in the 19th century – it may be much older. Miao people in these countries number well over 300,000; in Laos, the Miao of the Xieng Khouang area achieved some political recognition in the 1950s. From Vietnam a sizeable community of Miao speakers has now settled in the United States.

In spite of geographical fragmentation, Miao dialects have not diverged to the point of mutual unintelligibility. It is a predominantly monosyllabic language with around fifty initial consonants, six vowels and seven or eight tones. Miao speakers are frequently bilingual or multilingual; Lao serves as a lingua franca for many southern Miao, while Yunnanese and other dialects of Chinese serve a similar function further north.



Miao has borrowed heavily from Chinese, particularly terms needed in modern life – soap, matches, school, club – and political innovations such as cooperative, industrialisation, commune. Such words are often readily distinguishable: many diphthongs do not occur in native Miao words and identify Chinese loanwords immediately.

### Miao subdivisions

‘These communities are so scattered and so broken up into groups and subgroups, there are literally dozens of local names for them, most of which refer to something the Chinese found noteworthy about them, their appearance or location, say, or their dress or their hairstyle. Many of the names are indeed memorable. Here are a few examples: the Shrimp Miao, the Short-skirt Miao, the Long-skirt Miao, the Magpie Miao, the Pointed Miao, the Up-side-down Miao, the Steep-slope Miao, the Striped Miao, the Big-board Miao, the Cowrie-shell Miao. Only five such names are widely known, however: the Black Miao, the Red Miao, the White Miao, the Blue Miao (*Qing Miáo*), and the Flowery Miao (*Huá Miáo*).

‘The authorities of the People’s Republic do not . . . refer to the Miao by those colorful epithets of the past, holding such names to be degrading. Instead . . . they classify the people by the dialect that they speak.’

S. Robert Ramsey, *The languages of China* (Princeton: Princeton University Press, 1987)

p. 281

Miao was not a written language until Christian missionaries devised a Latin orthography at the beginning of the 20th century. After the Revolution a new romanisation was introduced, closer in style to the pinyin orthography for Chinese.

### The Miao-Yao languages

In China, Eastern, Northern and Western Miao dialects are recognised. In Vietnam, Laos and Thailand, dialect distinctions are still traditionally made with the help of dress and customs: here *Hmong Neua* (also spelt *Mong Njua*, Blue or Green Meo) and *Hmong Daw* (White Meo) are the most important groups.

YAO or *Mien* really consists of two languages, *Yao* and *Nu*, each of which has dialect subdivisions. Together they have about 700,000 speakers in China and well over 100,000 in south-east Asian countries.



### The rice harvest

Lù giể, mblễ txôu pàng;  
 Xì giể, mblễ troa lô;  
 Poa giể, mblễ giòa sá;  
 Chơu giể, hlái mblễ, nđầu mblễ.

In the sixth month, the rice flowers;  
 in the seventh month, the grains form;  
 in the eighth month, the rice ripens;  
 in the ninth month, we cut the rice, we thresh the rice.

from a Miao agricultural calendar collected by F. M. Savina

# MINANGKABAU

**6,500,000 SPEAKERS**

*Indonesia*

**O**ne of the AUSTRONESIAN LANGUAGES, and one of the regional languages of Indonesia that are closely related to MALAY (see map and table of numerals there), Minangkabau is spoken in the central part of the western coastal strip of Sumatra: it is the language of the province of Sumatera Barat, 'West Sumatra'. Carried by a busy sea-borne trade, it is a well-known lingua franca along the whole of this coast and its offshore islands.

Minangkabau and Kerinci form a single dialect continuum with Malay itself on its native

ground. However, the Minangkabau dialects have undergone big phonetic changes, so that they no longer look or sound much like Malay.

Moreover, Minangkabau, as a language of trading communities, has itself spread well beyond its original region. There are colonies of speakers in north-western and south-eastern Sumatra, in the Malay state of Negeri Sembilan, in Bandung and Surabaya on Java, and elsewhere in the archipelago. There are said to be half a million speakers in the Indonesian capital Jakarta.

## **The story of Anggun Nan Tungga**

A traditional cycle of stories is told by oral epic poets, *tukang sijobang*, in western Sumatra. While puppet plays, *randai*, are performed in standard Minangkabau, the epics, *sijobang*, are typically in local dialects. One performer of *sijobang* began his story with this preface:

Lusueh kulindan suto kujuk,  
sodang tajélo atéh karok,  
lusueh di pétak rang Malako;  
sungguéh kok bolun tolan tuntuk,  
umpamo kojo tukang kakok,  
niat sangajo iko juo aiii . . .

Worn is the thread, the silk is tangled,  
trailed over the heddles,  
worn in the room of the Malaccan;  
although, my friends, you have not asked me to,  
like a workman performing his task,  
this is just what I want to do . . .

Nigel Phillips, *Sijobang* (Cambridge: Cambridge University Press, 1981) pp. 42–3

# MINGRELIAN AND LAZ

**POSSIBLY 600,000 SPEAKERS**

*Georgia, Turkey*

**A** lone in the Kartvelian or South Caucasian family, GEORGIAN (see map there) is an official and literary language with a long written history. In western Georgia – the region closest to the Black Sea coast – the everyday languages are quite distinct from Georgian, though with a family resemblance. In ancient times this was Colchis, from which the Argonauts of Greek legend retrieved the Golden Fleece: later it was Lazica, the western neighbour of early Georgia.

In a compact area of north-western Georgia Mingrelian is spoken. To many it is a mother tongue, though Georgian is the language they learn at school, hear in the media and read in the press. To many speakers of ABKHAZ, in the far north-western extremity of Georgia (and locked in hostility to Georgian rule) Mingrelian functions as an everyday second language, quite unrelated to their own North-west Caucasian speech.

Mingrelian was never listed as a separate ethnic group or language in Soviet statistics: speakers were counted simply as Georgian, and they count themselves as Georgian. Independent Georgia has been too unsettled and too riven with nationalism to allow statistical research on minorities. Some estimates of Mingrelian speakers put them at about half a million. 'Estimates of numbers of Laz speakers in Turkey range from 46,987 (from a 1945 Turkish

census) to ten times that number' (Dee Ann Holisky in *The indigenous languages of the Caucasus* vol. 1 ed. Alice C. Harris [Delmar, NY: Caravan Books, 1991] p. 397). Mingrelian and Laz are sometimes grouped together as *Zan* or *Tsan*.

To the south, Georgian dialect areas now reach the sea in the Batumi region; once, it is believed, Mingrelian was spoken here too. Mingrelian dialects would then have shaded gradually into Laz, which is linguistically close to Mingrelian but is generally regarded as sufficiently distinct to count as a third member of the Kartvelian family. Laz is spoken along a strip of the Black Sea shore from Sarpi in Georgia to Pazar in Turkey. Most speakers are bilingual in Turkish. Laz, like Mingrelian, is unwritten and in general ignored by both Turkish and Georgian governments.

Both languages have five vowels and thirty consonants. As in other Caucasian languages, consonant clusters abound. The plural in nouns is marked by a suffix, *-ep* in Mingrelian, *-pe* or *-epe* in Laz.

*Denizi görmemiş, Laz'a Kürt denir*, 'A Laz who has not seen the sea is called a Kurd' – the Turkish proverb places two Turkish minorities geographically, blurring the linguistic and cultural divide between them.

Numerals in Mingrelian, Laz, Georgian and Svan				
	Mingrelian	Laz	Georgian	Svan
1	art'i	art'i	ert'i	eshkhvi
2	zhiri	zhur	ori	ervi
3	sumi	shum	sami	semi
4	'ot'khi	ot'khi	ot'khi	voshdkhv
5	khut'i	khu	khut'i	vokhvishd
6	amshvi	ashi	ek'vsi	mevsgve
7	shk'vit'i	shk'wit'	shvidi	ishgvid
8	ruo	ovro	rva	ara
9	chkhoro	chkhoro	tskhra	chkhara
10	vit'i	vit'	at'i	ieshd
Based on Adolf Dirr, <i>Einführung in das Studium der kaukasischen Sprachen</i> (Leipzig: Asia Major, 1928) p. 358				

# MIXTEC

**PERHAPS 250,000 SPEAKERS**

*Mexico*

**M**ixtec belongs to the Otomanguean family of AMERIND LANGUAGES. It is a language of Mexico, and one of those with a long written tradition, for Mixtec manuscripts are among the major sources on pre-Columbian culture and history.

The name of the speakers and their language identifies them as 'cloud people'. Their historic homeland is Mixteca Alta, Achiutla being the legendary place of origin of the Mixtec people.

Tilantongo is the site of extensive Mixtec ruins and was once their capital. The city is sometimes said to have been founded by non-Mixtec invaders who became assimilated to Mixtec culture. At any rate, at the Spanish conquest, Mixtec commoners would still refer to their rulers in terminology that was not Mixtec but apparently drawn from another Otomanguean language, Cuicatec of Puebla. There was, however, cultural mixing, in later pre-Spanish times, among Mixtec and Puebla peoples.

From Tilantongo Mixtec domination spread eastwards about 900 to Oaxaca, where they displaced their linguistic relatives, the Zapotecs, from Monte Albán. Later they expanded westwards into the Valley of Mexico. Southwards their coastal kingdom centred on Tututepec in the Sierra Madre del Sur.

Some Mixtec manuscripts ('codices') are religious: these have been interpreted by Edward Seler and others. Others are historical, or rather genealogical. They give royal dates of birth and death, ceremonial names of monarchs, lists of their relatives and wives; they record state

events, migrations, colonisations, conquests and sacrifices. Some deal with a single place: some, like the Bodley Codex now in Oxford, cover a wider political area.

For a long time it was unclear whether the non-ritual codices were truly historical (as United States researchers thought) or were linked to mythological tales (as German scholars believed). It was not even certain that Mixtec was the language hidden in the pictographs. Alfonso Caso made the breakthrough while studying a map of Teozacoalco made for the Spaniards around 1580. He found that the map included names, dates and genealogies of Mixtec kings – and these matched dates and pictorial versions of names in the codices. From this starting-point, by comparing the codices among themselves, Caso was able to build up a history of Mixtec kings from the 7th to the 17th century, and to show that the sequences of pictographs made sense only when read in Mixtec.

As the Zapotecs were driven eastwards and the Mixtecs took the valley sites, Mixtec borrowed from Zapotec the word for 'turkey' among other luxury items previously unfamiliar to Mixtec speakers. Mixtec also borrowed from Mixe-Zoque such words as proto-Mixe-Zoque *\*toto* 'paper', *\*kuku* 'turtle-dove'. The Mixtec medical loanwords in neighbouring Trique suggest some cultural dominance – that Mixtec shamans were called in to cure Trique diseases.

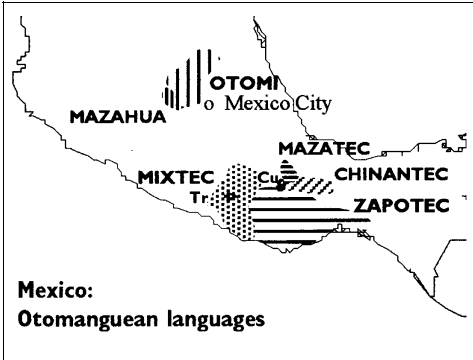
The first ten numerals in Mixtec are: *ĩ'ĩ'*, *uù*, *unì*, *kũũ'*, *ũ'ũ'*, *ĩ'nyũ*, *úshia*, *unà*, *ĩ'ĩ'*, *ushì*.

**Mixtec, Zapotec and some relatives**

The *Mixtec* dialects are spoken in Oaxaca state. Close relatives, with only a few thousand speakers, are *Cuicatec* and *Trique*.

ZAPOTEC is also a language of Oaxaca state. It has about 450,000 speakers.

Other Otomanguan languages, all spoken in Mexico, include *Otomí* (200,000), *Mazahua* (350,000), *Mazatec* (100,000) and *Chinantec* (60,000). There are still others with smaller numbers of speakers.



# MON

**PERHAPS 200,000 SPEAKERS**

*Burma, Thailand*

**O**ne of the Mon-Khmer group of AUSTRASIATIC LANGUAGES, Mon was once the ruling language of a powerful kingdom. Through Mon, Buddhist culture was transmitted to the early speakers of Burmese. By way of Mon script, Burmese and Shan first became written languages. Mon is now a minority language of southern Burma and central Thailand (see map at KHMER), and most of its speakers are bilingual in Burmese or Thai.

*Mon* is the speakers' own name for themselves and their language. In Burmese they are sometimes *Mun* but traditionally *Talaing*, and that term was used in older English sources, along with *Peguan*, which is now completely obsolete.

There are written sources for Mon history, in the form of stone inscriptions in Mon, from the 7th century onwards. The oldest are from central Thailand, once the Mon kingdom of Dvaravati. The Mon kingdom in Burma, with its old capital at Thaton, is known from the 10th century. It was a centre of Buddhist learning, with close literary and cultural links with Sri Lanka. The Old Mon language of this period had already many loanwords from Pali, and was written in a script of Indic type.

The Burmese monarch Anawrata, in 1057,

attacked Thaton and returned to Pagan with manuscripts of the Buddhist scriptures – they would have been in Pali, in Mon script, and perhaps with Mon commentaries.

A Mon kingdom in southern Burma, usually called 'Pegu' in European sources after its chief city, survived until its final conquest by the Burmese in 1757. From that date until 1852 the use of Mon was strongly discouraged: it was forbidden in religious education. The Mon-speaking country then became part of British Burma. In the last two centuries there appears to have been little desire among Mon speakers to emphasise their distinctness as against Burmese.

Although Mon is not an official language in Burma, it is now used again in traditional monastic education, which most boys experience. This has so far ensured the survival of a standard literary form of Mon, which is closest to the dialect spoken in the neighbourhood of Moulmein. Some publications continue to appear in the language.

Mon once influenced Burmese heavily. Recently it has been influenced in its turn – by Burmese and by English. It is not a tonal language but, like Khmer, has a 'clear' and a 'breathy' register, a feature which now forms part of the Burmese sound pattern too.

# MONGOLIAN LANGUAGES

**PERHAPS 5,000,000 SPEAKERS**

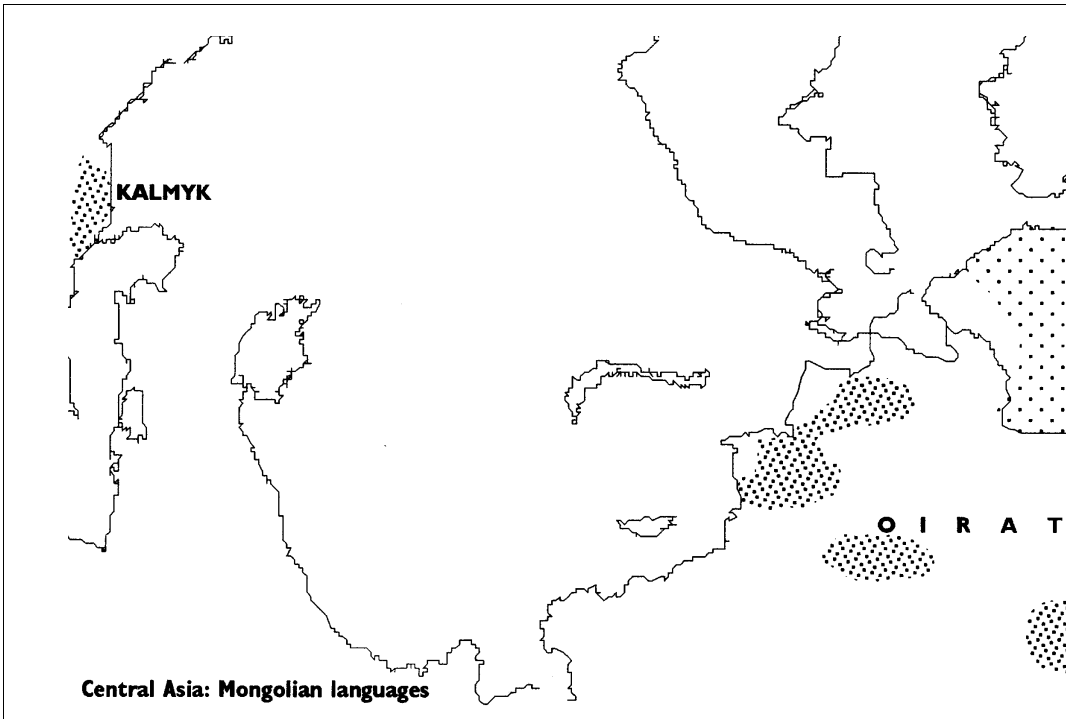
*China, Mongolia, Russia*

**A** group of closely related languages usually believed to be one of the constituents of the ALTAIC family.

In the 13th century horsemen from Inner Asia, ruled by the Mongol leader Genghis Khan and his son Ogedei, conquered much of the known world, from central Europe to the Pacific coast of Asia. The warriors were, some of them, Mongol speakers; but it is clear that the majority spoke Turkic languages, for it was Turkic languages

that spread widely as a result. In due course even the 'Mongol' hordes turned from Mongolian to Turkic as their language of government. Only in China did Mongolian remain for some centuries a court language – to be eventually supplanted by another language of northern invaders, Manchu (see TUNGUSIC LANGUAGES).

However, the great conquests left Mongolian dialects more widespread than they had been before. Their homeland was, and still is, the





country we now call Mongolia, with a region of Siberia just to the north and some parts of China to the south and east. Beside this, they can still be heard in Afghanistan, and even, far away to the west, in southern Russia on the conventional borderline between Europe and Asia.

Mongolia itself, once known by reference to China as 'Outer Mongolia', became an independent state – the first true 'Soviet satellite' – in 1924. 'Inner Mongolia' remains an autonomous region of China; about half the speakers of Mongolian live here. Most speakers in Siberia (about 300,000) are in the Russian republic of Buryatia, self-governing since 1992. The Kalmyks occupy the Russian republic of Khalkh Tangch. Its origins go back to the 16th century, when Kalmyk Khan signed a treaty, after the collapse of the Golden Horde, making his people a part of the Russian Empire. The First Kalmyk Regiment, mounted on Bactrian camels, made a great impression on western Europe when they entered Paris in 1814, inspiring Balzac's sketch 'Kalmyks in Paris'.

Mongolia itself, once a theocratic state, re-

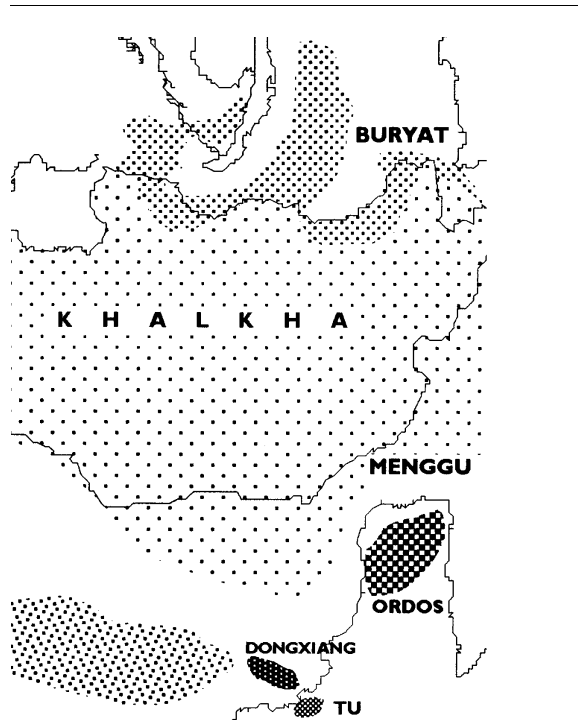
mains a stronghold of Mahayana Buddhism, a fact with which its Communist governments have had to come to terms. Buddhism came to the Mongols from Tibet, and Tibetan remains a sacred language taught in monasteries, alongside classical Mongolian. Mongolian Buddhists acknowledge the spiritual authority of the Dalai Lama. In Buryatia Buddhist monasticism was almost destroyed in Soviet times, though some young men were able to study in Mongolia. There is now a Buddhist revival. The Kalmyks, too, are Buddhist and traditionally nomadic. Numbering nearly 100,000, the Kalmyks were among Stalin's 'punished peoples', deported en masse to Central Asia in 1944; from 1957 onwards their territory was gradually restored, a process completed only in 1992. There are now 150,000 of them.

The first written records of Mongolian date from the 13th century, under Genghis Khan, when the script was devised (traditionally in 1204). But the first work of Mongolian literature, and one of the greatest, the *Secret History of the Mongols*, was written down in Chinese characters.

Classical Mongolian is the language of the Buddhist scriptures as translated from Tibetan in 1604–34 and printed from wood blocks in Beijing. It is still the literary language for all educated Mongolian speakers. The language displays its cultural history in its numerous loanwords from Sanskrit and Tibetan.

### Mongolian and its relatives

The most important language in the group is Mongolian itself, a literary language and group of spoken dialects known under six names in progressively differing varieties. First is *Khalkha*, the national language of modern Mongolia, which is very close to *Menggu*, the main language of 'Inner Mongolia' in China, and also close to *Buryat*, of the Buryat ASSR to the north. Far to their west is the variety called *Kalmyk* in Russia, where it is now spoken to the north of the Caspian – and called *Oirat* in the Xinjiang region of China. The last member is *Moghol* or *Mogul*, once spoken in Herat, Maimana and Badakshan provinces of



Afghanistan. Having separated from the rest some time ago, Moghol shows archaic features which it shares with old-fashioned literary Mongolian rather than with the other modern forms of Mongolian. Moghol is now almost extinct, but has left traces of its former importance in the form of Moghol loanwords in the Hazaragi dialect of PERSIAN.

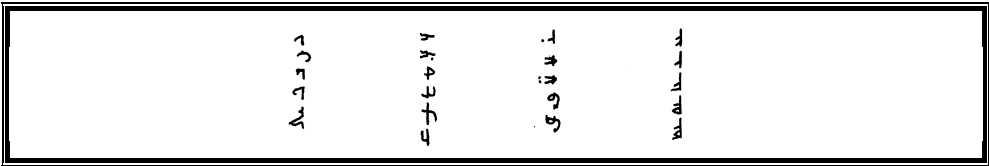
There are several other languages in the group, distinct from these partly because they do not share the common inheritance of Mongolian literary culture. Most notable are three local languages of China: *Tu* or Monguor, with about 100,000 speakers in Qinghai, *Dongxiang* or Santa, with 280,000 in Gansu, and *Dagur* or Daur, with 60,000 in Inner Mongolia and Xinjiang.

Numerals in the Mongolian languages				
Classical Mongolian		Khalkha	Buryat	Kalmyk
nigen	1	negen	negen	negn
qoyar	2	khoyor	khoyor	khoyr
ghurban	3	gurav	gurban	hurvn
dürben	4	dürüv	dürben	dörvn
tabun	5	tavan	taban	tavn
jirghughan	6	zurgaан	zurgaan	zurhan
dologhan	7	doloon	doloon	dolan
naiman	8	naiman	naiman	nəəmn
yisün	9	yesön	yühen	yisn
arban	10	arav	arban	arvn
From George L. Campbell, <i>Compendium of the world's languages</i> (London: Routledge, 1991) and other sources				

Mongolian and Oirat scripts

Mongolian is traditionally written in its own script, derived from the Uighur script that had been borrowed from Sogdian and came ultimately from ARAMAIC. This traditional script is still used in Inner Mongolia, while Oirat in China

is written with the related Oirat alphabet devised in 1648. In independent Mongolia the Cyrillic alphabet was introduced for Khalkha and Oirat in 1944, though the traditional script is still used in religious texts. In the Soviet Union a Latin alphabet was assigned to Buryat and Kalmyk in 1931 to be replaced by Cyrillic in 1938.



# MŌŌRE

3,000,000 SPEAKERS

*Burkina Faso*

**O**ne of the NIGER-CONGO LANGUAGES, Mōōre was the ruling language of the kingdoms of Ouagadougou and Yatenga, founded by conquering horsemen from the Lake Chad area in the 10th century. These two states in the Niger bend survived until the French occupation in the late 1800s, and their influence stretched over most of modern Burkina Faso and northern Ghana.

*Mossi* is the name of the people (singular *Moagha*) and is sometimes used for the language. The speakers themselves, however, call their language *Mōōre*, 'language of the *Moogho* country'. The usual French spelling is *Moré*.

The populations of these states were of mixed origin and language, including speakers of FULANI and MANDEKAN. Mōōre became the lingua franca, and for many the mother tongue, throughout much of the territory that was to become Haut-Volta under French rule, now the independent state of Burkina Faso. Probably as many as a million people speak it as a second or third language, in addition to the three million for whom it is the mother tongue. Its speakers often migrate to find seasonal and short-term work: thus there are large communities of Mōōre speakers in the southern cities of Ivory Coast, Ghana, Togo and Benin. There are also widely scattered Mōōre settlements in northern Ghana.

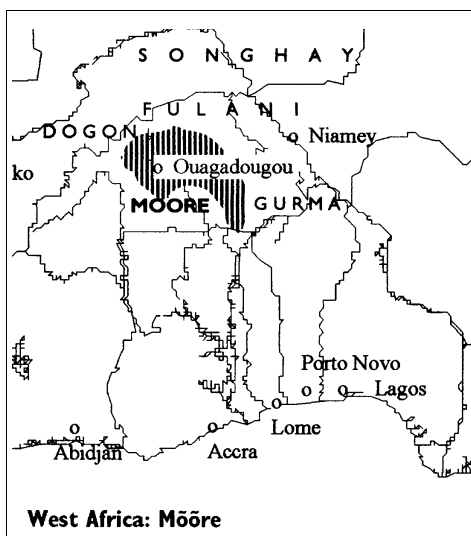
Mōōre is a tonal language, with three tones, high, low and falling-rising. A system of long distance drum messages, centred on the royal

palace at Ouagadougou, used three-tone drums to match the patterning of normal speech.

There are six noun classes in Mōōre, marked by terminations that are widely different in singular and plural. 'Horse' is *wèefoo*, 'horses' *widi*. There are numerous French loanwords – *livrè* 'book', *àlímètà* 'matches' (French *allumettes*) – and some Songhay loanwords, Songhay being the trading language of this part of the Niger valley. The first ten numerals are: *ié, yí, tá, náasee, nú, íoabee, íópóe, nú, ué, pūgá*.

## A Mōōre folk tale begins . . .

Em bā sō:āmbā nē ěm bā bá:gá dā kēgà dā:gā  
Uncle Hare and uncle Dog went to market –



# MORDVIN

850,000 SPEAKERS

*Russia*

**M**ordvin is the official name for two very closely related URALIC LANGUAGES, Erza and Moksha, spoken by a population group to the west of the middle Volga. Only a third of speakers live in Mordovia, now a self-governing republic within Russia: even here they are a minority, outnumbered by Russians. The republic is best known for its former secret city, Sarov (also called Moscow-2, Kremlyovsk, Arzamas-16), the centre of nuclear research which was once conducted largely by prisoners such as Andrei Sakharov.

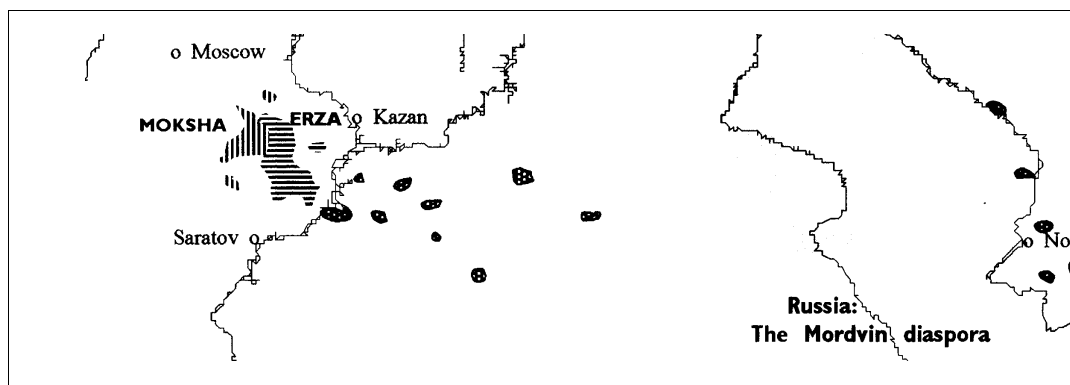
*Morden* were among the eleven 'Arctic peoples' who were conquered by the Gothic king Ermanaric in the 4th century, according to a Latin chronicler. Later they submitted to the rulers of Kiev. Russian interest in the Mordvin country had already begun with the capture and refounding of Nizhnii Novgorod, originally a Mordvin town, in 1221. ('Lower Novgorod' is so called to distinguish it from Novgorod the Great, the trading city north-west of Moscow. It was renamed Gorkii between 1932 and 1991.) Russian settlement gathered pace in

the 16th century after the Russians smashed the Tatar Khanate of Kazan, and it went hand in hand with Russian influence on the local languages.

Publishing, at first religious and educational, began in the 19th century. Saransk, capital of Mordovia, is the publishing centre. Literary Erza is based on the dialect of Kozlovka, while literary Moksha is based on the dialects of Krasnoslobodsk and Temnikov. Both use Cyrillic script.

Both languages are now heavily loaded with Russian loanwords, particularly Erza; Moksha shows greater Tatar influence. Mordvins have a long history of bilingualism in Russian and of assimilation to Russian society, yet the number of Mordvin speakers tended, at least until quite recently, to remain constant.

Verbs in Erza and Moksha have a complex range of forms, varying for both subject and object, for four tenses and for seven moods: Erza *vanok* 'look!'; *vanikselyiny* 'I wanted to look'; *vaninydyeryavlyiny* 'if I had looked'.



**Better than nothing?**

Er̥χ́í at̥iŋ́iĭt̥ bab̥iŋ́iĭt̥,  
 panda ṕ̥rasa kud̥iŋ́iĭt̥;  
 v̥e uš̥ tuma p̥eŋ́iĭt̥,  
 v̥e kraš̥t̥ima č̥eŋ́iĭt̥,  
 iũd̥ij š̥eŋ́χ̥ka š̥t̥iŋ́iĭt̥,  
 k̥eŋ́di p̥eķ̥e č̥oriŋ́iĭt̥,  
 k̥snav š̥el̥m̥e aŋ́ak̥s̥k̥iŋ́iĭt̥,  
 kaŋ́žur̥ š̥el̥m̥e avak̥s̥k̥iŋ́iĭt̥,  
 puluftuma p̥arn̥iŋ́iĭt̥,  
 š̥uruftuma sk̥aliŋ́iĭt̥!

There was an old man and old woman  
 In a little house on a hilltop;  
 They had one little log for burning,  
 With one little splinter for kindling,  
 One little girl with legs like reeds,  
 One little boy with a belly like a felt rag,  
 One little hen with eyes like peas,  
 One little chicken with eyes like mustard seeds,  
 One little foal without a tail,  
 One little cow without any horns!

This Moksha folk song is in seven-syllable lines, with a repeated rhyme supplied by the diminutive suffixes (English 'little . . .') on the last word of each line. The text is given in the standard transcription of the Société Finno-Ougrienne, in which an acute accent ´ indicates palatalisation, a very common feature of Moksha.

*Proben der mordwinischen Volksliteratur* gesammelt von H. Paasonen. Vol. 1:  
 Erza (Helsinki, 1891. *Journal de la Société Finno-Ougrienne*, 10) pp. 84–5

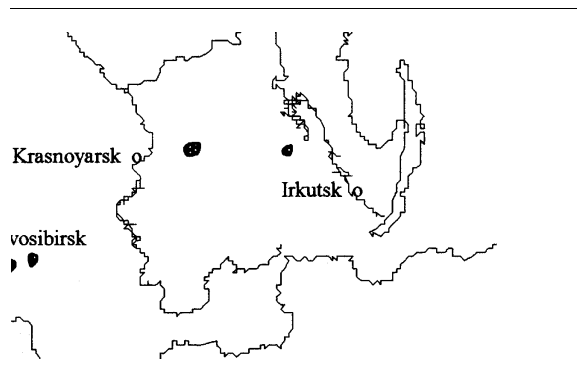
**Erza, Moksha and their neighbours: the map**

*Mordovskie yazyki*, 'the Mordvinian languages', is the official Russian term for this language pair. The Erza and Moksha are viewed by themselves as quite distinct, with independent traditions and no history of close relationships – or even of intermarriage. Moksha is now the language of the Moksha valley, while Erza is spoken in the Sura valley.

Mordvin migrations – often in reaction to Russian invasions and settlement – have made them one of the most scattered peoples of Russia.

**Numerals in Erza and Moksha**

Erza		Moksha
veyke	1	fkä
kavto	2	kafta
kolmo	3	kolma
nyilye	4	nyilyä
vetye	5	vetyä
koto	6	kota
syisyem	7	syisyäm
kavkso	8	kafksa
veykse	9	veyksa
kemeny	10	kemäny



# MOZARABIC

## *Extinct language of Spain*

U nique among the ROMANCE LANGUAGES, Mozarabic developed in an Islamic environment, in medieval southern Spain. Although the language of government was Arabic, Romance speech survived among the people of the Kingdom of Granada until eventually, after the Spanish *Reconquista*, it was overwhelmed by the Castilian Spanish of the Christian north.

Arabic *Musta'rib* 'Arabised' is the origin of the term *Mozarabic*.

For a long time it was thought that in Muslim Spain Romance speech had died out completely, to return from the Christian north during the reconquest – although there are local plant names, of Romance origin, in Arabic agricultural and botanical texts from Spain.

But ever since 1851, Mozarabic refrains have been known in Hebrew poetry from Arab Spain.

A hundred years later, S. M. Stern began to find Mozarabic refrains in Arabic poetry too. This is the oldest Romance literature from Spain – some texts are older than the *Poema del Cid* with which Spanish literature begins – and it is evidence of the language that was once spoken across all of central and southern Spain.

It is clear that many of the Arabic loanwords now found in Spanish, Portuguese and Judezmo must have reached those languages by way of Mozarabic. They include Spanish *alcachofa* 'artichoke', *arroz* 'rice', *albérrchiga* 'peach' and many other fruits, vegetables and flowers; *aldea* 'village', *alcalde* 'magistrate' and *almirante* 'admiral', which comes from the Arabic word known in English as *emir*.

Examples from W. D. Elcock, *The Romance languages* (London: Faber, 1960)

### Mozarabic lyrics: verse and refrain

A typical *muwashshah* verse, in Hebrew or Arabic, gives the atmosphere of a love affair –

‘When he is here the city is clothed in his glory,  
and she is transported to heaven with pride as long as he remains,  
She cries, on the day of his departure – ’

In Arabic poetics the rule of the *muwashshah* was that the refrain – the woman’s words – should be in colloquial language. In Spain, that meant Mozarabic:

ky fr'ýw	– ‘Que farayo	– ‘What shall I do,
‘w ky šýr'd dmyby	o que serad de mibi?	what will become of me?
ħbyby	ħabībī,	Lover,
nwn tytwlgś dmyby	Non te tolgas de mibi!’	do not leave me!’

Thanks to this rule, a forgotten language has been rediscovered. In Arabic script only the consonants are written (left column), and these brief texts pose enormous problems of interpretation.

After S. M. Stern, *Les chansons mozarabes* (Palermo: Manfredi, 1953) pp. 16–17

# MUNDARI

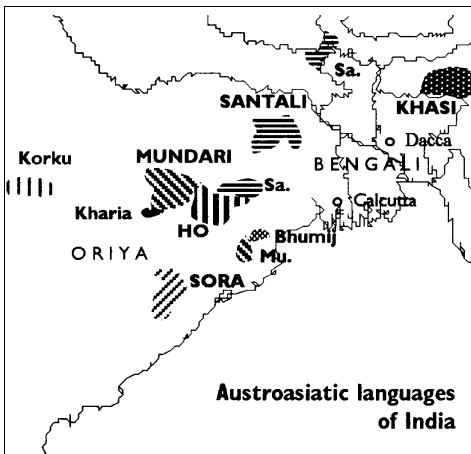
850,000 SPEAKERS

## India

**M**undari is one of the Munda group of languages, spoken by rural peoples of eastern India. These languages are now known to belong, with the Mon-Khmer group of south-east Asia, to the family of AUSTROASIATIC LANGUAGES.

By contrast with the Indo-Aryan and Dravidian groups, Munda languages have traditionally been unwritten and their peoples have little recorded history until they came into the sights of British administrators in India's colonial period. References can be found to forest peoples in classical Sanskrit texts, but their identification with speakers of Munda languages is a matter of faith rather than evidence.

There is now a limited periodical press in Mundari; publications have appeared in all three surrounding scripts, Devanagari (as for Hindi), Bengali and Oriya. The elaborate *Encyclopaedia Mundarica*, of which thirteen volumes were published in the 1930s and 1940s, is a storehouse of the language and its culture. It is in Latin script, as generally preferred by Christian missionaries.



As indicated by the accompanying Mundari song, Mundari are typically hill peoples while the related SANTALI (see table of numerals there) is spoken by lowland farmers.

Mundari and other Munda languages have been spoken alongside Indo-Aryan languages for two millennia. They show the signs of this long coexistence not only in loanwords – in both directions – but also in the structure of typical sentences, which can sometimes be paralleled, word for word, in languages of these quite independent groups.

### Major Austroasiatic languages of India: the map

KHASI has numerous dialect divisions. *Synteng* or Jaintia or Pnar, the language of a relatively large kingdom with a distinctive history, has perhaps 80,000 speakers. *Lynggam* is spoken to the west of other Khasi dialects by a people who show some cultural links with GARO speakers.

*Korku* has 320,000 speakers in southern Madhya Pradesh and northern Maharashtra.

*Kharia* has about 150,000 speakers in Bihar.

*Mundari* has 850,000 speakers, mainly in Bihar.

SANTALI has 4,000,000 speakers, mainly in West Bengal, but including 100,000 in western Bangladesh.

*Ho* is spoken to the south-east of Mundari, in Singbhum district of Bihar and Mayurbhanj district of Orissa. It has 750,000 speakers.

*Bhumij* has 100,000 speakers in Mayurbhanj district of Orissa. The number of speakers is declining rapidly.

*Sora* has 250,000 speakers in Ganjam district of Orissa. 'The British administrator's supreme

ignorance can be exemplified by the report of H. D. Taylor, Agent to the Governor in Ganjam, who in 1900 recommended that Government servants should be encouraged to learn the So:ra: language. He said that "the Sourah language

consists, I believe, of only 700 words and is not difficult to acquire; but it is spoken by hill-men" ' (D. P. Pattanayak in *Current trends in linguistics* vol. 5 ed. Thomas A. Sebeok [The Hague: Mouton, 1969] pp. 139-40).

### Who planted the mustard?

Búru re búru ré máni dó,  
 bérá re béra ré rái;  
 límáng lómongá máni dó,  
 kídár kóдорá rái.  
 Ókoé ge hérléda máni dó,  
 címaé ge pásiléđ rái?  
 límáng lómongá máni dó,  
 kídár kóдорá rái.  
 Múndakó ge hérléda máni dó,  
 Sántakó ge pásirleđ rái;  
 límáng lómongá máni dó,  
 kídár kóдорá rái.  
 Síde lége mónéña máni dó,  
 ʔóʔa' lége sánaña rái;  
 límáng lómongá máni dó,  
 kídár kóдорá rái.  
 Álo répe sídeá máni dó,  
 álo répe ʔóʔa' éa rái!  
 límáng lómongá máni dó,  
 kídár kóдорá rái.  
 Tĩire múndam gónóngte máni dó,  
 Káʔare póla sáʔite rái!  
 límáng lómongá máni dó,  
 kídár kóдорá rái.

Millet on every hill,  
 Mustard in every valley;  
 Millet dances,  
 Mustard sways.  
 Who planted the millet?  
 Who sowed the mustard?  
 Millet dances,  
 Mustard sways.  
 Mundas planted the millet,  
 Santals sowed the mustard;  
 Millet dances,  
 Mustard sways.  
 I want to pick the millet,  
 I want to nip the mustard.  
 Millet dances,  
 Mustard sways.  
 Don't pick the millet!  
 Don't nip the mustard!  
 Millet dances,  
 Mustard sways.  
 The millet costs a finger-ring,  
 The mustard costs a toe-ring:  
 Millet dances,  
 Mustard sways.

Ram Dayal Munda, 'Some formal features of traditional Mundari poetry' in *Austroasiatic studies* ed. Philip N. Jenner and others (Honolulu: University Press of Hawaii, 1976) pp. 844-71

*An or jadur* sung as one of a trilogy of songs during the Munda flower festival of February, March or April. In this text the accent ´ marks the strong beat of the verse; the repetition inherent in Mundari style (*buru re* 'on every hill') is offset by the irregular rhythm.



# MUONG

800,000 SPEAKERS

*Vietnam*

**M**uong is evidently related to VIETNAMESE (see map there); this 'Viet-Muong' group is now generally agreed to belong to the AUSTRASIATIC LANGUAGES.

*Muong*, the commonly used name for this people and their language, is a mistake – it is the word in Tai languages for a town and its territory. Speakers sometimes call themselves *Mon* or *Mwon*, however, like their distant relatives the MON of Burma. But neither in Muong nor in Vietnamese has there been any comprehensive term for the Muong. Near Vinh the Vietnamese call them *Nha Long*; in Quang Binh the name was *Nguon. Ao-ta*, *Sach* and *Tho* are also used – but the last serves also as the name for the Tai language of the northern Vietnamese plains.

Muong is spoken in inland northern Vietnam. Its cultural history has been completely different from that of Vietnamese, the language of a

centralised state, penetrated by Chinese culture, in which the arts of writing have been essential to everyday life.

Historically the most likely theory is that the speakers of early Muong and of early Vietnamese, two thousand years ago, lived not far apart. Chinese interest centred on the Vietnamese of the Red River delta: the Muong, in country areas to the south, remained untouched. Theirs is, and has no doubt always been, a rural, agricultural lifestyle, and it was only far later, three or four hundred years ago, that they came gradually to be incorporated in the Vietnamese state. This involved transferring some of the traditional power of the Muong hereditary noble families to Vietnamese-speaking civil servants.

Since this time Vietnamese has influenced Muong pervasively; there are now numerous French loanwords too. Like Vietnamese, Muong is a monosyllabic language. It has five tones.

# NA-DENÉ LANGUAGES

**T**he Na-Dené family of languages, apparently unrelated to any other, consists of one clearly identifiable group, Athapaskan, and three minor isolated languages which are much more distantly related, Tlingit, Eyak and Haida. The family grouping was postulated by the linguist Edward Sapir in a seminal paper in the *American anthropologist* in 1915: it remains controversial to this day.

The first element to be recognised, before Sapir's time, was that of the Athapaskan (or Athabaskan) languages, stretching from Alaska to Mexico in four separate geographical zones. NAVAHO, a language of Arizona and New Mexico, has the most speakers today, though its neighbours, the *Apache dialects* with around 15,000 speakers in Texas, New Mexico and Arizona, fill more of the map. The Athapaskan languages of California and those of Oregon are practically extinct ('there are perhaps no speakers under seventy years of age', said Michael Krauss in 1971). The most varied region is western Canada and central Alaska. Canadian languages of the group include *Carrier*, with about 3,600 speakers, Chipewyan with around 5,500, and *Slave* (pronounced, and sometimes spelt, *Slavey*) with about 2,000. In Alaska the number of speakers of Athapaskan languages is now very small but the variety greatest.

The three distantly related languages are also in Alaska and British Columbia. They are *Tlingit* (about 2,000 speakers in Alaska and British Columbia), Eyak and Haida. *Eyak* was first described by a Russian linguist in 1805. Long forgotten, and overlooked by Sapir himself, the Eyak were rediscovered in Alaska in 1930. By 1990 only one speaker remained alive.

If most specialists now agree that Tlingit and Eyak are somehow linked with the Athabaskan group, that is not true of *Haida*, with its 350

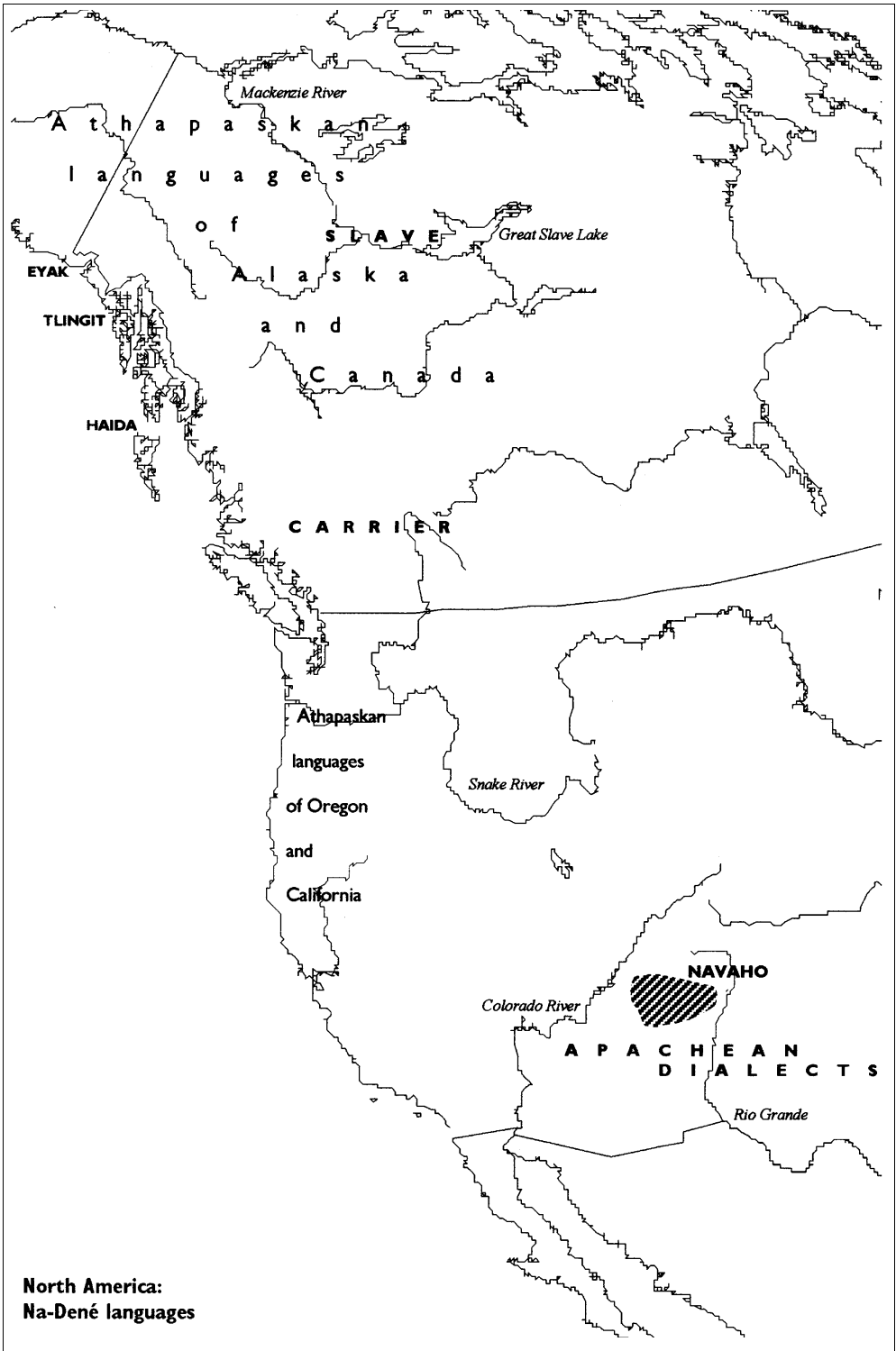
speakers. Here the controversy raised by Sapir has not died down: some, such as Krauss, consider Haida to be a linguistic isolate, with no demonstrated relationship to any other language.

Na-Dené, a relatively small family on the world scale, has not been convincingly linked to any wider grouping. In particular, Joseph Greenberg excludes Na-Dené, along with the Eskimo-Aleut languages, from the Amerind family in which he boldly classes every other language of native America. He suggests that Na-Dené speakers at their first arrival on the American side of the Bering Strait are to be identified with the Paleo-Arctic or Beringian culture of Alaska, dated by archaeologists to 8000 to 5000 BC.

Various proposed affiliations for Na-Dené, for instance with Sino-Tibetan or with Basque, though put forward by respected scholars, have not carried conviction.

## Edward Sapir

Sapir (1884–1939) studied under Franz Boas. He was a general linguist whose detailed research was largely in the area of American native languages. As part of a determined attempt to reduce the number of independent families into which American languages were then classified, Sapir renewed the idea, first proposed in 1798 and 1805, of a distant relationship between the Athabaskan languages on the one hand and Tlingit and Haida on the other. In 1929 Sapir published (in an article in the 14th edition of the *Encyclopaedia Britannica*) a classification of all Central and North American languages into six 'super-stocks', one of which was Na-Dené.



North America:  
Na-Dené languages

# NAHUATL

**PERHAPS 1,250,000 SPEAKERS OF NAHUA DIALECTS**

## *Mexico*

**N**ahuatl is the most important member of the Uto-Aztecan family, now believed by some to belong to a larger family of AMERIND LANGUAGES. Nahuatl was the language of the powerful Aztec civilisation so suddenly overwhelmed by the Spanish conquest in the 16th century. Nahuatl is still a major language of Mexico, a country proud of its Aztec inheritance – but Spanish is the only official language of the country.

Among the many languages of central Mexico it can be surprisingly difficult to make links with civilisations of the past. Who exactly were the Toltecs of AD 850 to 1100, masters of the Valley of Mexico in pre-Aztec times? The Aztec rulers, who took power after them, were invaders from the north: but it is unlikely that the Nahuatl language arrived only with the Aztecs. More probably the Toltec civilisation itself had already been Nahuaspeaking. The fame of the Toltec capital, Tula, destroyed about AD 1200, seems to survive in the legendary mother-city of *Tollán* known from 15th-century Nahuatl manuscripts.

The Toltecs are known to have sent colonies southwards around AD 900. *Pipil* is the language of this southern migration. Between them, Toltecs and Pipils destroyed much of classical Maya civilisation in their passing.

The Aztec intruders were in the Valley of Mexico in 1256. They founded Tenochtitlán in 1325, and began to expand their empire in the 15th century. Although the valley remained highly fragmented, linguistically and politically, the Aztecs soon dominated it – and were quickly

informed of the first Spanish landings, far to the east, in 1517.

Nahuatl, though little known to the mass of speakers of other Mexican languages, had spread widely in the previous two centuries and was by now the most useful long distance language in the region, essential to Spanish conquerors. There was certainly already some bilingualism in Nahuatl. That is why, even after the conquest, Nahuatl loanwords (as well as Spanish loanwords) multiplied in other Mexican languages: loanwords such as Nahuatl *tentzu* 'goat', Toton-tepec *teents*, Tzotzil *tentsun*. Enjoying Spanish favour, Nahuatl continued to spread while some other languages died away. It even became a kind of official language, recognised as such in a royal ordinance of 1570, though even then it was seen as a stepping-stone to Catholic conversion and eventual Hispanicisation.

From 1770 onwards, education in Mexico officially had to be in Spanish – but, after all, the reach of education was limited, and the position of Nahuatl was still not seriously undermined. The future of the language, two centuries later, is far from clear. In practice, the administration always favours Spanish and, crucially perhaps, Spanish is demanded by the new population mobility of the late 20th century.

Already at the Spanish conquest Nahuatl was being recorded in writing: in sculpted inscriptions, usually brief; and in picture books, which acted as a mnemonic for the reciter of orally remembered texts. The Spaniards destroyed many of these picture books, but the more enlightened among them encouraged the making of others, a record of Nahuatl beliefs and

culture, useful (at the very least) to priests who wished to understand the people whom they planned to convert to Christianity.

Certain manuscripts offer special help in understanding classical Nahuatl pictographic writing. Compiled under Spanish domination, *Codex Mendoza* is 'bilingual'. Pictograms of the ancient type, whose meaning would soon be lost, are here accompanied by interpretations in Spanish. This not only helps to explain the individual symbols and their variations. It also shows that they never corresponded to a precise text (which could have been *translated* into Spanish). Instead, Mexican pictograms stood for a sequence of ideas. Their makers certainly expected them to be read back in Nahuatl or Mixtec. But one who held the key could interpret them in any language. Thus the forgotten maker of *Codex Mendoza* has assisted modern scholars to reconstruct the key.

Bernardino de Sahagún, a Franciscan friar active in 'New Spain' from 1529 to 1590, gathered information from his Mexican students about their own culture and wrote it down in his 'General record of matters of New Spain', which is an encyclopaedia of Aztec religion, astronomy, astrology, economics, daily life, medicine, rhetoric, mineralogy and history. This vast work includes hymns to the Aztec gods, riddles and proverbs, and reproduces selected passages from the old picture books. It survives in two fairly complete versions, in manuscripts at Florence and Madrid.

Some of the pictographic manuscripts are historical, recording – with careful calculation of dates – migrations into the Valley of Mexico, wars between cities, and genealogy. Nahuatl traditional literature included formal speeches of praise and family history, given on public occasions. The manuscripts contain songs and myths: the so-called *Mexican Songs* include lyrics, war songs, mourning songs, burlesque, drama. It was a literature of great complexity, with certain styles specially belonging to different towns and regions. Poetic language involved metaphor, multiple compounds, parallel expres-

sions, and imagery based on jewels, feathers and flowers.

The range of Nahuatl literature shrank with the Spanish ascendancy; but a genre of Catholic religious plays flourished, and they are still performed in Nahuatl-speaking villages.

At a crossroads of Mexican culture, Nahuatl naturally contains loanwords from other American languages, such as Mixe-Zoque (because the early Olmec civilisation spoke a Mixe-Zoque language): from this source comes *nixtamalli* 'maize dough', a staple food; from Mixe-Zoque, too (proto-Mixe-Zoque \**kakawa*) comes the Nahuatl word for 'cacao' that was eventually borrowed into European languages (e.g. English *chocolate*, *cocoa*). Other important loanwords from Nahuatl in European languages include *tomato*, *avocado*. For things newly introduced after the conquest, Spanish loanwords in Nahuatl have multiplied: from Spanish *naranja* comes Nahuatl *laxa* 'orange'.

Because of its past importance as a lingua franca, loanwords from Nahuatl are to be found in many of the languages of Mexico, including Tepecano, Cora, Huichol, Tequistlatec, Huastec, Yucatec, Quiché and Cakchiquel. A Nahuatl-Spanish pidgin (see also SPANISH) was once widely used in central America. This is how Nahuatl *tequell* 'work' came to be borrowed into the Lenca language of distant Guatemala, just as Spanish *trabajo* 'work' was borrowed into other Amerindian languages.

### Printing in the New World

The first American printed book (1539) was a bilingual catechism in Spanish and Nahuatl. The bishop of Mexico, Francisco de Zumárraga, had imported a printing press to assist the spread of Christianity. The first Nahuatl grammar, by A. de Olmos, *Arte para aprender la lengua mexicana*, appeared in 1547.

Trained in Latin grammar, the Spanish-speaking friars who first codified Nahuatl grammar, in the 16th century, were open-minded enough to realise that they had to devise a new terminology for this utterly different tongue. To classify verb



forms they introduced terms such as ‘compulsive’ (causative) and ‘applicative’ (benefactive). The first ten numerals in Nahuatl are *ce, ome, yei, nahui, macuilli, chicuace, chicome, chicuei, chicunahui, matlactli*. It can be seen that the numbers ‘6’ to ‘9’ are formed on the pattern ‘5 plus 1’, ‘5 plus 2’ and so on.

### Uto-Aztecan languages

*Nahuatl* (or ‘Aztec’) is the only Uto-Aztecan language with a large number of speakers – and the only one which was the vehicle of a great civilisation. The others are scattered across the western United States and northern Mexico.

Nahua dialects are classified according to a particular phonetic trait. The lateral affricate or click, *tl*, is typical of the central group of dialects, which are thus properly called *Nahuatl*. Others are known as *Nahual* and *Nahuat*. Central Nahuatl is also known as *México* – language of the Valley of Mexico, after which the great modern city, and the whole nation state of which it is the capital, are both named.

*Pipil*, once a major language of colonial El Salvador and neighbouring states, is on the verge of extinction: a recent survey gives only twenty elderly speakers, the Pipil ethnic group having adopted Spanish as its mother tongue. Pipil is a

Nahua dialect, apparently established as a result of eastward conquests by the Toltecs in the 10th century.

*Mayo* is the next largest Uto-Aztecan language, spoken by about 50,000 people in Sonora and Sinaloa provinces of Mexico.

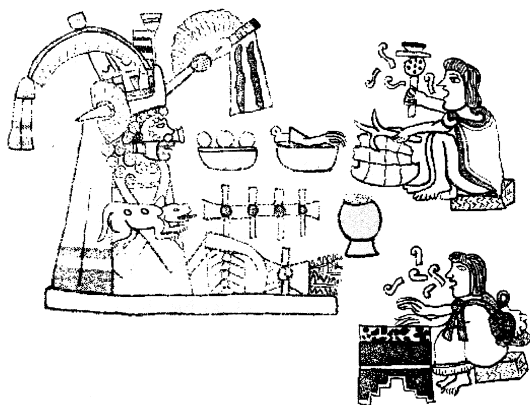
*Tarahumara* also has around 50,000 speakers in Mexico.

*Ute*, *Comanche*, *Hopi*, *Shoshone* and *Northern Paiute* are among the thirty remaining Uto-Aztecan languages with no more than a few thousand speakers each. Most are on the way to extinction.

### The language of human sacrifice

Emphases on war and human sacrifice are well-known features of pre-Columbian Mexican culture. These ideas seem, from archaeology, to have belonged to the conquering Toltecs, around AD 1000, and to have been learnt from them by peoples such as the Aztecs.

In *Codex Magliabechiano*, a second 'bilingual' manuscript, feast days are described. The Spanish text explains: 'This picture shows the unspeakable ritual of the Yndians on the day that they sacrificed men to their idols. Then, before the demon called Mictlantecutl, which means "Lord of the Dead" as explained elsewhere, they placed many cooking pots full of the human flesh, and shared it among the nobles and rulers and those who served in the demon's temple, called *tlamagatl*; and these shared their portions among their friends and families. They say that it tasted as pork tastes now, and this is why they like pork so much.'



# NAVAHO

PERHAPS 120,000 SPEAKERS

*United States*

Navaho is spoken in Arizona and New Mexico. As one of the NA-DENÉ LANGUAGES (see map there), it is closely related to the various Apache dialects, but – as far as anyone knows – totally unrelated to the majority of its Amerindian neighbours.

An anatomical atlas of Navaho names for parts of the body has over 400 entries. The Navaho and English-speaking compilers realised as they worked that while it is natural in English to start with the head and to work through shoulders, arms, hands, trunk, legs, in Navaho precisely the opposite order comes naturally: you start with the feet and finish with arms, hands and head.

Navaho has a rich oral literature, consisting partly of ceremonial poetry. This belongs to rituals sometimes known (in attempted English renderings) as Blessingway, Beautyway, Windways, Red Antway, Mountainway, Nightway, Shootingway. The Horse Songs that belong to the Blessingway ceremony are especially notable. There are some modern publications in Navaho:

there has been a newspaper in the language, *Adahooniligii*. The *Navajo historical series* has built up into a record of Navaho history and traditions.

Navaho is famous – among linguists – for its resistance to borrowing. Automobile terms, for example, tend to be new coinages in Navaho, not borrowings from English: some of them are formed by extending the meaning of anatomical words, such as ‘liver’ which now also means ‘automobile battery’ (according to Basso and Adams in the *American anthropologist* 1967–8). It has also been claimed that most American Indian languages seem to borrow rather little from English – because speakers are bilingual, and simply switch to English when speaking of topics that local languages do not cover.

Navaho once served as a lingua franca for the speakers of the isolated Zuni. When Hopi speakers fled as refugees to the Zuni, Navaho served as the common language between the two peoples.

Navaho distinguishes vowel length, nasality and tone. The first ten numerals in Navaho are: *lá ’í, nāki, tā’, dī’, ’acdlā’, xastā’ ’h, tsost’sid, tsēbī’’, náxást ’ēī, nēznā’ ’h.*

**Invocation**

Hayolkál behogán,  
nahotsoí behogán,  
kósyildil behogán,  
niltsabaká behogán,  
a’dilyíl behogán,  
niltsabaád behogán,  
thaditdín behogán,  
aniltani behogán,

House made of dawn,  
House made of dusk;  
House made of dark cloud,  
House made of male rain;  
House made of dark fog,  
House made of female rain;  
House made of pollen,  
House made of grasshoppers;

Opening invocation of a Navaho prayer. After Washington Matthews in *Navaho myths, prayers and songs* ed. Pliny Earle Goddard (*University of California publications in American archaeology* vol. 5 ii, 1907)



# NAXI

250,000 SPEAKERS

*China*

**N**axi is one of the SINO-TIBETAN LANGUAGES. It is sometimes said to be a member of the Burmese-Lolo group, but it may be better regarded as an independent descendant of proto-Tibeto-Burman. It is spoken in northern Yunnan, in and around Lijiang, in a district strongly defended by the encircling Yangtze gorges (for map see BAI).

The forms *Na-khi* and *Nahsi* are also found. It is the native term, and it means 'black man', a reminder of the relatively dark complexion of typical Naxi speakers. In older English writings *Naxi* is sometimes called *Moso*: this is still used as the name of one of the subdivisions of the Naxi people in the neighbourhood of Lugu lake.

The Naxi were apparently independent until their conquest by the Mongols in the 14th century. Thereafter they were subject to China as a tributary people, still with their own rulers, and still prosperous until recent times. The last Naxi prince was killed soon after the communist takeover in 1949. They are still one of the official nationalities of China, and Naxi is used in elementary classes in schools. In Naxi villages, as yet, few people speak Chinese.

Naxi is an essentially monosyllabic language – its open syllables and lack of consonant clusters give it a superficial resemblance to ƳI and the other Burmese-Lolo languages of China. Its vocabulary and basic structure seem quite different, however. It has been suggested that Naxi speech came from further north, perhaps two thousand years ago, and, in the course of its spread to local speakers in Yunnan, took on the sound patterns of their mother tongues.

Naxi used to be written in its own pictographic script, quite different from that of Chinese. According to legend, the script was invented by King Moubao Azong in the 13th century – yet it seems to preserve some reminiscences of an older time, when Naxi speakers lived in north-western China, with different fauna around them.

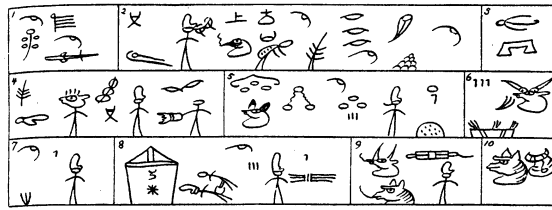
The script was used mainly by hereditary poet-priests, *dto-mba*, as an aid in the recitation of texts. 'Anyone can appreciate the graphic beauty of a Naxi text, but only someone well-versed in the mystical lore of Naxi religion can interpret its meaning and translate it into language' (S. R. Ramsey, *The languages of China* (Princeton, 1987) p. 267). The belief system of the Naxi is related to the pre-Buddhist Bon religion of Tibet, and to the Tibetan form of Buddhism itself. Naxi manuscripts can be found in many collections across the world, but its use has now almost died away. A standard romanisation, 'Naxi Pinyin', has been promulgated in China.

The Naxi writing system acts as an aid to the recitation of poetic texts, such as the tragic love story of K'a-mä of which this is the beginning. Not every word needs to be written – just enough to remind the reader what to say.

The poem is in five-syllable lines. A box of characters represents a sentence or sense-unit, usually several lines long. In this illustration the boxes have been numbered so that they can be compared with the translation.

Many of the pictograms are realistic, like the figure with a high headdress which means 'girl' (in the passage above it usually has to be translated 'I', since K'a-mä is talking about herself).

### Naxi pictographs



'Ghügh-daw muan ch'ung ho,  
muan-daw t'o dgyu ho.' . . .

1. 'Whether I weave gently does not matter;  
better not to weave at all.'

2. K'a-mä saw a crow and asked it to carry a message. 'A body is heavy, a message is not; a tree is heavy,  
its leaves are not; water is heavy, its foam is not.

3-4. To my lover's father Sse-shi, my lover's mother Ch'ung-ch'ung, my lover's friend La-yu,  
give a message:

5. In the sky above us, three male stars have not yet caught their three female stars; I am one of the three.

6-7. Above the village La-ler-dü the sheep have grazed on all but three tufts of grass; I am one of the three.

8. In my village, a handsome boy has made love to all but three girls; I am one of the three.

9. As you would call your cattle gently home, so call me.'

After J. F. Rock in *Bulletin de l'Ecole Française d'Extrême-Orient* vol. 39 (1939) pp. 25-9

Most characters correspond not to a sound but to an idea. For example, the crescent sign which occurs several times on this page is the negative. Sometimes the word *muan*, 'not', will occur in the corresponding line of verse, but not always; just as, in the English translation above, the negative is usually translated *not* but sometimes by *all but*.

Some characters are used phonetically, in a procedure which may be thought of as punning. For example, in box 4 the monkey-headed character means 'father-in-law', *yü-p'a*, because by coincidence *yü* is the word for 'monkey'. This is one of the methods by which, in the distant past, the Chinese writing system, too, developed its range and flexibility.

# NDEBELE

1,600,000 SPEAKERS

*Zimbabwe, South Africa*

**O**ne of the Nguni group of BANTU LANGUAGES, Ndebele is the second language of Zimbabwe (after Shona). Southern Ndebele has an equal number of speakers in the Transvaal, where it is one of the eleven official languages of South Africa (see map at ZULU).

Speakers call their language *isiNdebele* and themselves *amaNdebele*. The word is correctly pronounced with three short *es* of which the first is stressed, *Ndē'bělē*.

*AmaNdebele* appears to have been their own version of the name given to them by the Sotho and Tswana speakers in whose land they were settled – *maTebele*. Early English-speaking explorers heard and adopted this Sotho form, calling the people *Matabele* and their northern domain *Matabeleland*.

Southern Ndebele is a language of the Nguni group, close to Zulu. It may represent a gradual spread or a migration northwards of Nguni-language speakers: at any rate it was probably established in the southern and central Transvaal by the 16th century.

The Ndebele language of Zimbabwe arrived there as the result of historical events of the early 19th century. Its origin is in the adventurous story of Mzilikazi, a captive prince who was enrolled as one of the warriors of the Zulu king Shaka. Having offended Shaka, Mzilikazi fled northwards with his followers and at first settled in the Marico valley of Transvaal, in a region inhabited by speakers of Tswana and Southern Ndebele. He ruled there from 1832 to 1837.

After fighting with the advancing Boers, Mzilikazi led a second armed migration, this time probably with larger numbers, and (so legend says) he was advised on what route to take by the missionary Robert Moffat, known to Ndebele speakers as Mshete. The leaders of the expedition were to seek a strangely shaped hill: near modern Bulawayo, on the high Zambezi-Limpopo watershed, in modern Zimbabwe, this hill – *Intaba yezinduna*, the 'hill of the viceroy' – was found. It is at the

centre of the present territory of Ndebele speakers.

Originally the language of a military caste, Ndebele spread rapidly among the conquered population of the immediate neighbourhood – most of them probably speakers of Shona dialects. The changeover was speeded by the fact that the captives and the conquered were able to become warriors themselves – for which purpose they had to adopt the Ndebele way of life and the Ndebele language.

Mzilikazi and his successor Lobengula dominated Shona-speaking country and raided far and wide, eventually coming into conflict with the British South Africa Company, which at last drove Lobengula out and annexed his kingdom. Thus, after only fifty-five years of Ndebele supremacy, 'Southern Rhodesia' was under English-speaking rulers from 1893 to 1980. The great city of Bulawayo grew up, and much farming land was seized. Yet, at the end of this period, when Zimbabwe achieved independence, Ndebele remained solidly established as the regional language of the Bulawayo district, just as it had been at the beginning.

Zulu and Ndebele are still to some extent mutually intelligible, though idioms differ and Ndebele has clearly borrowed numerous terms from the languages previously spoken in its territory. It retains the click consonants typical of the Nguni group, written *c*, *q*, *x* – the last 'the same as the sound made by a carter to urge on his horse', as one linguist observed.

Most speakers of Ndebele are now bilingual in Shona, the majority language of Zimbabwe. Although the two languages are quite distinct, there is certainly a convergence at work between them. Here, modern political conditions reinforce an earlier tendency, for at the very beginning of its history a large proportion of the speakers of Ndebele had learnt a Shona dialect as their mother tongue.

The first ten numerals in Ndebele are: *nye, bili, tatu, ne, hlanu, tandatu* or *capakanye, capakabili, capakatatu, capakane, itshumi*.

# NEPALI

8,000,000 SPEAKERS

*Nepal, India*

**N**epali, one of the Pahari group of INDO-ARYAN LANGUAGES (see map there), is the national language of Nepal, also spoken in the Indian states of West Bengal and Sikkim and in south-west Bhutan.

It is still sometimes called by its old name, *Khas Kurā*, 'Khasa language', a reminder that it first spread across the foothills of Nepal as the everyday language of the Khasa kingdom, which ruled here in the 13th and 14th centuries. Its collapse led to the emergence of numerous petty princedoms using the same mother tongue, including Porbat (whence the language has often been called *Parbatiyā*) and Gorkha. Since Rajput dynasts founded several of these small states, it is often said that the Nepali language itself was brought to Nepal by the Rajputs, but this is definitely untrue.

In 1768–9 Pṛthvīnārāyaṇ Śāh, prince of Gorkha, conquered the NEWARI-speaking Nepal valley, including the city of Kathmandu, which he made his capital. This was the foundation of the modern kingdom of Nepal – and it explains why his own state language, and not the language of Kathmandu itself, became the national language. The term *Nepalese* was first used in a grammar published in Calcutta in 1820 – but *Parbatiyā* remained the official name in Nepal itself until the 1930s.

Since it spread from the state of Gorkha Nepali has also been called *Gorkhali*. This became the special name of the military lingua francas, slightly altered varieties of the language, which came into everyday use in the Nepalese Army (*Gorkhālī bhāṣā*, 'language of Gorkha', in a drill manual of 1874), then the British army (*Gurkhali*) and the modern Indian army (*Gork-*

*hali*). These soldiers are universally called 'Gurkhas'. For Nepali, already widely used in the lower valleys, has continued to spread rapidly, partly as a first language, partly as a lingua franca. The higher valleys of Nepal, Sikkim and Bhutan, where communication is so difficult, speak a bewildering range of *Rāī* and other Sino-Tibetan languages.

Nepali shows the influence of its Sino-Tibetan neighbours. It has an unusual variety of personal pronouns. The second and third person pronouns ('you, he, she') have three alternative forms for degrees of respect. In the second person, for example, an inferior is addressed as *tañ*, an equal as *timi*, a superior as *tapaiñ* or *āphu*. For a table of numerals see KUMAUNI.

जति राई उति कुरा || *Jati Rāī uti kurā*, says the Nepali proverb: 'There are as many languages as there are Rāīs'. The linguistic complexity of the higher valleys, where Rāī and other Sino-Tibetan speakers live, explains the rapid spread of Nepali as a language of communication in the southern Himalaya.

## 'Gurkhali' and Nepali in writing

**The British Army's standard Roman script and the Devanagari equivalents**

a ā i i u e ai o au k kh g gh ng ch chh j jh ñ ṭ ṭh ḍ ḍh ṇ t t h d dh n p ph b bh m y r l v sh s h

अ आ इ ई उ ऊ ए ऐ ओ औ क ख ग घ ङ च छ ज झ ण ट ठ ड ढ ण त थ द ध न प फ ब भ म य र ल व श स ह

# NEWARI

500,000 SPEAKERS

*Nepal, India*

One of the SINO-TIBETAN LANGUAGES, Newari is the historic language of the Nepal valley, with its metropolis of Kathmandu. It is only recently that the Indo-Aryan speech of the states of Porbat and Gorkha has come to challenge the position of Newari and has taken the name NEPALI.

By its own speakers Newari is called *Newā: bhāy* or *Nepā: bhāy*, 'Nepal language'. They prefer to call Nepali by an older name, *Khay bhāy* 'Khasa language'.

As one of the row of Sino-Tibetan languages of the southern slopes of the Himalaya, Newari has probably been spoken in its present location for a long time. It is less closely related to Tibetan than are some of its neighbours. Records show that for several centuries at least it was the ruling language of a kingdom at Kathmandu – a kingdom that had participated fully in the Buddhist culture of north India, and retained Buddhism when India lost it.

Newari inscriptions are very numerous from the long reign of Jayayaksa Malla (1428–80) down to the Gorkha conquest of the Nepal valley in 1769. Early inscriptions are often in a mixture of Sanskrit and Newari: the language differs from its relatives to the north precisely in the extent of Sanskrit influence and the great number of Sanskrit and Indo-Aryan loanwords.

Newari was already cultivated as a literary language in the 14th century and perhaps earlier. 'Classical Newari [*le névari de la belle époque*]' shows a harmonious balance between the Himalayan dialects, still undeveloped owing to their isolation, impoverished, rough, incapable of expressing advanced thought and abstract ideas,

and the dialects that had become entirely Hinduised through borrowing from the Indo-Aryan languages of the plains. Newari expanded its vocabulary through its native resources, and although it did borrow from the modern Indo-Aryan languages it was able to assimilate these loans and to enrich itself in the process. A considerable number of Newari commentaries on Sanskrit Buddhist texts, and also of Newari translations, survive' (Sylvain Lévi, *Le Népal* (Paris, 1905) vol. 1 pp. 251–2).

The Gorkha conquest marked the end of Newari's role as a state language. It was actually banned from public use between 1846 and 1950, and the number of speakers has probably been in decline for some time.

According to the chronicle *Kirāt mundhum* the goddess Sarasvatī appeared to a 9th-century Limbu king, Sirijanga, took him into a cave, and showed him a series of stone tablets that told the story of the creation – the story with which the chronicle itself begins. Since that time, it is said, the script in which these miraculous tablets were written has been lost and rediscovered several times, the most recent rediscoverer being Imānsiṃha Cemjong, the 20th-century champion of Limbu culture.

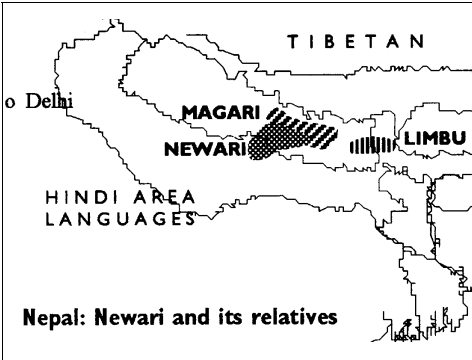
Newari was traditionally written in a local script, known in many variants from different periods, related to medieval forms of the Nagari and Bengali scripts of north India. The same script was used locally for Sanskrit texts. When Newari is printed it now appears in the familiar Devanagari alphabet that is used for Nepali. Unlike Nepali, however, Newari has a distinction of vowel length. Long vowels are indicated

by : in the script. The first ten numerals in Newari are: *chi, nasi, sō, pi, ngā, khu, nhaye, cyā, gū, sānha*.

**Newari and its neighbours**

*Newari* is the local language of Kathmandu and the Nepal valley. There are some speakers in Bettiah in the Indian state of Bihar.

*Limbu* or Kirāt has 200,000 speakers in eastern Nepal and Sikkim. It claims an independent literary tradition, but no early texts survive except for a religious chronicle, *Kirāt mundhum*, which has been given the grandiose name ‘the Veda of the Kirāts’. The disappearance of old literature is ascribed to an order of the King of Nepal, in



1788, that all Limbu books were to be burnt.  
*Magar* or Magari has 500,000 speakers in the central mountains of Nepal.

# NIGER-CONGO LANGUAGES

As they studied the languages of Africa south of the Sahara, 19th- and 20th-century scholars noted pervasive similarities that, just as in other parts of the world, led to the identification of families each descended from a single postulated proto-language. Some groupings, the result of more recent migration, were easier to recognise than others: the relationship among the Bantu languages, for example, was clear by the 1850s. Wilhelm Bleek, in 1856, was an early proponent of the wider grouping, covering both southern and western African languages, that is now usually known as Niger-Congo: he called it *Bantu*.

The Niger-Congo language family, or a similar grouping, has had many names: *Bantu*, *West African*, *Western Sudanic*, *Western Nigritic*, *Niger-Kordofanian*. Names like *Mixed Negro* (and *Semi-Bantu*, see below) were applied as a result of a theory – now seen to be quite unhistorical – that they originated in a mixture between Bantu languages, with their typical noun classes, and AFROASIATIC LANGUAGES which typically have noun gender. The term *Niger-Congo* was invented by Joseph Greenberg, who developed his seminal reclassification of African languages between 1949 and 1963.

Diedrich Westermann had already clearly demonstrated in 1911 and 1927 the relationship of most of the languages now counted in the Niger-Congo family, and later linguistic study of Niger-Congo stems from his work.

Westermann, however, like other scholars of his time, tended to assume two coordinate divisions of the family, Bantu and ‘Western Sudanic’. Even Malcolm Guthrie, who worked so fruitfully on Bantu classification in the 1950s and 1960s, could not accept Greenberg’s demonstra-

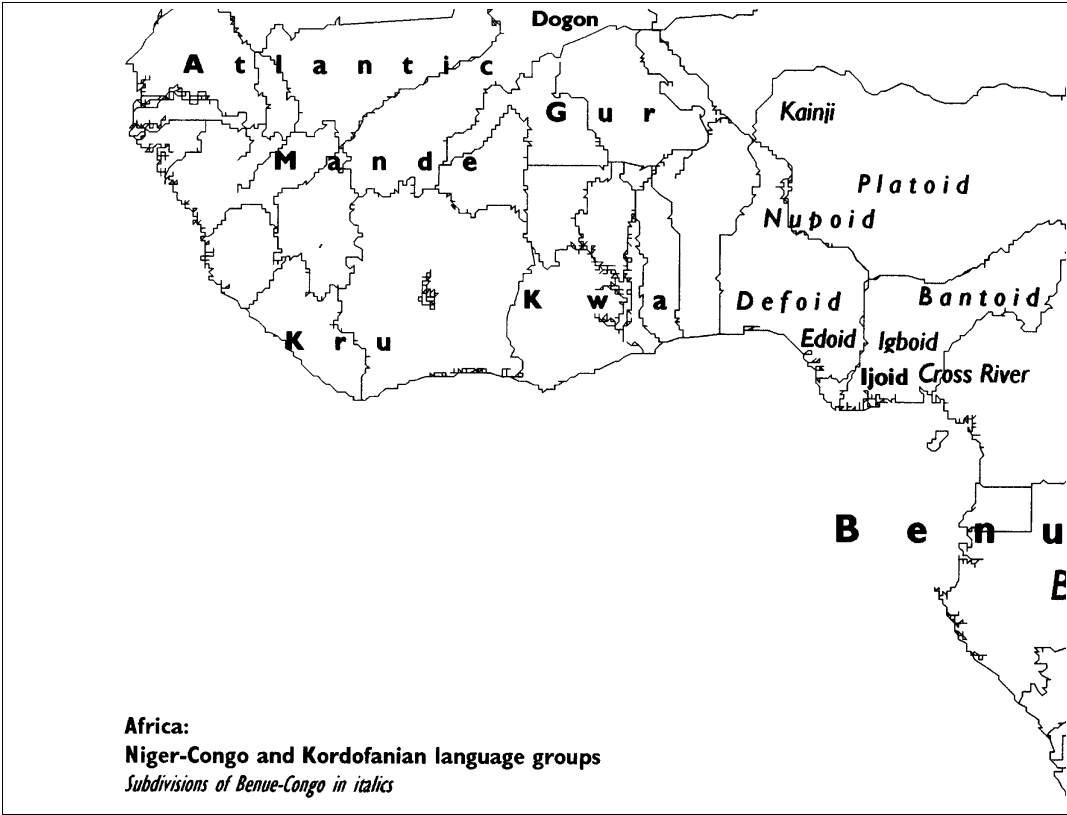
tion that the huge group of Bantu languages was historically an offshoot of a subgroup of a group of one of the main divisions of the Niger-Congo family.

There is a great deal of controversy over the higher levels of classification – which represent population divisions and migrations that took place many thousands of years ago. Greenberg counted West Atlantic, Mande, Gur, Kwa, Benue-Congo and Adamawa-Eastern as six coordinate divisions. Kordofanian languages (see below), Greenberg considered, had separated from the family at an earlier date: hence his alternative name for them, Niger-Kordofanian. It is quite widely agreed now that not only Kordofanian, but also the Mande and Atlantic groups and possibly the Ijoid group, split off at a very early stage from the remainder of the family – and the remainder is sometimes called Volta-Congo. Some believe that Niger-Congo and NILO-SAHARAN LANGUAGES, long seen as entirely distinct, will turn out to be distantly related; and that the Mande languages and SONGHAY belong neither to the one family nor to the other, but together form a third family distantly linked to both.

There are hundreds of Niger-Congo languages, though many have only a few thousand speakers each. The lists, here and at BANTU LANGUAGES, show all those with more than 100,000 speakers. The most important of all, typically with over a million speakers, have separate entries in the *Dictionary*, and these are signposted in the following survey.

## Kordofanian languages

Isolated and with no close relatives, the Kordofanian languages are spoken by small communities in the Nuba Mountains of Southern



Kordofan province, Sudan. Thilo Schadeberg, the linguist who has spent most time on this group, believes that there are about 165,000 speakers. The four dialect groups, centring on the towns of Heiban, Talodi, Rashad and Katla, are quite distantly related: Rashad was the capital of Tegali, a significant kingdom in the 19th century.

Close at hand to the west are the 90,000 speakers of Krongo and other Kadugli languages, spoken to the north-west and south-east of Kadugli in the same province. It is geographically handy to link Kadugli with Kordofanian, but in fact the relationship is quite unproved.

**Mande, Atlantic and Ijoid groups**

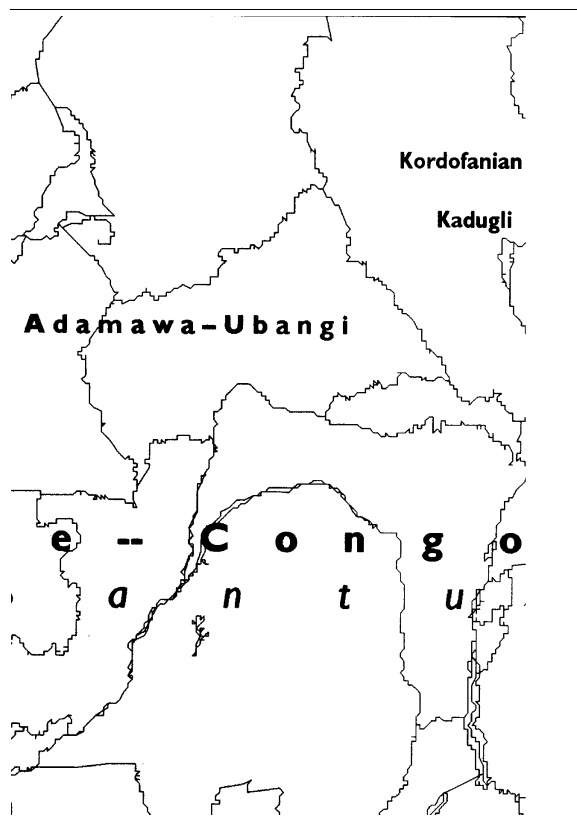
Mande-speaking peoples now live in widely scattered districts of inland West Africa. This group of languages may have been developing separately from the main Niger-Congo family for

twelve thousand years or more, and there must have been many unrecorded migrations in that time. But the large area now covered by MANDEKAN, SUSU and SONINKE is certainly the result of fairly recent historical expansion. Although they had probably begun to differentiate from one another by about 2000 BC, the Mande languages (which also include MENDE) were easily seen by 19th-century linguists to be related: Sigismund Koelle's *Polyglotta Africana* (1854), a collection of comparative wordlists, made the relationship obvious.

At least five Mande languages – Mende, Loma, Kpelle, Vai (see below) and Mandekan – have their own scripts, invented by local speakers in the course of the 19th century. There are also indigenous scripts for at least two languages – Fulani and Wolof – in the Atlantic group.

The Atlantic languages (Greenberg's West Atlantic) were probably the next to branch off from Niger-Congo. They are now spoken along





the Atlantic coast of West Africa, from Mauritania to Liberia, and include Serer, TEMNE and DIOLA. WOLOF, the language of coastal Senegal, was the first African language encountered by Europeans in their seaborne exploration of the Atlantic coasts in the 15th century. The most important of this group is FULANI, whose speakers, traditionally nomadic herdsman, founded an Islamic empire in eastern Nigeria in the 19th century. Fulani now has the widest geographical range of any single African language.

The IJO language of the Niger Delta seems to be an isolated member of Niger-Congo. It also separated very early, and has sometimes been thought to be an aberrant relative of the Mande group.

### Central Niger-Congo

Perhaps ten thousand years ago the proto-Volta-Congo language, from which Kordofanian,

Mande, Atlantic and Ijoid had already separated, began to split into several dialects which are the origin of the remaining divisions of Niger-Congo.

The KRU LANGUAGES (sometimes included in the Kwa group, below) are spoken in the forests of southern Liberia and Ivory Coast. Kru people worked on European ships as early as the 16th century, and there are still Klao, Grebo and Bassa-speaking communities in several West African port cities. Bété, of the Ivory Coast, has the largest number of speakers of any Kru language today.

Senufo, sometimes spoken of as a single language rather than a dialect group, straddles the frontiers of Ivory Coast, Burkina Faso and Mali. It is hard to place in any Niger-Congo subgroup. *Senari* is the dialect with the largest number of speakers. The first ten numerals in the Bamana dialect are: *nene*, *shyoni*, *tare*, *tityere*, *kaguru*, *ghani*, *gharashyo*, *shyolake*, *untu-lake*, *kantoke* (*gharashyo* means 'second 6'). These numerals show scarcely any similarity with those of Mòôre.

The Gur or Voltaic languages, spoken in the savanna country of Burkina Faso and the northern parts of Ivory Coast, Ghana, Togo and Benin, form a highly disparate group, difficult to classify. Apart from MÔÔRE, widely used as a second language in Burkina Faso, major members of the group are Dagaari, Gourmantché, Frafra or Gur-enne and (some say) the Senufo dialects. Dogon, a language of eastern Mali, used to be counted among the Gur languages, but the relationship is doubtful: it has also been considered a member of the Mande group.

The Adamawa-Ubangi languages (Greenberg's Adamawa-Eastern) are the easternmost of any in the Niger-Congo family except Kordofanian and Bantu. They divide into two subgroups. Historically the most important Adamawa language is Mbum, which used to be a lingua franca of the Tibati-Ngaoundéré region but has now given way to Fulani. Languages of the Ubangi subgroup were carried far to the east by recent migrations of conquering peoples, notably ZANDE. Ngbandi is important as the basis

of the creole SANGO, now the national language of the Central African Republic. Two further members of this group are GBAYA and BANDA.

The Kwa languages (merely a subdivision of the much more miscellaneous Kwa group postulated by earlier scholars) include three of the most important Ghanaian languages, AKAN, EWE and GA, and a major Ivory Coast language, BAULE.

**Benue-Congo Languages**

By far the largest subdivision of the Niger-Congo family is Benue-Congo. Once known as Semi-Bantu or Benue-Cross, the nature of its relationship with the BANTU LANGUAGES was

clarified by Greenberg: the latter are simply an offshoot of the Bantoid group of Benue-Congo. In this *Dictionary* they are given a separate entry because of their large number and their geographical extent. The remaining Bantoid languages are spoken in a much smaller, well-defined region of Nigeria and Cameroun: the best known of them is TIV.

EFIK belongs to the Cross River subgroup. Two of the national languages of Nigeria, YORUBA and IGBO, are Benue-Congo languages, along with EDO, language of the Empire of Benin. Proto-Benue-Congo may have been spoken around the Niger-Benue confluence, from where the existing languages have gradually spread.

**The subdivisions of Niger-Congo**

**Major Niger-Congo languages**

AKAN	5,000,000	Ghana, Ivory Coast
Attié	220,000	Ivory Coast
Awutu	100,000	Ghana
Balanta	300,000	Guiné
Bamendjou	100,000	Cameroun
Bamun	220,000	Cameroun
BANDA	1,000,000	Central African Republic
Bassa	300,000	Liberia
BAULE with Anyin	2,000,000	Ivory Coast, Ghana
Bété	700,000	Ivory Coast
Biom	200,000	Nigeria
Bobo Fing	175,000	Burkina Faso, Mali
Boomu and Bwamu, or Bobo Wule	450,000	Mali, Burkina Faso
Borgu or Bariba	250,000	Benin, Nigeria
Buli	150,000	Ghana, Burkina Faso
Bullom or Sherbro	175,000	Sierra Leone
Busa and Boko	100,000	Benin, Nigeria
Dagaari with Birifor and Waali	650,000	Ghana, Burkina Faso
Dan and Toura	500,000	Ivory Coast, Liberia
DIOLA	500,000	Senegal
Ditammari and Tamberma	120,000	Benin, Togo
Dogon	500,000	Mali and Burkina Faso
Ebira or Igbira	500,000	Nigeria
EDO	1,350,000	Nigeria
EFIK and Ibibio	2,750,000	Nigeria
EWE and Fon	3,500,000	Togo, Benin, Ghana
Frafra or Gurenne	500,000	Ghana, Burkina Faso
FULANI	15,000,000	West African countries

GA	1,000,000	Ghana
Gbagyi	250,000	Nigeria
GBAYA	850,000	Central African Republic, Cameroun
Ghomala'	250,000	Cameroun
Gokana and Kana	150,000	Nigeria
Gonja	120,000	central Ghana
Gourmantché or Gurma	400,000	Burkina Faso, Togo, Benin, Niger
Grebo and Krumen	250,000	Ivory Coast, Liberia
Gua	120,000	Ghana
Guéré, Wobe and Krahn	450,000	Ivory Coast, Liberia
Guro and Yaouré	220,000	Ivory Coast
Idoma	300,000	Nigeria
Ife or Ana	100,000	Togo, Benin
Igala	800,000	Nigeria
IGBO	12,000,000	Nigeria
Igede	120,000	Nigeria
IJO	600,000	Nigeria
Isoko	300,000	Nigeria
Itsekiri	500,000	Nigeria
Jarawa	150,000	Nigeria
Jju or Kaje	300,000	Nigeria
Jukun dialects	125,000	Nigeria, Cameroun
Kabiyé or Kabré	400,000	Togo, Benin
Kambari	100,000	Nigeria
Kissi	450,000	Liberia, Sierra Leone, Guinea
Klao	200,000	Liberia and West African port cities
Kom	120,000	Cameroun
Konkomba	350,000	Togo
Kono	120,000	Sierra Leone
Kpelle	600,000	Guinea and Liberia
Kulango or Koulan	175,000	Ivory Coast, Ghana
Kusaal	200,000	Ghana, Burkina Faso
Lamnso'	120,000	Cameroun
Limba	250,000	Sierra Leone
Lobi	200,000	Burkina Faso, Ivory Coast
Loma	120,000	Liberia
Mande of Burkina Faso	400,000	Burkina Faso, Ghana
MANDEKAN	5,000,000	Mali, Senegal, Guinea, Ivory Coast, Burkina Faso, Sierra Leone
Mandyak, Mankanya and Papel	250,000	Guiné
Manja	120,000	Central African Republic
Mano	150,000	Liberia
Mbanza	220,000	Congo (Kinshasa)
Medumba	200,000	Cameroun
MENDE	1,000,000	Sierra Leone, Liberia
Moba and Bimoba	200,000	Togo, Burkina Faso, Ghana
MÔÔRE	3,000,000	Burkina Faso
Nawdm	100,000	Togo

Ngbaka	900,000	Congo (Kinshasa)
Ngbandi	210,000	Congo (Kinshasa), Central African Republic
Ngyemboon	100,000	Cameroun
Northern Samo	125,000	Burkina Faso
NUPE	1,000,000	Nigeria
Nzema and Ahanta	350,000	Ivory Coast, Ghana
Rubassa	100,000	Nigeria
SANGO	5,000,000, mainly as second language	Central African Republic
Senufo of Ivory Coast, or Cebaara or Senadi	450,000	Ivory Coast
Senufo of Mali, or Mamara or Bamana	300,000	Mali
Serer	650,000	Senegal, Gambia
Sisaala	120,000	Ghana, Burkina Faso
SONINKE	1,000,000	Mali, Senegal, Mauritania
SUSU	700,000	Guinea, Sierra Leone
Tarok	140,000	Nigeria
Tem	300,000	Togo, Benin
TEMNE	950,000	Sierra Leone
TIV	1,500,000	Nigeria
Toma	120,000	Guinea
Urhobo	340,000	Nigeria
Vai	75,000	Liberia, Sierra Leone
WOLOF	2,000,000	Senegal, Gambia
Yemba-Nwe	350,000	Cameroun
YORUBA	20,000,000	Nigeria, Benin
ZANDE	1,200,000	Congo (Kinshasa), Central African Republic, Sudan

**The Vai syllabary**

*Vai*, a Mande language of about 75,000 speakers in inland Liberia and Sierra Leone, is most notable for its special script, invented by Momolu Duwalu Bukele in the 1830s. It came to him in a dream, or so he told Sigismund Koelle (perhaps the first European to hear of the script, in 1849). Though not taught in schools, the two hundred characters of the Vai syllabary continue to be used in letter-writing, diaries and Bible translations.

Vai was the language of a warrior people of the mid 16th century, known in European writings of the period as *Mani* or *Kquoja*.

**Polyglotta Africana**

Sigismund Koelle made a great contribution to African linguistics with the publication of *Polyglotta Africana* in 1854. This was a collection of almost 300 words and phrases in well over a hundred languages – all collected from freed slaves settled in Sierra Leone. It was criticised by some 19th-century scholars, but the real test is that Koelle’s book was used then and is still used now. For some languages, even now, no better vocabulary exists. Practically all the languages he listed can be identified (though he had never visited the places where they originated).

Koelle’s work was published by the Church Missionary Society in London. He also worked on Vai and its syllabary, and on KANURI.

# NILO-SAHARAN LANGUAGES

**T**he linguist Joseph Greenberg, in 1963, proposed the recognition of a Nilo-Saharan language family, grouping together numerous languages of Uganda, Ethiopia, Sudan and Chad, for many of which no generally accepted family links had previously been suggested. Greenberg's proposals have partly been confirmed by later research. The languages and language groups do differ quite radically from one another, suggesting that many thousands of years have passed since the postulated proto-Nilo-Saharan language may have been spoken.

Across a belt of north central Africa, all the way from Lake Turkana to the middle Niger, archaeologists have identified a highly specialised way of life that came into existence in the eighth millennium BC, when this area was a good deal wetter than it is today. This 'aquatic tradition', making full use of lake and river resources, has been linked by J. E. G. Sutton and Patrick Munson, in the 1970s, with the initial expansion of Nilo-Saharan languages. Not all linguists agree: some believe that proto-Nilo-Saharan must be dated rather more than ten thousand years ago. But it is a striking fact that the modern languages of the group, from Songhay to Turkana, actually map this 'aquatic tradition' rather accurately.

## The isolates of Nilo-Saharan

Far distant from the other members of the group, and long considered quite unrelated to other known languages, SONGHAY is spoken by over two million people in the middle Niger valley. Half a million people in the southern Sudan speak Fur. The little-known Gumuz is spoken at the confluence of the Diddesa and Blue Nile in Ethiopia. Kunama, a language of Eritrea, is usually classified in the Chari-Nile

group, but is very different from its claimed relatives.

Catholic missionaries have competed with Norwegian Lutherans for converts among the Kunama. The result: a good supply of Kunama books in the Latin alphabet – very unusual for former Ethiopia, in which most languages are printed in the Ethiopic script – but with two different sets of spelling rules.

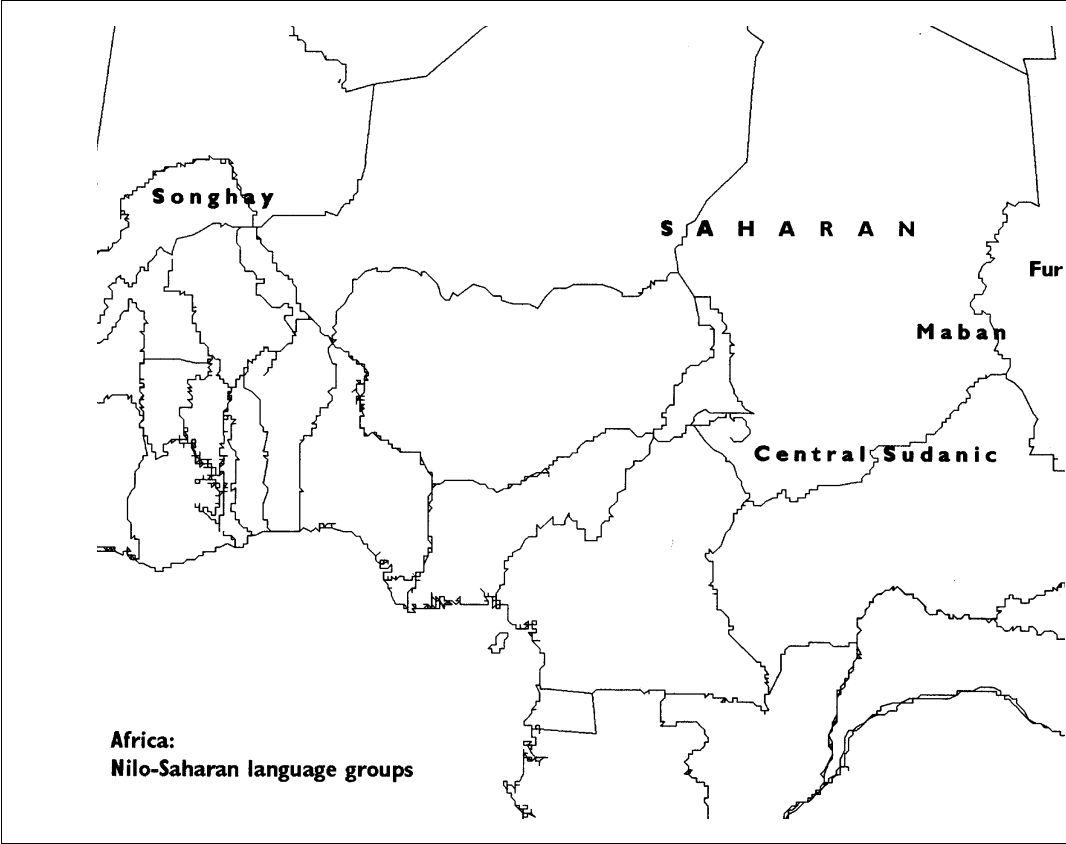
The Saharan languages include KANURI, a major language of northern Nigeria, along with Teda and Zaghawa. The small Maban language group includes MABA, Massalat and Mimi. There is another small, distinct group called Koman, the largest member of which, Kwama, is spoken on the Ethiopia–Sudan border.

## The Chari-Nile group

Greenberg's Chari-Nile group of languages, which he made the main constituent of his Nilo-Saharan family, is not accepted by all scholars as a real unity. Its two subgroups and one language isolate (Berta, an Ethiopian language) can instead be regarded as immediate components of Nilo-Saharan.

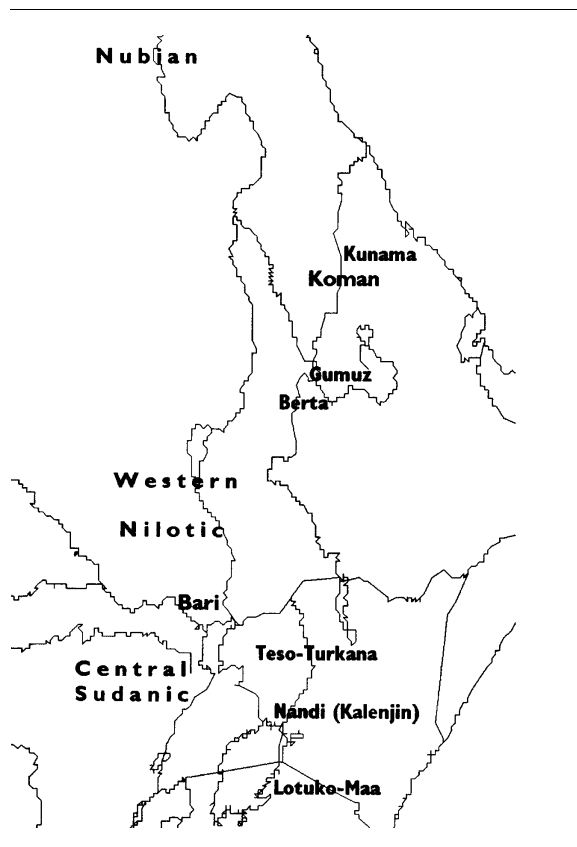
The Central Sudanic subgroup includes languages of Congo (Kinshasa), Uganda and Chad: Lendu, Logo, Lugbara, Madi, Mangbetu, and SARA.

The East Sudanic subgroup is the most extensive geographically, extending from Sudan to Tanzania. Many of these languages have only small numbers of speakers. There are separate entries in this dictionary for KALENJIN (the major *Nandi* language) and the NUBIAN LANGUAGES; Pökoot and Datoga also belong here. But two subdivisions within East Sudanic are especially important.



**Major Nilo-Saharan languages**

Acholi	700,000	Uganda
Adhola	235,000	Uganda
Alur	800,000	Congo (Kinshasa) and Uganda
Bari	340,000	Sudan and Uganda
Berta	60,000	Ethiopia and Sudan
Daju and Sila	150,000	Chad and Sudan
Datoga	400,000	Tanzania
DINKA	1,350,000	Sudan
Fur	500,000	Darfur District, Sudan
Gumuz	90,000	Ethiopia
Kakwa	148,000	Sudan and Congo (Kinshasa)
KALENJIN	1,400,000	Kenya
KANURI	4,000,000	Nigeria, Niger, Chad, Cameroun
Kuman	150,000	Uganda
Kunama	50,000	Eritrea
Kwama	15,000	Ethiopia
Lango of Uganda	800,000	Uganda
Lendu	490,000	Congo (Kinshasa)



First, the Nilotic or *Western Nilotic* languages include DINKA, Nuer, the LUO languages and others.

Secondly, a subdivision of languages of East Africa was once called *Nilo-Hamitic*. 'Nilo-Hamitic' or even *Half-Hamitic* seemed appropriate for speakers of Teso, Karamojong, Turkana, Masai and others – at a time when many African languages were explained through theories of racial mixture – because Cushitic (i.e. Hamitic or AFROASIATIC) components were seen in them and their culture. The term was in common use till the 1950s. In 1966 a new name, *Paranilotic* 'not quite Nilotic', was suggested for these same languages. *Eastern Nilotic* is now the preferred name. Together they form a continuous geographical sequence, with mainly Afroasiatic languages to their east and mainly Nilotic and Bantu languages to their west.

Eastern Nilotic (see map at TURKANA) is itself divided into three groups: *Bari*, mostly in Sudan, including Bari and Kakwa; *Teso-Turkana*, in Uganda and north-western Kenya, including TESO and the Karamojong group (one of which is Turkana); and *Lotuko-Maa*, including MASAI and Otuho.

Logo	210,000
Lugbara	920,000
LUO	3,200,000
MABA with Massalat and Mimi	300,000
Madi	233,000
Mangbetu	650,000
MASAI with Samburu	750,000
NUBIAN LANGUAGES	1,000,000
Nuer	850,000
Otuho	185,000
Pökoot or Suk	170,000
Sabaot	100,000
SARA including Mbai and Ngambai	700,000
Shilluk	175,000
SONGHAY	2,000,000
Tama and Sungor	165,000

Congo (Kinshasa)
Uganda
Kenya, Tanzania
Chad
Uganda
Congo (Kinshasa) and Uganda
Kenya and Tanzania
Sudan, Egypt
Sudan, Ethiopia
Sudan
Kenya and Uganda
Kenya
Chad
Sudan
Niger, Mali, Nigeria, Burkina Faso, Benin
Chad and Sudan, most speakers bilingual in Arabic

Teda	200,000	Chad and Libya
TESO	1,300,000	Uganda, Kenya
Thuri	154,000	Sudan
Tugen	150,000	Kenya
TURKANA and Karamojong	650,000	Kenya, Uganda, Sudan, Ethiopia
Zaghawa	123,000	Sudan and Chad



# NKORE

## 4,600,000 SPEAKERS OF 'RUTARA' LANGUAGES

### *Uganda, Tanzania*

**N**kore, one of the major languages of Uganda, belongs to the Rutara or Western Lacustrine group of BANTU LANGUAGES. The languages of this group – which also includes Kiga and Nyoro in Uganda, Haya and Zinza in Tanzania – are so close to one another that speakers of any one can easily understand the others, so they are dealt with together here.

Detailed genealogical histories are important to Nkore and Nyoro speakers, but their relation to movements of population and to linguistic changes is at best indirect. At any rate, these are the dialects of a group of kingdoms to the west and south of Buganda, in a region where Bantu languages have certainly been spoken for about two thousand years.

Their written history begins with J. H. Speke's travels through the region in 1862. At the establishment of the British protectorate, in 1896, buNyoro was a small but historic kingdom. BuToro had become a separate state only around 1830. Under British influence the four kingdoms ruled by the baHima caste (see box), Nkore, Buhweju, Buzimba and Igara, were united into a single district of Ankole, ruled from Nkore, in 1914.

*Nyankole* or *Nkore* is thus a 20th-century concept, the language of the new district of Ankole. However, the dialects of the four old kingdoms were almost identical.

The term *ruTara* is also a recent invention, named after the early kingdom of Kitara which, according to traditions widespread in the region, once held sway on the western shores of Lake Victoria from its capital at Bwera in Buganda.

In colonial times these kingdoms were given significantly fewer privileges than the favoured kingdom of Buganda. And GANDA's special position lasted into the early years of independence. Together the Rutara group forms the second largest linguistic community in Uganda, amounting to a fifth of the population. Yet by comparison with the slightly larger Ganda language group, Nkore and its relatives are very little seen or heard in press and media.

For a table of numerals see GANDA.

#### **Names with a meaning**

In Nkore-Kiga, personal names often take the form of nicknames based on a detail of personal history. For example:

*Ke-emerwa*, from the saying *akaizire keemerwa*  
'you have to take what comes';

*Nkagwe-guhira*, from  
*omuriro nkagweguhira gwanyotsya*  
'I got some fire and it burnt me',  
a saying hinting at an unhappy marriage.

After Charles Taylor, *Nkore-Kiga*  
(London: Croom Helm, 1985)

#### **Farmers and servants**

There are traditionally three castes among the speakers of Nkore, Tooro and Nyoro. The *ba-Hima* or *baHuma*, a traditionally cattle-keeping people, have higher status than the *baIru*. A third group of recent arrivals, the *baBito*, formed the

royal caste. There is no linguistic difference among the three.

There is a Nyoro myth that explains and validates their stratification of society into rulers or 'people of the drum', *baKama*, cattle-farmers, *baHuma*, and servants, *baIru*. They are said to be the descendants of the youngest, second and eldest sons respectively of the creator god, Ruhanga. BaNyoro go on to say that the baHuma have never ruled the country; in other words, that the three successive royal dynasties of bu-Nyoro, *baTemбуzi*, *baCwezi* and *baBito*, were none of them baHuma. The Nkore say, however, that the brief but powerful baCwezi dynasty, which ruled the great kingdom of Kitara, was baHuma in origin. Most people accept that the founders of the third dynasty, baBito, were speakers of a LUO language.

Unlike the Tutsi among speakers of RUNDI and Rwanda, the baHuma in buNyoro and bu-Toro now retain little of their traditional political and economic dominance. The distinction between the two castes is known to everyone – but no longer important in everyday life. The baHima of Ankole are somewhat more powerful; and here there is a middle caste, the *baMbari*, children of baHima men and their baIru concubines.

Based on Brian K. Taylor, *The Western Lacustrine Bantu* (London: International African Institute, 1962) and other sources

### Nkore and its relatives: the Rutara group

*Nkore* or Nyankole or Runyankore (1,500,000 speakers) is the language of the kingdom of Nkore (Ankole), now part of Uganda. Nkore and Kiga are now generally regarded in Uganda as a single language, *Nkore-Kiga*.

*Kiga* or ruKiga or Chiga or Ciga (1,000,000 speakers) and the closely related *Hororo* or ru-Hororo are spoken in what is now Kigezi district of Uganda. There was no kingdom and no castes here: traditional power was held by priests of the Nyabingi cult. Under British rule they were regarded as subversive and replaced by Ganda-speaking agents.

*Nyoro* or Runyoro (1,000,000 speakers) is the language of the kingdom of Bunyoro. It is close

to *Tooro* or Rutooro of the old kingdom of Toro. *Hema-Sud*, 'South Hema' of Congo (Kinshasa), is a continuation of the Tooro dialect continuum. It is so called to distinguish it from its neighbour *Hema-Nord* 'North Hema', a Nilo-Saharan language.

In Tanzania, *Haya* or luHaya or Kihaya (1,100,000 speakers) includes the dialects *Nyambo* (or Karagwe) and *Zinza*. This region was annexed by Germany in 1890 and first occupied by Britain in 1916.

*Kerebe* has 100,000 speakers on Ukerewe Island in the southern part of Lake Victoria, and in neighbouring Kibara. It is close to Haya linguistically.

#### BaHima cattle raids and their poetry

The praise poems of the baHima are an outstanding feature of Nkore literature. This is the opening of an 88-line poem describing an incident in 1949. Each line begins with a new praise-name of the hero or another participant.

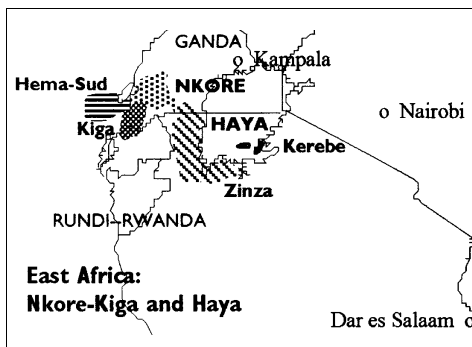
Rugumyána nkahiga!

rutashoróórwa bakándekura nkabanza;  
rutarimbíka ébikoomi nkabyetuuramu . . .

I who give courage to my companions  
made a vow!

I who am not rejected was sent out in advance;  
I who do not hesitate descended upon  
their camps . . .

H. F. Morris, *The heroic recitations of the Bahima of Ankole* (Oxford: Clarendon Press, 1964) pp. 66–7



# NORWEGIAN

**5,000,000 SPEAKERS**

## *Norway*

**A** descendant of OLD NORSE, Norwegian is the language of the long, narrow western strip of the Scandinavian peninsula. It is now the official language of Norway – but it has only recently acquired this status.

Old Norse was the language of the once-powerful medieval kingdom of Norway, with its successive capitals at Nidaros (Trondheim), Bergen and Oslo. Its heyday lasted from the 9th to the 14th century. The literature of this period, much of it poetry and historical prose inspired by the royal court, is in a language which cannot really be distinguished from early ICELANDIC: some of it, such as the royal ‘sagas’ or histories collected as *Heimskringla*, is by Icelanders.

Norway then became subject in turn to Sweden and to Denmark. The influence of Danish was strengthened with the Reformation, for no translation of the Bible into Norwegian appeared. There was no printing press in Norway until the 17th century, and no university until the foundation of the University of Christiania (Oslo) in 1811.

Through this period there was heavy influence, too, from German: High German, because of its cultural and educational pre-eminence, but particularly the Low German of the Hansa merchants, who dominated Norwegian trade from their establishments at Oslo, Bergen and Tønsberg.

Thus, for centuries, Norwegian was a series of local dialects of which none formed a standard. Danish was the government language, the literary language, and the language of higher education. The language spoken and written by educated Norwegians was in essence Dan-

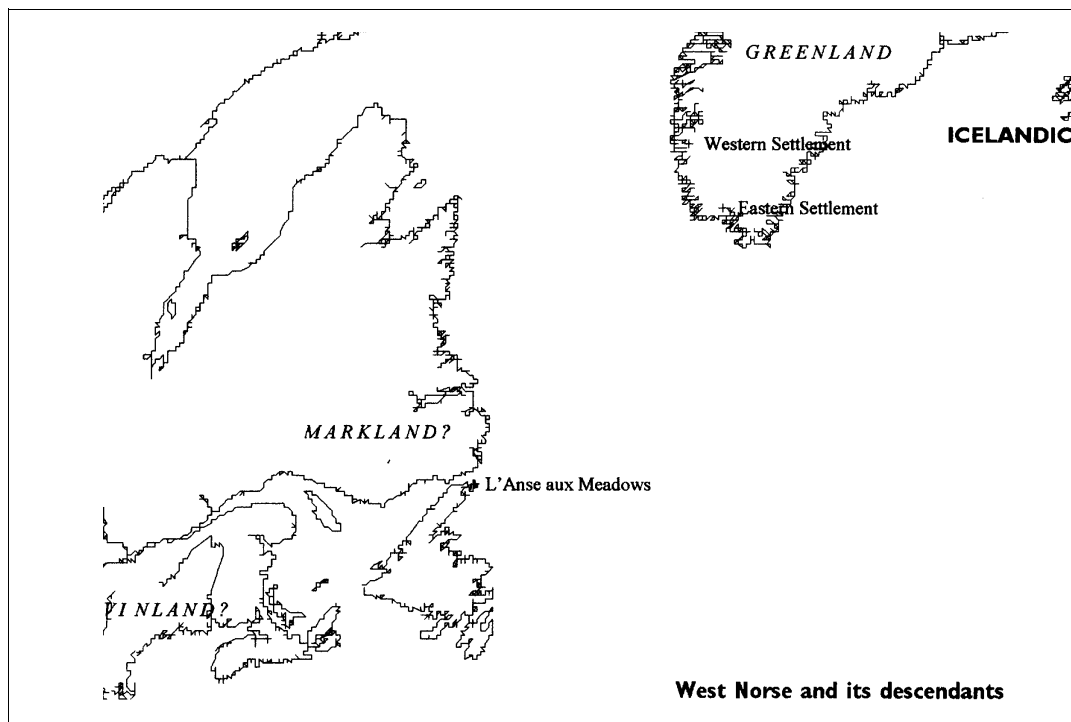
ish – though with a distinctive Norwegian accent.

Norway separated from Denmark in 1814. Education at first continued in Danish (re-named *Norsk* or *Modersmaal*) but in the 1830s a movement developed to create a national language.

Why? Because the Danish standard was too different from the everyday speech of any Norwegian and thus difficult for schoolchildren to learn; and because, so nationalists thought, every nation should have its own language.

How was this national language to be created? The poet Henrik Wergeland argued for the enriching of the existing standard with elements drawn from colloquial Norwegian and the various regional dialects. Essentially this was the route followed in the great folk tale collection by P. C. Asbjørnsen and Jørgen Moe, *Norske Folkeeventyr* (1842–4). Others, such as Ivar Aasen, wanted to build a new language on the basis of selected dialects, looking back to Old Norse when enrichment was needed and cutting out all possible foreign forms, especially those that had a Danish or German feel. Both points of view have had their supporters – from the 1830s to the present day.

So there are two standard languages in Norway. One, *Landsmål* ‘language of the country’ (the term was Aasen’s invention), is closely and very consciously based on the Norwegian dialects, particularly the country dialects of western Norway. Since 1929 it has been officially called *Nynorsk* ‘New Norwegian’. The other, *Riksmål* ‘national language’, now officially *Bokmål* ‘book language’, was developed, in sympathy with



West Norse and its descendants

Wergeland's ideas, by Knud Knudsen and others. It has a strong Danish flavour – but its resemblance is to written, not to spoken Danish – and a firmer base in the educated speech of 19th-century Oslo and Bergen. The 19th-century playwright Henrik Ibsen, the best-known figure of Norwegian literature internationally, eventually favoured Knudsen's Riksmål. So did the novelists Sigrid Undset and Knut Hamsun.

Schools can choose either, and many civil servants need to be able to handle both forms. Both are subject to occasional change by parliamentary commission, and (not surprisingly) a movement for *Samnorsk*, 'Union Norwegian', has gained some ground.

The pronunciation of Norwegian (of all varieties) seems closer to Swedish than to Danish: hence the most unfair cliché that Norwegian is 'Danish with a Swedish accent'.

*Russenorsk*, now extinct, was a mixed pidgin language used among Norwegian and Russian fishermen along the coast of the White Sea and northern Norway in the 19th century. Alongside

Russian and Norwegian words *Russenorsk* included a few each from French, English, Dutch, Swedish, Finnish and Lapp. Unusual among pidgins, the speakers were social equals: this no doubt explains why the two main languages contributed almost equally to the vocabulary of *Russenorsk*. In 1917 the Allied blockade of Russian ports ended the trade, and the pidgin soon disappeared.

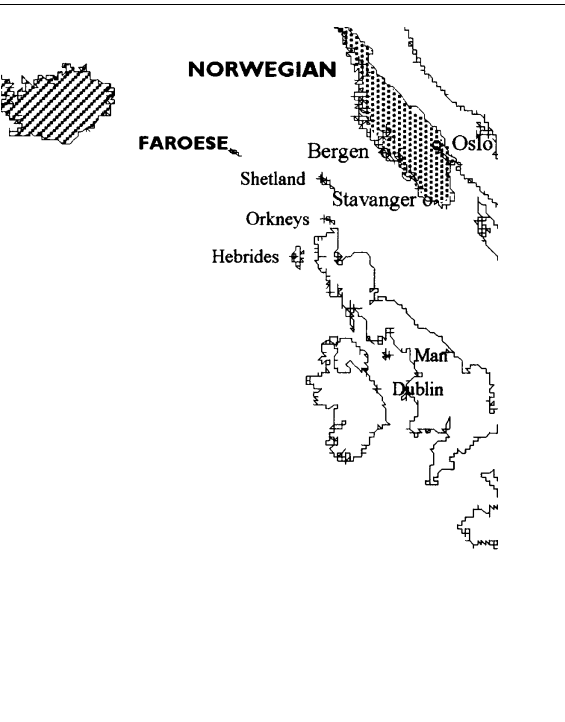
### Descendants of West Norse

The modern descendants of West Norse are FAROESE, ICELANDIC and NORWEGIAN.

The dialects of Norwegian can be grouped as West Norwegian (including Bergen and Stavanger) and East Norwegian, which stretches inland from Oslo. The Central and North Norwegian dialects of the northern coast are subdivisions of East Norwegian.

East Norwegian, like Swedish, often has pure vowels where West Norwegian has diphthongs, e.g. *sten* for *stein* 'stone'.

Crosses mark former centres of West Norse



speech where the language is now extinct. They include Man, Dublin, the Orkneys (*Orkneyjar*) and Shetland (*Hjaltland*). Greenlanders established a settlement briefly on the coast of New England (*Vinland*), but abandoned it after three years.

As a result of migration in the 19th century there are more than half a million speakers of Norwegian in the United States.

# NUBIAN LANGUAGES

**PERHAPS 1,000,000 SPEAKERS**

*Sudan, Egypt*

**N**ubian is a close-knit group of NILO-SAHARAN LANGUAGES, and has a much longer recorded history than any other member of the family.

From 850 BC until AD 300 there was an independent kingdom, south of Egypt, in the middle Nile valley. It was known to the Egyptians, and to its own rulers, as *Kush*; its name in Greek and Latin was *Meroe*. Its stone inscriptions were at first in EGYPTIAN; later ones are in another language, known to scholars as Meroitic, which seems to be unrelated to any of its neighbours. The writing system was deciphered by F. Ll. Griffith – there are twenty-three alphabetic symbols, borrowed from Egyptian scripts – but no one has yet made sense of the language.

With the coming of Christianity to the middle Nile, in the 6th century, the kingdoms of Nubatia (or Meris) and Makouria emerge into history. Their records are in ‘Nubian’, an early form of modern *Nobiin*. So what happened to Meroitic? When did Nubian begin to be spoken in the Nile valley, and how did this Nilo-Saharan language first reach the middle Nile? The answers are far from clear – but it seems most likely that the early homeland of Nubian is some way to the south-west of the modern Nile Nubians: in Darfur, where one minority Nubian language, *Meidob*, is still to be heard, or in Kordofan, where there are more. From here, perhaps not long before the time of Christ, they spread to the Nile valley itself and gradually displaced Meroitic.

People called *Noubai*, *Nubae*, *Noba*, *Annoubades* and *Nobatai* are named in Greek and Latin sources of the early centuries AD. South of them, in modern southern Sudan, were the *Makoritai*. It seems likely that two modern Nubian languages represent these early medieval peoples. The northern is *Nobiin*: its speakers are called *Mahas* in Sudan, *Fadija* in Egypt. The southern is *Dongolawi* or ‘Dongolese Nubian’, language of the *Danagla* and their recent northern offshoot the *Kenuz*.

The Christian kingdoms of the Nubians disappeared six centuries ago. Modern Nubian speakers are Muslims, and they are bilingual in Arabic. The traditional habitat of speakers of the Nile Nubian languages has largely disappeared under the waters of the High Dam at Aswan, and they have been resettled – a move which is likely to speed the disappearance of their historic languages.

Nile Nubian languages have five vowels and a distinction of vowel length, which may be used in forming derivatives: *baag* ‘divide’, *bag* ‘distribute’; *aab* ‘catch in a net’, *ab* ‘catch in the hand’.

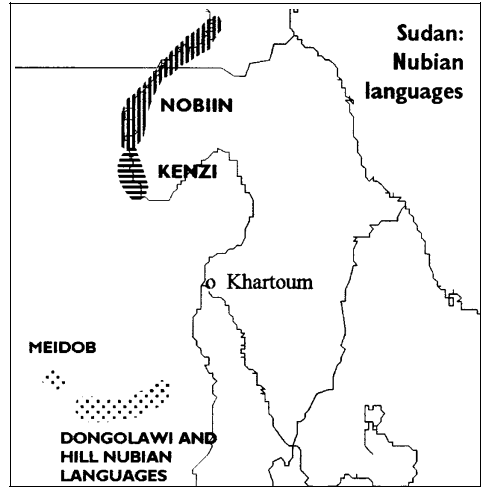
The first ten numerals in Dongolawi Nubian are: *wer*, *oww*, *tosk*, *kemis*, *digh*, *gorigh*, *kolod*, *idiw*, *iskod*, *dimin*.

Based on papers by William Y. Adams and Robin Thelwall in *The archaeological and linguistic reconstruction of African prehistory* ed. C. Ehret, M. Posnansky (Berkeley: University of California Press, 1982) and other sources

## Hill Nubian and Nile Nubian

The Nile Nubian languages are *Dongolawi* (with its offshoot *Kenzi*) and *Nobiin*. On the apparent origin of the Kenzi dialect see BEJA.

The Hill Nubian languages, including *Ghulfan* and *Meidob*, are spoken by small groups in southern Kordofan and Darfur.



# NUPE

1,000,000 SPEAKERS

*Nigeria*

One of the NIGER-CONGO LANGUAGES of the Benue-Congo group, Nupe is spoken along the Niger and Kaduna rivers in the Emirates of Bida, Agaie, Lapai and Patigi-Lafiagi.

Nupe speakers call themselves *Nupei* ʒi, their language *ezi Nupe* and their country *kin Nupe* 'land of Nupe'. In Hausa they are called *ba Nupe* or *Nufawa*, in Yoruba *Takpa* or *Tapa*. As *Tapa*, Nupe-speaking slaves were to be found in Brazil and Cuba in the 19th century.

Nupe tradition tells of the origin of their ethnic identity in a migration from what is now Igala country to their present location north of the Niger, led by the hero Tsoede or Egedi, five hundred years ago. He is said to have united the indigenous people, till then disorganised. The Nupe kingdom can be found on 18th-century European maps as *Noofy*. It came under Hausa and Fulani domination in the 19th century: after 1870 their ascendancy was followed by that of the Niger Company and the British Government of Nigeria.

Long used in the Niger river trade, Nupe is nowadays a language of radio and television. However, numerous speakers are bilingual in Hausa, the lingua franca of the area and the national language of Northern Nigeria. Yoruba is also known by many. When Christian missionaries began work in Nupe, in 1858, they approached the language by way of Hausa and Yoruba-speaking polyglot interpreters. The majority of Nupe speakers are Muslims: some therefore study Arabic, or are at least able to recite from the Koran.

Nupe is a richly tonal language, with three

level tones, a rising tone and a falling tone: *bá* 'be sour', *ba* 'cut', *bà* 'pray', *bǎ* 'not', *bâ* 'slander'. The first ten numerals are: *niní*, *gúbè*, *gútá*, *gúnni*, *gútsun*, *gútswànyì*, *gútàbà*, *gútótá*, *gútwani*, *gúwo*. '20' is *esi*. Counting in Nupe is by fives and twenties: but for '30', '50' and '70' Yoruba loan-words have been inserted into the system.

*Ndǎ mi de fitila niní, a gá lá 'na da u yìzè kpátá le  
bà yé* – 'My father has only one lamp, but  
when he lights it the whole world is lit up!  
The answer? The moon.

## Early Nupe literature

'For a long time it was thought that the Nupes possessed no literature of their own, until the discovery was made a few years ago that there were in existence in the Nupe country a number of songs written in the Aljemi character, and dating back about one hundred years. This bastard Arabic character called Aljemi is in general use in parts of north Africa, and all over the western Sudan; and although it is not at all suited to the Nupe language, still it has been used as the medium of circulating poems and songs intended to be committed to memory, and embodying . . . Mohammedanism . . . with curious allusions and statements which are certainly not to be found in the Koran.'

A. W. Banfield, J. M. Macintyre,  
*A grammar of the Nupe language* (London: SPCK, 1915) p. 122

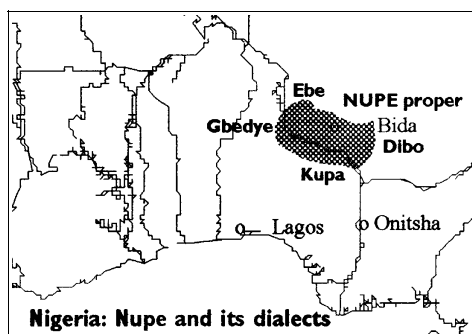
This is a couplet from a Muslim song from one of these manuscripts:



Abà Jiyà gǎ, Ye kpe ze ètá Nupe ci èjĩ yèbó?  
 Abũ Bākārì egì Anàsì ci èkóní, ci èjĩ yèbo.  
 Aba Jiya says, Do you know those who  
 speak Nupe and are giving thanks?  
 Abu Bakari son of Anasi is singing,  
 he is giving thanks.

## Nupe and its dialects: the map

The major dialects of Nupe are *Nupe* proper, *Ebe*, *Dibo* or *Zitako*, *Kupa* and *Gbedye*. The river people, *Kyedye*, with colonies as far south as Onitsha, speak Nupe proper. The Nupe spoken in Bida, ‘higher



*Nupe* with its Hausa loanwords, is recognised as the prestige form of the language; it, or at any rate Nupe proper, is known to speakers of other dialects.

# NYANJA

**4,000,000 OR MORE SPEAKERS**

*Malawi, Mozambique, Zambia, Zimbabwe*

**U**nder its various names, Nyanja is the language of central and southern Malawi and of surrounding districts of Zambia, Mozambique and Tanzania. One of the BANTU LANGUAGES, Nyanja remained the language of local administration when, under the name Nyasaland, the heart of this country became a British protectorate.

*Nyanja, chiNyanja, ciNyanja, Nyasa and Niassa* are all variants of the usual name for the language and its speakers, meaning 'Lake' or 'Lake people'. *Cewa* or *chiChewa* is the western Nyanja dialect. *Cewa* and *Chichewa* are sometimes used, in Zambia and Malawi, as names for the language as a whole – or for the true regional language, as opposed to the lingua franca of British administration and of Zambian cities.

The speakers of Nyanja and neighbouring languages were grouped by 19th-century explorers, including David Livingstone, as *Maravi* or *Malawi*; this was revived as the name of the newly independent state.

Nyanja-speaking people have a tradition of origin in the Zaire basin, several hundred years ago. What is certain is that the language has been spoken along the southern shores of Lake Nyasa and the Shire river for some centuries; that it was once the language of an extensive political federation; and that in the last 150 years several successive population movements have redrawn some details of the ethnic map.

At the end of their long march from Zululand several groups of the Ngoni settled in Nyanja-speaking country in the 1830s and 1840s and ruled it as conquerors. Their language is no

longer spoken here, but people of some Nyanja-speaking areas regard themselves today as Ngoni rather than Nyanja.

In southern Malawi and neighbouring districts of Mozambique, Nyanja speakers are intermixed in a complex pattern with speakers of Yao, Makua, Lomwe and Chwabo. This is partly the result of recent migrations (in particular, Yao incursions in the late 19th century), but the evidence suggests that such movements have been going on for rather longer than that.

To the north-west, the extensive territory where the Cewa dialect of Nyanja is now spoken results from a migration that began about two hundred years ago.

## Language of the police

Missionaries and tea and tobacco planters came to live in the Nyanja-speaking region in the late 19th and early 20th centuries. From working with them, many Nyanja had become useful white-collar workers just at the time when British domination was extending over what is now Zambia and Zimbabwe.

Nyanja, already the language of a large and relatively well-educated population, spread over great distances. It was used in local government offices, in the police and the army, all over the British Central African territories that were briefly united as the Federation of Rhodesia and Nyasaland. Nyanja became one of the four languages of the Zambian Copperbelt (see BEMBA), and it remains the lingua franca of the Zambian capital, Lusaka. But the European tide has receded, and in the long term, as with Fanakalo (see ZULU), the police and government

overtones of Nyanja can be seen to have lessened its popularity outside its homeland.

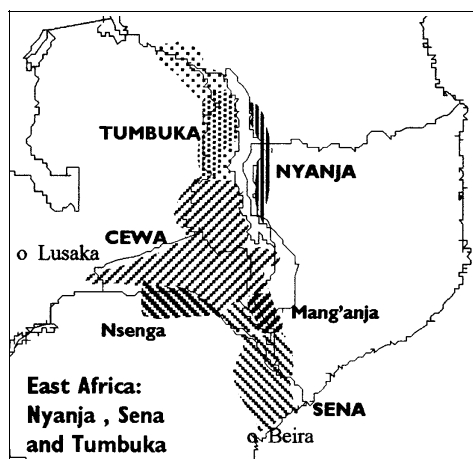
Rival missionaries set up three literary standards for Nyanja. The Blantyre Mission was established on the Shire plateau in 1876 and published in Mang'anja; the Dutch Reformed Church mission worked in Cewa, while the Universities Mission to Central Africa favoured the Lake Nyanja dialect. The artificial 'Union Nyanja' of the Bible translation published in 1922 was not a success.

Cewa or Nyanja is the major African language of Malawi and the second most widely known of the eight official languages of Zambia, spoken by fewer people than Bemba but by more people than English. The first ten numerals are: *-modzi*, *-wiri*, *-tatu*, *-nayi*, *-sanu*, *-sanu ndi -modzi*, *-sanu ndi -wiri*, *-sanu ndi -tatu*, *-sanu ndi -nayi*, *khumi*.

## Nyanja, Sena and Tumbuka

The dialects of Nyanja are *Cewa* or Chewa or Ancheya in the west, *Peta* in the south-west, *Nyanja* near Blantyre, *Mang'anja* in the lower Shire valley and *Nyasa* on the eastern shore of Lake Nyasa.

To the south and west, the closely related *SENA* is the language spoken along the Zambezi in Mozambique. There are 1,200,000 speakers. *Nyungwe* or Tete, a northern variant of Sena, has



250,000 speakers. *Kunda* or *ciKunda* (100,000 speakers) is the language of a trading people, metalworkers who also dealt in slaves and elephants, with old-established settlements along the Zambezi–Luangwa trade route. Their name may be Nyanja in origin: *mcikunda* 'fighter'. Their language is close to Sena and Nyanja, but their ancestry is said to be varied, with links to many of the peoples of central Mozambique.

Outside the group usually regarded as Maravi or Malawi, TUMBUKA is the main member of a group of small dialects spoken in northern Malawi and eastern Zambia.

# OCCITAN

## SEVERAL MILLION SPEAKERS

### *France*

**P**rovençal, now more often called Occitan, was in medieval times one of the major ROMANCE LANGUAGES, known and cultivated over much of western Europe as the vehicle of beautiful lyric poetry. The survival of Occitan is now threatened by the inexorable spread of French.

The term *Provençal* is often used of the medieval literary language, one of whose main centres was the court of Provence. It is not a suitable term for the modern language, which is actually spoken in a much wider area of France than the region of Provence (whose name comes from Latin *provincia*, 'the province', a by-name dating from the period before Julius Caesar's conquests, when this was still the only part of Gaul that was under Roman rule).

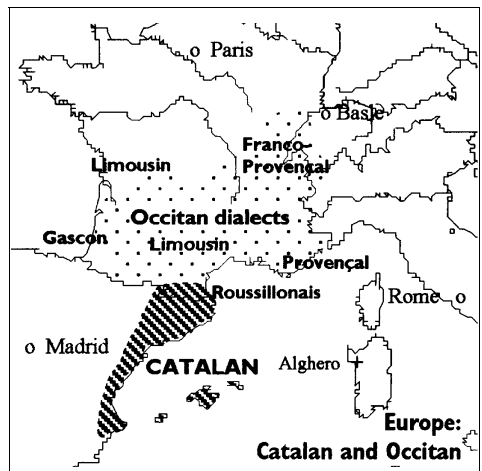
Traditionally the medieval languages of France were distinguished by the word for 'yes': the northern language, French, was called the *langue d'oïl* (but in modern French 'yes' is now *oui*) and the southern was the *langue d'oc*. From this derive the old regional name *Languedoc* and also the modern name for the language and its territory, *Occitan*, *Occitanie*.

Recorded from the 10th century onwards, as learned Latin began to give way to the local spoken tongue in documents and poetry, Provençal is best known as the language of the troubadours. Their poetry of love, satire and war was performed in royal and noble courts in France, in Spain and across Europe. In that multilingual environment Provençal was, from the 12th to the 14th century, the language of lyric: kings were not ashamed to compose and

sing in it. Provençal poetry was later imitated in Sicilian and other Italian dialects, in Galician-Portuguese and in German. Some trace the origin of Provençal lyric song to the influence of Arabic culture, spread across Spain and into southern France by the Islamic conquests.

As France became a unified nation state, Provençal rapidly declined in importance after the 15th century. The prophet of the language's revival was Frédéric Mistral (1830–1914), whose long poem *Mireiò* was much admired. Mistral gained the Nobel Prize for Literature in 1904. But no one literary standard was agreed on, and the writing and reading of Occitan literature remains a minority interest in southern France. In schools and in all official and public contexts French is required: fewer and fewer children in each generation have Occitan as their mother tongue.

Like medieval French, Provençal retained two of the six Latin noun cases: for example,



nominative *iois* ('joy', subject of a clause) and accusative *ioi* ('joy', direct or indirect object) both appear in lines 1–2 of the poem by the Countess of Die (see box). In other ways, medieval Provençal and modern Occitan are closer to Catalan than to French.

The first ten numerals in Occitan are: *un, dous, tres, quatre, cinq, sièis, sèt, vne, nòu, dèu*. *Shuadit*, the language of Provençal Jews, was a special form of Occitan with borrowings from Hebrew and other Jewish languages. It has been extinct since 1977.

### Playing with words

This 12th-century poem shows how word forms could be neatly varied to give the rhyme and parallelism that was at the centre of Occitan poetic style. It is by one of the earliest Occitan poets, the 'Countess of Die'. Scholars have puzzled over her identity, but she remains mysterious.

Ab ioi et ab ioven m'apais  
 e iois e iovens m'apaia,  
 que mos amics es lo plus gais  
 per qu'eu sui coindeta e gaia,  
 e pois eu li sui veraia  
 be.is taing q'el me sia verais,  
 c'anc de lui amar no m'estrais  
 ni ai cor que m'en estraiä.

I pleasure myself with joy and youthfulness,  
 And joy and youthfulness pleasure me,  
 For my lover is the gayest  
 Through whom I am happy and gay,  
 And since I am true to him  
 It is well that he is true to me,  
 And I never escape from loving him  
 And I have not the heart to escape this.

In lines 3–4 and 5–6 the final adjectives alternate between masculine (*gais, verais*) and feminine (*gaia, veraia*). Lines 1–2 and 7–8, using the same rhymes, end with verbs that alternate between 1st person (*-ais*) and 3rd person (*-aiä*).

Text from *Songs of the women troubadours* ed. Matilda Tomaryn Bruckner, Laurie Shepard, Sarah White (New York: Garland, 1995) p. 2

# OJIBWA

80,000 SPEAKERS

*Canada, United States*

One of the Algonquian family of Amerind languages, Ojibwa was once spoken north of Lake Huron and on both sides of Lake Superior, extending as far as what is now North Dakota.

Alternative names are *Chippewa*, *Chipeṡṡyan*.

Ojibwa speakers were traditionally semi-nomadic hunter-gatherers, relying on wild rice gathering, on migratory hunting in winter, and on fishing at settled sites in summer. As with the Cree, their territory was at its greatest extent in the 18th century, when they took land formerly claimed by the Sioux and Fox peoples. They were firm allies of the French through this period. Meanwhile the fur trade with Europeans in the 18th and 19th centuries was destabilising Ojibwa culture, enriching the chieftains and turning their office into a hereditary one.

Ojibwa is still spoken from Ontario westwards to North Dakota in scattered Indian reservations (see map at ALGONQUIAN LANGUAGES).

The language of the Ottawa or Odawa, after whom Canada’s national capital is named, is sometimes called *Eastern Ojibwa*. In historical tradition

Ojibwa, Ottawa and Potawatomi, speakers of three related Algonquian languages, together made up an alliance known as the ‘three fires’.

The first ten numerals in Ojibwa are: *pēshik-wan*, *nīshan*, *nissan*, *nīwan*, *nānan*, *ninkotwāssan*, *nīshwāssan*, *nīshwāssan*, *shānkassan*, *mintāssan*.

It was on the myths of the Ojibwa – collected in the *Algic researches* of Henry R. Schoolcraft (1839) – that Longfellow based his epic *Hiawatha*.

## Writing Ojibwa

In writing Ojibwa, a syllabary (based on the one used for CREE) has competed with the Roman alphabet. Its inventor, the missionary James Evans, had worked first with Ojibwa, for which he devised a Latin transliteration. By the time he was transferred to Cree territory he had concluded that a newly designed syllabary would work better. Invented in 1840, the Cree syllabary was soon adopted in Ojibwa writing and printing. It is partly inspired by English short-hand.

# OLD NORSE

## EARLY LANGUAGE OF SCANDINAVIA

**O**ld Norse represents one of the three early branches of the GERMANIC LANGUAGES. It is the oldest recorded form of what are now the five modern Scandinavian languages.

The earliest period of Old Norse, 'proto-Nordic', extends from about AD 300 to 800. The only records of it at this period are Runic inscriptions (see box) along with some Scandinavian words and names in Latin and Anglo-Saxon texts. There is as yet no sign of dialect differences.

Between 800 and 1050 dialects begin to emerge. The principal division is between East and West Norse: the great mountain chain of Scandinavia, the 'Keel', *Kjølen* in Norwegian, *Kölen* in Swedish, marks the dividing line.

West Norse was spoken in Norway and a growing area of Viking settlement in the north Atlantic. About thirty Runic inscriptions survive on the Isle of Man as a reminder that Old Norse, not Manx, was that island's first written language. More Norse inscriptions are to be found in Ireland and in Orkney. But the most important of the Viking colonies was certainly Iceland. West Norse is the immediate ancestor of modern NORWEGIAN (see map there), FAROESE and ICELANDIC.

East Norse was spoken in Denmark, Sweden, coastal Finland and Estonia, and the eastern territories once ruled by Vikings. The Danelaw, the Norse conquests in eastern England, were occupied and settled from Denmark more than from Norway. Had a Norse language survived in East Anglia, Lincolnshire and Yorkshire, it would have been East Norse! As it is, numerous Norse words have been borrowed into English, including *skip*, *skirt*, *egg*, *steak* (Norwegian *steke* 'to roast'). More are to be found in north-eastern English dialects: *laik* 'to play'. Viking trade and conquest in western Russia has left plenty of linguistic evidence, including Christian names

such as *Igor* and *Olga* and the name of *Russia* (*Rus*) itself. East Norse is the ancestor of modern SWEDISH and DANISH (see map there).

The earliest name for what we now call Old Norse was *Dānsk tunga*, 'Danish tongue'. *Norrønt mál* was the term for what is essentially the same language as spoken in Norway and Iceland, 'West Norse'.

The best-known collection of Old Norse poetry is the *Edda*, sometimes called 'Elder' or 'Poetic Edda' to distinguish it from a later prose work on Norse mythology. The *Edda* poems reflect pagan beliefs. Old Norse literature shades imperceptibly into Old Icelandic and Old Norwegian.

The Scandinavian languages as a group have several features that mark them off from other Germanic languages. The standard languages have two genders, common and neuter (though Icelandic and Faroese still have three). In the standard languages and nearly all the dialects the definite article 'the' is suffixed to the noun. Faroese and Icelandic differ from the rest, appearing in many ways more archaic: this is partly the result of the very heavy Low German and High German influence on the 'continental' Scandinavian languages, an influence which the island languages escaped.

'Each of the Scandinavian languages has its own "flavour". Swedish can be very formal and "correct", but also has a surprising amount of grace, and has now shed most of the syntactical heaviness it used to share with German. Norwegian is rugged and salty, and infinitely varied, while Danish has an amazing lightness and a pronunciation which positively lends itself to comedy' (M. O'C. Walshe, *Introduction to the Scandinavian languages* (London: André Deutsch, 1965) p. 13).





# OLD SLAVONIC

## Classical language of Eastern Europe

In the 9th century Christian missionaries from the Byzantine Empire began the large-scale conversion of speakers of SLAVONIC LANGUAGES – which, at that date, had not yet differentiated very widely from the ancestral form of speech that linguists reconstruct as ‘proto-Slavonic’.

The court to which the missionaries were first invited, in 862, was that of Prince Rastislav of Great Moravia. The local tongue must have been an early form of Czech. But the first missionaries, Saints Cyril and Methodius, had learnt their Slavonic speech in Macedonia, near Saloniki. Thus the language of the early translations – Bibles and religious manuals – in what we now know as Old Slavonic is closer to modern Macedonian and Bulgarian than to the other Slavonic languages.

*Old (Church) Slavonic* or *Old Slavic* is sometimes called *Old Bulgarian* to emphasise its geographical origin (the Macedonian language being regarded as a form of Bulgarian). The name *Church Slavonic* is often reserved for the later ‘Russianised’ form of the language.

Old Slavonic is essentially the language of the 9th- and 10th-century texts. A developed form of Old Slavonic, with some features of spoken Bulgarian, was written in the monasteries of Bulgaria and Macedonia, and notably at the Bulgarian capital of Trnovo and the great monastic centre at Ohrid. Old Slavonic also spread, with Christianity, in Serbia, in the Romanian principalities, in Ukraine and eventually in Russia. But as a result of lengthy warfare ending only with the Turkish conquests and the revival of Russia, the Bulgarian, Serbian, Romanian and Ukrainian-Russian varieties of Old Slavonic tended to develop in isolation, each with an increasing number of local features and with

very different pronunciations.

Old Slavonic reigned in Bulgaria and Serbia until the 14th century, when a linguistic reform and literary revival led by the Bulgarian Patriarch Evtimii was closely followed by the Turkish victories, notably the Battle of Kosovo, celebrated in Serbian epic, in 1389 and the fall of Trnovo in 1393. Ukrainian Church Slavonic was for four centuries the principal administrative and literary language of Ukraine, but it lost this role with the partition of Ukraine in 1387.

Evtimii’s linguistic reform, then, came as Old Slavonic was about to lose its dominant role in Bulgaria, Serbia and the Ukraine. However, as an official and literary language it long remained in regular use in the Romanian principalities of Wallachia and Moldavia: in modern Romanian a great deal of everyday vocabulary is still of Old Slavonic origin. And the linguistic reform had its fullest effect in Russia, where the Bulgarian scholar Kiprian was appointed Metropolitan of All Russia in 1389. Church Slavonic in its 14th-century Russian rebirth was by now a very different language from the spoken language that surrounded it.

After an abortive revolt at Pec in 1690 Serbian refugees settled in Hungary and were succoured by Russian missionaries. From this time onwards the Russian variety or ‘Church Slavonic’ was adopted in Serbian churches, while a Serbian-Church Slavonic mixture, *Slavenoserbian*, became the language of education. Russianised Church Slavonic spread to Bulgaria in the 18th century, and it is still to be heard in Orthodox church services in Russia and eastern Europe. Old Slavonic and Church Slavonic loanwords can be identified in all the modern Slavonic languages.

Old Slavonic has three numbers, singular, dual and plural. Verbs have two series of forms for

perfective and imperfective aspect. The first ten numerals are: *jedinə*, *dəva*, *tr'ye*, *chetire*, *pēt'*, *shest'*, *sedm'*, *osm'*, *devēt'*, *desēt'*.

## The alphabets of Old Slavonic

When a previously unwritten language was to be used for religious conversion, it was quite normal to invent a wholly new alphabet for it. The Gothic, Armenian and Georgian alphabets, among others, originated in this way. Thus St Cyril, one of the two missionaries to the Slavs of Great Moravia, invented the *Glagolica* – the Glagolitic alphabet, see the middle rows of the box – to be used for texts in Slavonic.

Great Moravia soon collapsed and the focus of the mission shifted south, to Bulgaria. Here, probably, it seemed best to use a writing system that was closer to Greek and easier for Greek-educated missionaries to learn. This may have been the origin of the ‘Cyrillic’ script (see the middle row). This is now used in a modernised form for many of the Slavonic languages. The earliest Cyrillic inscriptions are found in Bulgaria and Romania, and date from the 10th century.

This view of the origin of the Cyrillic alphabet is controversial – but it does explain why Cyrillic letter shapes are like Greek for sounds that exist in Greek, and are like Glagolitic for sounds that are unknown in Greek. Others believe that Cyrillic was in existence before the time of

the mission as a way of writing Slavonic names in Greek documents: but there is no real evidence for this. The traditional story is that St Cyril invented both Glagolitic and Cyrillic (and the latter is, after all, named after him) but it is difficult to see why he would have invented two alphabets to serve essentially the same purpose.

In Catholic Croatia the Glagolitic alphabet continued to be used until the late 19th century. Elsewhere in Eastern Europe it was soon forgotten.

Two manuscripts of the 11th century are thought to preserve Old Slavonic in its purest form. Known as Codex Zographensis and Codex Marianus, they contain Gospel translations from the Greek. They are in Glagolitic script: in later texts the Cyrillic alphabet was adopted.

## The Cyrillic and Glagolitic alphabets in Church Slavonic

А Б В Г Д Е Ж З И Й К Л М Н О П Р С Т У Ф Х Ц Ч Ш Щ Ъ Ы Ь Э Ю Я а б в г д е ж з и й к л м н о п р с т у ф х ц ч ш щ ъ ы ь э ю я

ተረፈው በጋራ ጥያቄው ላይ ለሚገኙት ሌሎችም ማስተላለፍ ይገባል፡፡

a b v g d e z h d z z i i g y k l m n o p r s t u f k h o t s  
ch sh sht ə 1 ʲ yɛ yu ya ye ě õ yě yõ x ps f i

# OMOTIC LANGUAGES

**T**he Omotic group are AFROASIATIC LANGUAGES spoken in Ethiopia. They differ strongly from the Semitic languages like Amharic and the Cushitic languages such as Somali and Oromo. The researches of Harold Fleming and Lionel Bender in the 1970s first demonstrated that Omotic was a separate division of the Afroasiatic family. It includes some of the languages once grouped under the name *Sidama*.

*Omotic* is named after the River Omo, along whose tributaries most speakers of Omotic languages live. *Sidama* is an OROMO word: whatever its origin, it came to mean 'foreign, not Oromo'.

Most Omotic languages are little known and spoken by small communities. The two important groups are Gonga, which includes the historic Kefa, and Ometo, whose major representative is WOLAYTTA. For a map see GONGA.

## Janjero and its royal language

*Janjero* is the language of the little-known former kingdom of Janjero, north-east of Kefa. It is first mentioned in Ethiopian records in 1427

and was conquered by the Emperor Menelek in 1894. Its remarkable customs, including human sacrifice, were described by the two Europeans who visited it – in 1613 and Antonio Cecchi in the 1870s.

'The people call themselves *Yamma* or *Yem*. The name by which they are better known is the [Oromo] form of the Amharic *zenjaro*, "ba-boon" . . . (the Italian spelling is *Giangero*).

'The royal language of Janjero consisted of a special vocabulary for parts of the body, weapons, and verbs of action referring to the king. Thus "eye" in common Janjero is *afa*, but *kema* in the royal language; "eat" is *ma* in common speech, *bos* in the royal language; and "spear", *ebo* in common speech, is *me'a* in the royal language. The language of respect used special words to describe the ordinary actions of notables: "eat" is *ma* in common speech, but *ta* in the language of respect. Improper use of the royal language was punished by death.'

G. W. B. Huntingford, *The Galla of Ethiopia*.

*The kingdoms of Kafa and Janjero* (London: International African Institute, 1955) p. 137

# ORIYA

22,000,000 SPEAKERS

*India*

**O**riya is the language of the Indian state of Orissa. It is one of the Eastern INDO-ARYAN LANGUAGES, closely related to neighbouring Bengali.

There are a few inscriptions in Oriya from dates as early as the 10th century, though at this date the language is not really distinguishable from Bengali. Longer inscriptions appear in the 13th and literary texts in the 15th. Rule of Orissa by the TELUGU-speaking Telingas and the MARATHI-speaking Nagpur dynasty of Bhonsla were followed by Bengali cultural and political dominance in the 19th and early 20th centuries, for the British paid little direct attention to Orissa and its culture (it became a separate state within British India in 1936). All three of these episodes have left their mark on the language.

In a world context Oriya is one of the least known of major Indian languages. Its poetic literature is extensive but little studied outside the borders of the state. Older texts are almost all inspired by Hindu mythology: there are more than a dozen reworkings of the Sanskrit *Rāmāyana* in Oriya. One of them is due to the founder of modern Oriya literature, Fakirmohan Senapati (1843–1918), who also wrote novels notable for their evocation of real everyday Oriya speech patterns.

Standard Oriya shows relatively little influence from the Austroasiatic languages, such as SANTALI, which are spoken in western Orissa.

The inland dialect Sambalpuri (for map see ASSAMESE) is much more heavily influenced by them.

For a table of numerals see BENGALI.

Orissa is best known outside India for the great temple of *Jagannātha*, a name of Krishna, at Puri. Thanks to the processions in his honour, this god's name is the origin of the English word *juggernaut*.

## Oriya script

The alphabet is historically related to Bengali, but it looks wholly different – because whereas Bengali script is suited to writing with a pen, Oriya is perfectly adapted to writing with a stylus on palm leaves, the traditional material for manuscript texts. There must be no long horizontal strokes – like the ‘washing-line’ on which Devanagari and Bengali scripts appear to depend – because these would split the leaf. Oriya substitutes the half-circles built into almost every character.

### The Oriya consonants

କ ଖ ଗ ଘ ଙ ଚ ଛ ଜ ଝ ଞ ଟ ଠ ଡ ଢ ଣ ତ ଥ ଦ ଢ ନ ପ ଫ

k kh g gh ṅ c ch j jh ñ ṭ ṭh ḍ ḍh ṇ t th d dh n p ph  
b bh m y r l ṣ ś s h

# OROMO

7,500,000 SPEAKERS

*Ethiopia, Kenya, Somalia*

**O**ne of the CUSHITIC LANGUAGES of the Afro-asiatic family, Galla is the major regional language of Ethiopia, though the national language is AMHARIC.

The name *Oromo*, now standard among linguists and official in Ethiopia in the form *Orominya*, is what the speakers call their language and themselves (the plural is *Oromota*). In local genealogy, it was Oromo son of Omer who long ago crossed the sea from Arabia and became the founder of the nation. *Galla* is the term found in most older sources. It was used only by Amharic and other foreign speakers.

The historical nucleus of Oromo speech is at the eastern end of its present territory, around and to the south of Harar. It is clear from Ethiopic and foreign records that Oromo speakers conquered much of central and western Ethiopia in the 16th and 17th centuries, including large areas of former GONGA speech. Oromo spread partly through these conquests, partly through slavery, for Oromo societies both kept slaves and sold them to neighbouring peoples. Slavery survived in eastern Ethiopia until the late 1930s.

Until the 19th century Oromo speakers, pastoralists and traders, dominated what is now south-eastern Ethiopia and the hinterland of eastern Kenya: they formed the essential link between the Swahili and Somali speakers of the coast (whom they called *Hamara* or *Abba shuffa*, 'people with clothes') and inland populations such as the KAMBA. 'Galla' is still a lingua franca in eastern Kenya, e.g. for speakers of Pokomo, but more now use Swahili.

Wello, Eastern and Southern Oromo speakers are largely Muslim. In these areas men often know some Arabic, but only those who have studied the religion are able to read and write it well.

Oromo is a language with a rich oral literature, including poetry, songs and historical traditions. Henry Salt's *Voyage to Abyssinia* (1814) contained a word list of Oromo, but the first published work by a native Oromo speaker was *The Galla spelling book* by Onesimus Nesib (see box). Until the 1970s Oromo was not favoured as a written language in Ethiopia. It was scarcely to be heard on the radio and seldom seen in print. The Amharic alphabet is normally used for written and printed Oromo.

Oromo has six vowels, with a long-short distinction, and high and low tones. The first ten numerals are: *tokko*, *lama*, *sadii*, *afur*, *shan*, *ja'a*, *torba*, *saddeet*, *sagal*, *kudan*.

## ***The Galla spelling book***

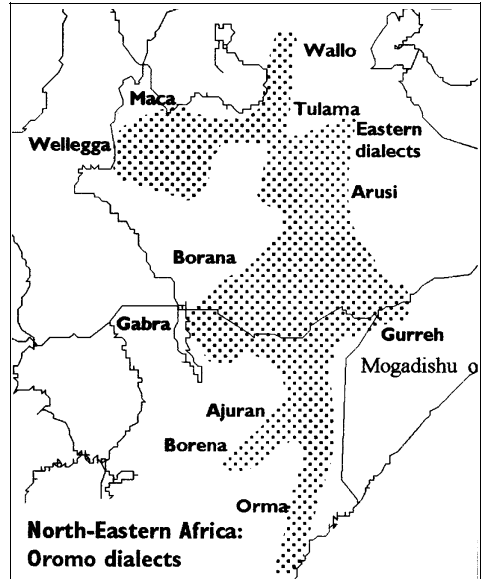
'Of written literature perhaps the most interesting is *The Galla spelling book*, by Onesimos Nsib, a native Galla. Printed at the Swedish Mission Press, Moncullo near Massowah, 1894. This is the English title; it has also a Galla title written in Amharic characters which reads, in English, "The beginning of teaching, that is, a book of words for those who teach the Galla language, to show the people of Galla land the way to God (*Waq*), collected and printed by Awaj Onesimos and Ganon Aster, made this side of Massawa in the village of Monkullo 1894 years after our Lord Jesus Christ was born." The texts in this book are in the Mača dialect of Limmu,

from which Onesimos came. Its purpose seems to have been to discredit the Swedish Mission, for the texts ostensibly intended to "show the Galla the way to God" include songs of love and war, and ritual songs relating to pagan worship and the cult of Atete, the Galla fertility goddess.'

G. W. B. Huntingford, *The Galla of Ethiopia* (London: International African Institute, 1955) p. 18  
 Onesimos (c. 1855–1931) actually spent many years translating the Bible and other religious texts into Oromo. He was a slave, purchased and freed in 1870 and sent by the Swedish Evangelical Mission to study in Sweden 1876–81.

### The dialects of Oromo

In Ethiopia the major dialect divisions are *Well-egga*, *Mača*, *Tulama* (the dialect of Shoa, the heartland of modern Ethiopia), *Wallo* (where Oromo speakers are much intermixed with Amharic), *Eastern* (centring on Harar), *Arusi* and – to the south – *Borena*. The city of Harar itself has its own language, *Harari*.



In Kenya, three dialects can be distinguished: that of the Borena, Sakuye and Ajuran (a 'Somali' tribe in which many speakers are bilingual in Somali and Oromo); that of the Orma or Wardai, also spoken in Somalia; and that of the Gabra and Gurreh (who, again, consider themselves Somali). The Gurreh tribe extends into Ethiopia and Somalia.

# OSSETE

500,000 SPEAKERS

*Russia, Georgia*

**S**peakers of Ossete are separated by hundreds of miles from their nearest linguistic relatives. They are the last remnant of the steppe nomads of eastern Europe and central Asia who were speakers of IRANIAN LANGUAGES – Sarmatians, Scythians, Saka, Alans. They once ranged from the middle Danube valley to the foothills of the Pamir, and were prominent in Greek, Latin and Chinese history.

Politically divided, Ossetia lies across the centre of the northern Caucasus. Southern Ossetia is a district of newly independent Georgia. Political tensions there led to armed conflict in 1989–92. Northern Ossetia is within the borders of Russia, and has been nurtured as a bulwark of Soviet Russian influence among its neighbours.

Many Ossete speakers are bilingual in their national language, either Russian or Georgian. These are now a strong influence on the language. Other neighbouring languages, Turkic and especially Caucasian, have also affected Ossete: the large number of consonants, 33 in the Iron dialect, is a typical Caucasian trait. It has had little contact with other Iranian languages for as long as two thousand years. Like them, Ossete has lost the old noun declensions of Indo-European – but unlike them it has developed a new declension system with nine cases, nominative, genitive, dative, allative, ablative, inessive, adessive, equative and comitative.

The two main dialects of Ossete, Iron and Digor, are so different that they are scarcely mutually intelligible. There is some written literature in Digor, which is spoken on the western edges of North Ossetia. Iron is the basis of

standard Ossete, used in literature, education and the press. It is written in Cyrillic script, like Russian, with the addition of the letter Æ æ for a front *a* sound. Until 1954 South Ossetia used the Georgian alphabet: this provided symbols for the ejective or glottalic consonants *k' c' č' t' p'*, which Ossete shares with its neighbours the CAUCASIAN LANGUAGES. In Cyrillic these are written къ цъ чъ тъ пъ.

The first ten numerals in Ossete are: *iu, dyuö, örtö, tsyppar, fonz, ökhsöz, avd, ast, farast, dös*.

## The Alans and the Zelenchuk inscription

The language of the Alans was the direct ancestor of Ossete. They first emerge into history as they crossed the Caucasus, in the 1st century AD, to raid Armenia and the Parthian Empire (Josephus, *The Jewish War* 7.7.4).

The only surviving text in the Alan language is an inscription in Greek characters, from the 10th century AD, found beside the River Zelenchuk.

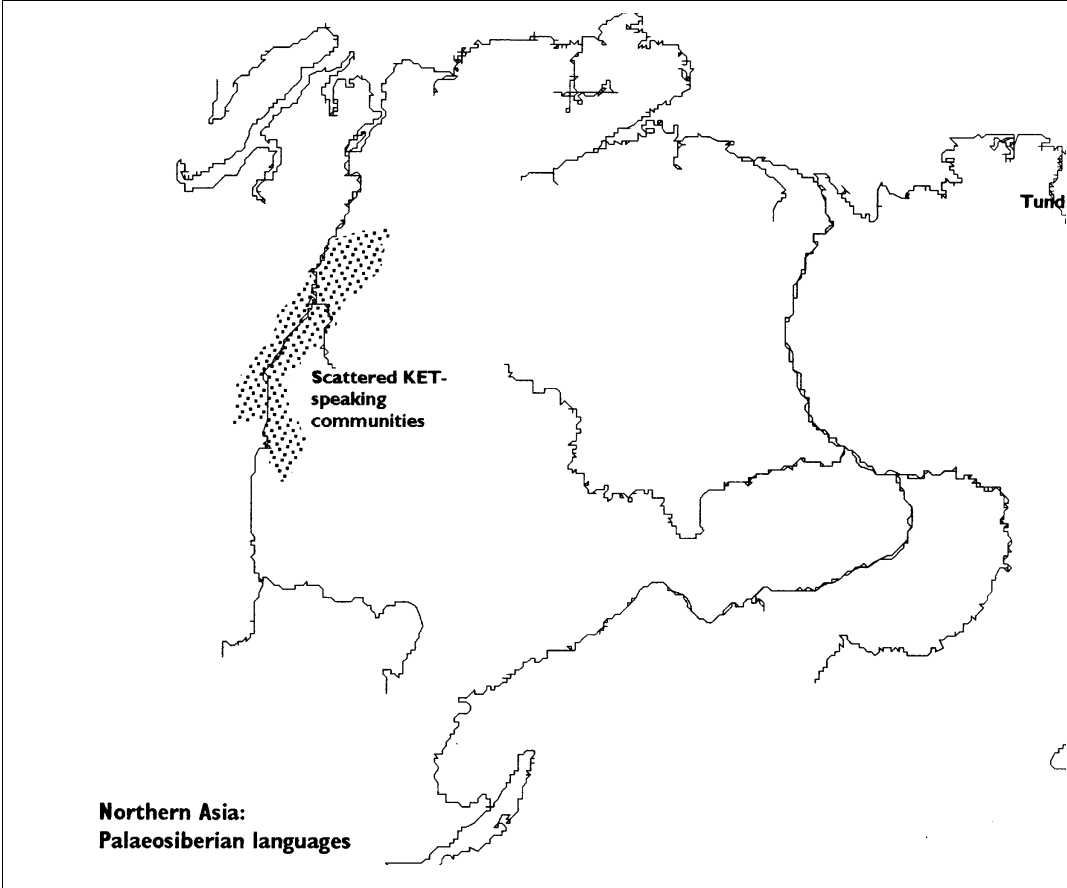


# PALAEOSIBERIAN LANGUAGES

**T**he minority languages of Siberia that do not belong to any of the larger language families are grouped, for convenience, under the label *Palaeosiberian*, ‘old Siberian’, or *Palaeoasiatic*. Some linguists used to call them *Hyperborean*, a word originally used in classical Greek for a mythical people who lived ‘beyond the North Wind’.

These labels cover three entirely isolated languages – and one very small family, Chukotko-Kamchatkan, so named after the two Pacific

peninsulas of north-eastern Siberia where these languages are spoken. Counting all together, they are spoken by only twenty-five thousand people. As with the SAMOYEDIC LANGUAGES and TUNGUSIC LANGUAGES, and some others of northern Russia, it is unlikely that the number of speakers of any individual language was ever more than a few thousand: Siberia has never been a hospitable environment. Yet, as with some of these others, there are strongly marked dialect distinctions within the languages, and





there is typically a cultural and linguistic divide between speakers who are nomadic and those who are settled. The facts show that languages can survive, for hundreds or thousands of years, with a tiny, highly scattered and mobile population base. In the past, at least, size did not matter.

The Chukotko-Kamchatkan languages include Chukchi or Luorawetlan, Koryak or Ny-mylyan, and Itelmen or Kamchadal, this last probably on the way to extinction. Together they occupy the far north-east of Siberian Russia. They are now threatened with extinction, by Russian development and migration.

Chukchi has 11,000 speakers on the far eastern Chukotka peninsula, and is used in primary schools and occasional publications. The small number of speakers belies Chukchi's regional importance. It was once the lingua franca of the north-eastern Siberian coasts: Yupik (see ESKIMO-

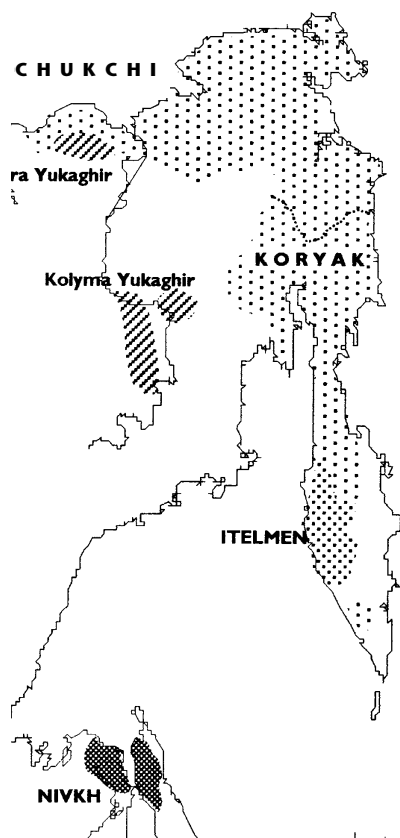
ALEUT LANGUAGES) has many Chukchi loanwords. There are Tundra Chukchi, nomadic reindeer-breeders, and Sea Chukchi settled on the coasts of the Bering Strait.

Koryak has about 6,000 speakers, settled on the Kamchatka peninsula. It is less favoured politically: children are taught in Russian.

The remaining three 'Palaeosiberian languages' have little in common with one another, and very little in common with any other known languages: they are true isolates. They are also far apart on the map. In fact the only reason for linking them together under one heading is the very dubious one of linguistic tradition.

Nivkh or Gilyak is spoken – in two very distinct dialects – by about 2,000 people on the banks of the lower Amur and on Sakhalin island. Yukagir has only 300 speakers, with two main dialects. The Tundra Yukagir lead a nomadic life in the valleys of the Alazeya and Chukoch'ya; the Kolyma Yukagir hunt and fish on the banks of the Yasachnaya and Korkodon.

Ket, or Yenisei Ostyak, is unique among the languages of Siberia in having a tone system: in this point, at least, it resembles Chinese, and it has actually been argued that Ket is a Sino-Tibetan language. It also has gender, masculine, feminine and neuter – again an oddity among Siberian languages. Ket is an isolate now, but two hundred years ago it was not: three other languages of 18th-century Siberia, Arin, Assan and Kott, were related to it. All died out long ago.



### Nivkh numerals

Nivkh has no close link with Japanese: but it does share one remarkable feature with Japanese, the numeral classifier system. For the numerals '1' to '5' there are twenty-six different forms depending on the class of object that is being counted. If it is a boat, the numerals are *ńim*, *mim*, *ńem*, *nəm*, *t'om*. If it is a dog-sledge, they are *ńiř*, *miř*, *ńeř*, *nəř*, *t'oř*.

Example from Bernard Comrie, *The languages of the Soviet Union* (Cambridge: Cambridge University Press, 1981) p. 269

# PALI

## Classical language of south and south-east Asia

**P**ali is one of the Prakrits (see INDO-ARYAN LANGUAGES), the one in which the dialogues of the Buddha, who died in the early 5th century BC, were recorded and recited. For a table of numerals see SANSKRIT.

*Pāli* meant 'series of texts, canon': *Pālibhāsā* was 'the language of the Canon'. Outsiders later misunderstood this compound word: Simon de la Loubère (*The Kingdom of Siam* (London, 1693)) took *Pali* to be the proper name of the language, and so did some Burmese and Sinhalese writers.

The text of the Buddha's teachings, and the other scriptures of the Theravada tradition, were – according to later sources – fixed at the first, second and third Buddhist Councils, the last of which took place under the Emperor Aśoka, who reigned c. 274–232 BC. But they were transmitted only in the memory and through the mouths of believers until, in Sri Lanka just before the time of Christ, they were at last written down.

If this is true, it may help to explain why Pali, the language of these scriptures, is a mixed dialect, not identical with the speech of any district of northern India, and showing signs of interference both from SANSKRIT and from the Dravidian languages of the south. It is certainly not – as tradition insisted – the *Māgadhī* or colloquial of north-eastern India, some form of which the Buddha himself must have spoken.

Once fixed in writing, Pali became the fairly stable medium in which Buddhists of the southern, Theravada tradition studied and wrote. The religion, and with it the Pali canon, spread successively to speakers of Sinhalese, Mon, Burmese, Thai and various languages of inland south-east Asia. Buddhism eventually disap-

peared from India proper, but Buddhists from Sri Lanka, Burma, Thailand, Laos and Cambodia still continue to learn and use this now-classical language of ancient India. Later Pali naturally shows the influence of its writers' varied mother tongues.

The dialogues of the Buddha are among the masterpieces of world literature. Another Pali classic is the *Questions of King Milinda*, recording a semi-legendary encounter between the Greek king of Bactria, Menander, and the Buddhist philosopher Nagasena. The stories of the Buddha's former lives, *Jātaka*, make up a far more complex text than these. Each of the 547 stories consists of brief verses (in a form of Pali with some archaic features), a prose retelling, a frame narrative and a detailed commentary on the verses.

Like the *Jātaka*, the corpus of Pali literature in its full form tends to be an interweaving of text, paraphrase, translation, commentary and sub-commentary. Many texts are bilingual. Most later writing comes from Sri Lanka, Burma (which adopted the Pali canon in the 11th century), then also Thailand and Cambodia. In all these countries the texts were normally recorded on palm leaf manuscripts, in local scripts. Local scripts are still used in printing Pali texts. European Buddhists and scholars – following this tradition – use the Latin alphabet.

As Pali became a classical language, one that had to be learned, a tradition of linguistic works developed. The dictionary *Abhidhanapadipika* 'The lamp of nouns' was compiled by Moggallāna, in the 12th century, on the model of the Sanskrit dictionary *Amarakoṣa*. A Sinhala translation of it was made in the 13th century.

**Literary rule**

Sahassam api ce pāthā anattapadasaṃhitā  
 ekaṃ pāthāpadaṃ seyyo yaṃ sutvā upasammati.  
 Better than a thousand anthologies of the path of wickedness  
 Is one single verse whose hearing brings peace.

*Dhammapada*, verse 101

The *Dhammapada* 'Path of morality' is an ancient collection of verse sayings attributed to the Buddha. Like the verses of the *Jātaka*, the *Dhammapada* gave rise to a later collection of prose tales, intended to explain why each verse was spoken. These stories, known as *Dhammapadaṭṭ-akathā* 'Dhammapada commentary', have been translated into English as *Buddhist legends*.

# PAMPANGAN

**1,850,000 SPEAKERS**

*Philippines*

**O**ne of the AUSTRONESIAN LANGUAGES, Pampangan is the main language of Pampanga province, in the central Luzon plain and immediately to the north-west of Manila (see map at TAGALOG). Pampangan is one of the eight ‘major languages’ of the Philippines.

Alternative forms of the name are *Kapampangan* and *Pampango*.

Pampangan was once written in a native script, a descendant of the Brahmi script of ancient India. This remained in use until fairly late in the Spanish period. However, printing in Pampangan – in Latin script – began as early as 1618.

Spelling was at first close to that of Spanish: a new orthography, similar to that adopted for Tagalog, was introduced in 1965.

Outside its heartland, there is a *barrio Kapampangan* ‘Pampangan suburb’ in Paco in metropolitan Manila. Pampangan is also spoken in four cities of Tarlac province (Bamban, Concepcion, Tarlac itself and Capas) and two cities of Bataan province (Dinalupihan and Hermosa) near the western borders of Pampanga.

The first ten numerals in Pampangan are: *métung*, *adwá*, *atlú*, *ápat*, *limá*, *anáam*, *pitú*, *walú*, *siám*, *apúlu*.

# PANGASINAN

1,650,000 SPEAKERS

*Philippines*

One of the AUSTRONESIAN LANGUAGES, Pangasinan is spoken in the central part of Pangasinan province in north central Luzon (see map at TAGALOG). It is particularly the language of the Lingayen gulf region, spoken in San Carlos, Dagupan, Lingayen (the provincial capital) and other nearby market towns. Around this heartland, the closely related Ilocano is the everyday language: Pangasinan speakers are almost equalled by Ilocano speakers in numbers, even in their home province.

*Pangalato*, the alternative name of Pangasinan, is a derogatory term now little used.

### Numerals in Pangasinan

In Pangasinan there are two sets of numerals. The inherited ones are used in traditional contexts: Spanish loan numerals are used in telling the time and, often, in trade.

Pangasinan		Spanish
sakey, isa	1	uno
dua	2	dos
talo	3	tres
apat	4	kuatro
lima	5	singko
anem	6	sais
pito	7	siete
walo	8	ocho
siam	9	nueve
samplo	10	dies

Richard A. Benton, *Pangasinan reference grammar* (Honolulu: University of Hawaii Press, 1971)

# PANJABI

**PERHAPS 60,000,000 SPEAKERS**

*Pakistan, India*

It is agreed among linguists that there are two Panjabi languages, one (Standard or Eastern or simply Panjabi) spoken in both India and Pakistan, the other (Lahnda or Western or sometimes 'Punjabi') entirely in Pakistan. Panjabi belongs to the INDO-ARYAN LANGUAGES, and to the dialect continuum of northern India and Pakistan: it has no firm frontiers with its neighbours, Hindi, Rajasthani and Sindhi.

*Panjab* is a place name of Persian origin: it means 'the five rivers', for the region is defined by the five great rivers, Jelam, Chenāb, Rāvī, Biās and Satlaj, which join the lower Indus. *Punjab* is an old-fashioned Anglo-Indian spelling of the word. The term *Lahnda* that was invented by Sir George Grierson, compiler of the *Linguistic Survey of India*, for the western dialects, comes from a Panjabi word for 'western'.

Indo-Aryan languages were spoken in this region from the early first millennium BC if not before. There are records of a language identifiable as Panjabi dating from the 11th century AD, but Panjabi literature really begins in the 15th century, and, from the start, follows two traditions. First came that of the Sufis, Muslim mystics, whose poetry – in Arabic script – was influenced by Persian traditions. The religious writings of the Sikhs, beginning almost at the same time, drew inspiration both from Sufi and from Hindu traditions. The Sikh scriptures, written by Guru Nanak (1469–1539) and his successors, make up the *Guru Granth Sahib*, the 'holy book of the Gurus', which was compiled by the fifth Guru, Arjan, in 1604. These early forms of Panjabi were very similar to the Braj that has now developed into

Hindi: indeed, Panjabi and Hindi are to some extent mutually intelligible even now.

In Pakistan the Sufi and Muslim tradition of Panjabi literature is nowadays weak, supplanted by Urdu, though the language is still pre-eminent in local Sufi contexts. Panjabi is little used in broadcasting and the press: Urdu (and English) take its place. In India, on the other hand, a modern standard of literary Panjabi was established in the 19th century, and Sikh authors write and publish copiously in the language.

In India, Panjabi now has about 15,000,000 speakers. Panjabi language and Gurmukhi script are identified with the Sikh religion; so much so that non-Sikhs are tempted to describe their language as Hindi, and some write in the Devanagari script that is typical of Hindi.

In Pakistan, where its speakers are Muslims, Panjabi or Lahnda is even more seriously under-reported in census figures. These dialects are actually spoken by the majority of the inhabitants of Pakistan, but many prefer to give their language as Urdu, Pakistan's official national language, in which they are for practical purposes bilingual. It is also under-reported because there has been no agreed name to cover the western dialects, differing fairly strongly as they do from the modern standard written language which is based on the dialects of Amritsar and Lahore. Terms like Western Panjabi and Lahnda are for linguists; Jatki, Multani and Hindko belong to individual dialects or dialect groups.

Sikhs are traditionally mobile. As soldiers in the British Army many served in Singapore, Hong Kong and Australia. In the early years of this century considerable numbers went as railway workers to East Africa, as farmers to British

Columbia and California. From about 1950 there was massive migration to Britain. The 400,000 or more Panjabi speakers now in Britain cluster in certain larger cities, including London (the boroughs of Southall and Newham), Birmingham, Wolverhampton, Coventry, Leeds and Bradford.

Texts from the *Guru Granth Sahib*, now some four hundred years old, are used in Sikh religious ritual. The language is still comprehensible without much difficulty to worshippers. Whatever their language of education, Sikhs need to know the Gurmukhi script in order to read and understand the scriptures.

Panjabi is unusual among Indo-Aryan languages in having three tones on stressed syllables: falling, as in *ਖੋੜਾ* *kòṛa* 'horse'; high rising, as in *ਕੋੜਾ* *kōṛa* 'leprous'; level, as in *ਕੋੜਾ* *kōṛa* 'whip'. The scripts (like those of south-east Asia) indicate the tones by varying not the vowel signs but the consonant symbols.

Example from G. A. Zograph, *Languages of South Asia* tr. G. L. Campbell (London: Routledge, 1982) pp. 49–50

## Panjabi and its dialects

*Western Panjabi* or *Lahnda* can be divided into Southern, North-western and North-eastern dia-

lects. The Southern dialects (south of the Salt Range) include *Shāhpūrī* and *Mūltānī*. These merge into the SIRAĪKĪ of Sind. The North-western dialects, sometimes called *Hindkō*, include the dialects of Attock, Kohat, Peshawar and Abbottabad. The North-eastern dialects include the *Pōthwārī* of Rawalpindi.

*Mājhī* is the central Panjabi dialect, spoken in the regions of Lahore and Amritsar. This is the dialect that forms the basis of standard Panjabi and the language of literature. There is naturally considerable variation in the standard as spoken by Muslim and by Sikh speakers, the former in Pakistan, the latter in India.

*Eastern Panjabi* includes *Dōābī*, dialect 'of the two rivers'; *Pōwādhār*, *Rāthār*, *Mākwār*, and *Bhaṭṭiānī*, linguistically close to Rajasthani and spoken in Ferozepore and Ganganagar Districts. Eastern Panjabi dialects shade into Western Hindi.

Also sometimes regarded as a dialect of Panjabi is DOGRI, the language of Jammu, with its dialects *Kāṅgrī*, *Kaṇḍālī* and *Bhaṭālī* (the last-named spoken in Chamba District, Himachal Pradesh).

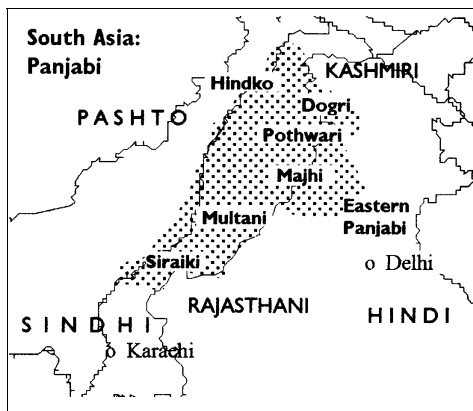
## The Gurmukhi alphabet

॥ ॥॥ ਇ ਈ ਓ ਓ ਏ ਐ ਓ ਐ ਕਖਗਘਙ ਚਛਜਝਞ ਟਠਡਢਣ ਤਥਦਧਠ ਪਫਬਭਮ ਯਰਲਵ ਸਸਹ

a ā i j u e ai o au k kh g gh ñ c ch i j h ñ t th d dh n t th d dh n p ph b bh m v r l v s s h

## The Gurmukhi script

Panjabi is most often written and printed in *Gurmukhi* script, ‘from the mouth of the Guru’. This was devised by Angad, the second Sikh guru, in the 16th century. It is a formalised and extended version of the *Landa* script that was and still is used by tradesmen in Panjab and Sind. The usual Roman transliteration, given here, follows the normal style for south Asian romanisation and does not attempt to mark tones. In Pakistan Panjabi is written in the Perso-Arabic script familiar from URDU.



Numerals in Panjabi and Romani				
	Standard Panjabi	Pothwari	Multani	Romani
1	ek	hikk	hekk	ek
2	do	do	d ṁ	dui
3	tin	tre	trae	trin
4	car	cār	cār	shtar
5	pañj	pañj	pañj	pañj
6	che	che	che	sho
7	sat	satt	satt	efta
8	aṭh	aṭṭh	aṭṭh	okhto
9	nau	naum	nam	enya
10	das	das	ḍāh	desh
In some Romani dialects, as here, the numerals '7' to '9' are borrowed from Greek.				



# PAPIAMENTU

**250,000 SPEAKERS**

*Curacao, Bonaire, Aruba*

**P**apiamentu is a creole based on PORTUGUESE and SPANISH. Spoken on three of the islands of the Dutch Antilles, it naturally shows Dutch influence on its vocabulary.

Linguists dispute the relative strength of Portuguese and Spanish influences in the make-up of Papiamentu. In early Dutch Curacao, in the late 17th century, the strongest linguistic presence may have been that of Spanish and Portuguese Jews, previously settled in Latin America, refugees from the Inquisition. They became active in the Dutch trade in African slaves, for which Curacao was the main entrepot.

Papiamentu might well help in studying the early history of JUDEZMO.

Developing largely independently of standard Spanish, and spoken by the great majority of inhabitants of the three Dutch islands, Papiamentu is now the essential lingua franca on all three, widely used among all social classes. It was recently made an official language of Curacao and Bonaire, though not of Aruba, where Dutch alone has that status.

Of the three island dialects that of Aruba, which lies close to the Venezuelan coast, shows continuing influence of Spanish.

# PAPUAN LANGUAGES

'P<sup>apuan</sup>' is a catch-all term for the languages of New Guinea (Irian) and nearby islands that do not belong to the family of AUSTRONESIAN LANGUAGES.

*New Guinea* was so called, in the 16th century, by the Spanish explorer Ortiz de Retes. He likened the dark-skinned, frizzy-haired people of this great unexplored island to those of the Guinea coast in West Africa. *Papua*, the general term for the inhabitants and their languages, derives from Malay *papuah* 'frizzy'. The Malay name for the island, however, is *Irian*.

This is linguistically the most complex region of the world. In mountainous, forested and swampy country, full of obstacles to travel, the languages of New Guinea have been developing and interacting for 40,000 years. It is only relatively recently that languages from the outside world, Austronesian and then Indo-European, have become established on the island. These provide its three best-known lingua francas – MALAY, HIRI MOTU and the English-based pidgin, TOK PISIN.

Communities in New Guinea are typically small: even in the heavily populated Highlands, villages seldom have more than three hundred inhabitants. The language map is an elaborate patchwork, and it will appear still more fragmented when exploration is complete. Languages move as groups migrate; they merge or disappear as one community comes to dominate another; they split as villages lose contact with one another.

Ternate, the island sultanate off Halmahera, once ruled parts of the northern New Guinea coast. In the 19th century the Dutch claimed the west, the British the south-east, and the Germans the north-east of the island. British and

German territories, brought together under Australian government, became independent as Papua New Guinea or PNG. Dutch territory was transferred to Indonesia in 1962 and forms the province called Irian Jaya.

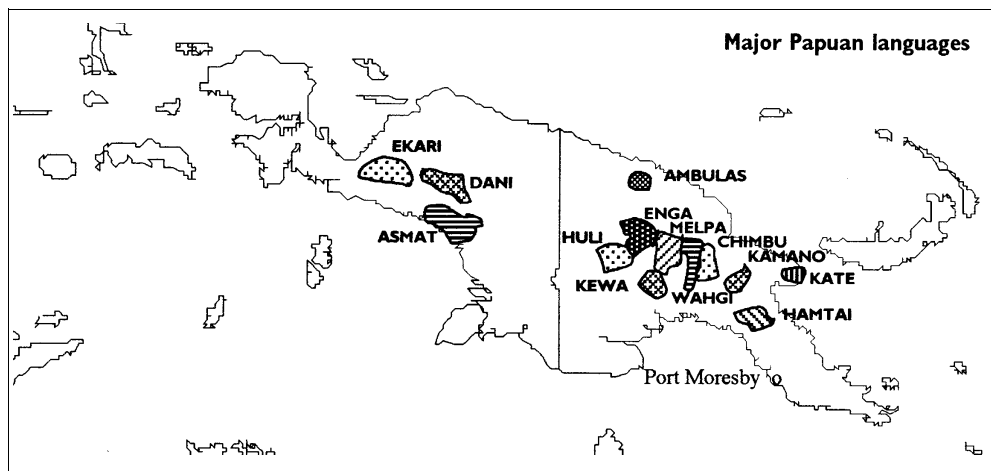
The first surviving wordlist of a Papuan language (excluding TERNATE) was made by two passengers on the Dutch ship *Triton*, when it called at Utanata, where Miriam was spoken, in 1828. Papuan languages were shown to be historically distinct from Austronesian by S. H. Ray in a paper on 'The languages of British New Guinea' presented to the Ninth Orientalist Congress in 1893.

It is a massive challenge to historical linguistics to trace language relationships that may date back 40,000 years or more. Genealogical trees have been drawn that link all the languages of New Guinea into a very few 'phyla', but for the present these all-embracing families are little more than blueprints for further research.

The administrative centre of Irian Jaya, once called *Hollandia* ('Netherlands' in Modern Latin), is now *Jayapura* ('Victory City' in Sanskrit), both languages quite alien to the culture of this remarkable island. *Port Moresby*, capital of PNG, is so called because Captain John Moresby mapped this natural harbour in 1873 and named it, modestly, after his father, Admiral Sir Fairfax Moresby.

## Non-Austronesian languages of New Guinea

About 750 Papuan languages are known, some of them very patchily. The great majority have fewer than 2,000 speakers apiece. Only those spoken by relatively large communities are listed here. For the West Papuan family see TERNATE.



For the Timor-Alor-Pantar family see box at MALAY.

*Hamtai*, Kamea or Kukukuku (40,000 speakers) belongs to the Angan family of Gulf, Morobe and Eastern Highlands provinces.

*Asmat* or Yas, with Citak and Kamoro, forms an important group (70,000 speakers) of the Central and South New Guinean family of the south coast of Irian Jaya.

The Dani-Kwerba language family includes as many as 300,000 speakers of the dialects of *Dani*, in inland northern Irian Jaya.

The East New Guinea Highlands family is a very large one. Languages of this family include *Enga* (165,000 speakers) and *Huli* (70,000), both of Southern Highlands and Enga Provinces; *Melpa* or Medlpa (70,000) and Wahgi (45,000), of Western Highlands Province; the *Chimbu* group (70,000), in Chimbu Province; the *Kewa* languages (60,000), in Southern Highlands Province; *Kamano* (50,000), in Eastern Highlands Province.

*Kâte*, of Morobe Province, is used as a religious language by the Lutheran church, and is known to a total of 80,000 people. It belongs to the Huon-Finisterre family.

The Sepik-Ramu family, a very extensive one, includes *Ambulas*, with 35,000 speakers in East Sepik Province.

*Ekari* (100,000 speakers) is a language of the Highlands. It belongs to the Wissel Lakes-Kemandoga family.

### Language pride in Papua New Guinea

Neighbouring groups in Papua New Guinea had contact through intermarriage, trade and warfare, leading to a certain amount of bilingualism or competence in other dialects. A sizeable minority of New Guinean women have had the experience of being linguistic 'foreigners' in the village into which they have married.

'We might well ask why such contacts did not lead to a *lessening* of linguistic differences. A partial explanation probably lies in the fact that New Guineans often make use of other-language and other-dialect knowledge in rhetoric and verbal art, highlighting the known differences between their own and neighbouring speech varieties. It appears that contacts with and awareness of other languages have led not to levelling but to heightened consciousness of and pride in difference.'

Gillian Sankoff, *The social life of language*  
(University of Pennsylvania Press, 1980)  
pp. 9–10, abridged

# PASHTO

PERHAPS 14,000,000 SPEAKERS

*Afghanistan, Pakistan*

**P**ashto is one of the two official languages of Afghanistan. Among modern IRANIAN LANGUAGES it is second only to Persian in the length of its written history. For the other official language of Afghanistan, see PERSIAN, DARI AND TAJIK.

The usual name *Pashto* (written *Pushtu* in older English sources) may in origin be identical with *Persian*, originating in a proto-Iranian form *\*parsawā* 'Persian language'. The speakers are called *Pakhtūn* in northern Afghanistan, *Pashtūn* in the south, and *Pathān* in Pakistan. The name *Afghan* is occasionally used for the language.

Pashto literature is recorded in manuscript form from the 16th century onwards. A 17th-century author, Khushḥāl Khān Khaṭak, is now regarded as Afghanistan's national poet. But Pashto has become a language of education and the press only in the course of the 20th century. The Dari variety of PERSIAN was the language of government in Afghanistan until 1933, when Pashto was given this role. Arabic script is used, in its Persian form, with several added symbols.

This state belonged to a people called *Patane*, lords of the hill country. Just as those who live on the skirts of the Pyrenees, on both sides, control the passes by which we cross from Spain to France, so these Pathans control the only two land routes into India.

*Décadas da Ásia* (iv.vi.1) by the Portuguese historian João de Barros, 1553

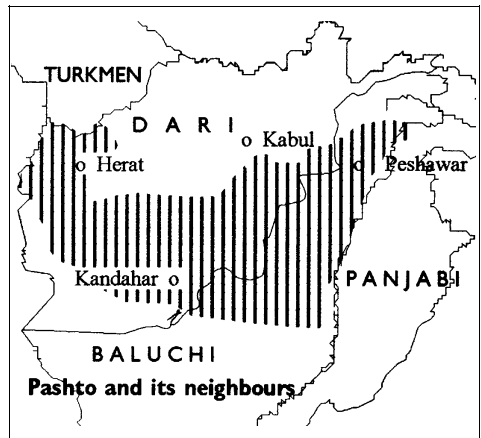
The first ten numerals are *yau*, *dwa*, *dre*, *calor*, *pinjə*, *špaž*, *owə*, *atə*, *nəh*, *las*. Pashto

has borrowed heavily from Persian, the former ruling language in this region, and from languages of Pakistan including Western Panjabi and Urdu. It has acquired the retroflex consonants *ʈ ɖ ʀ ʎ* that are typical of the Indian linguistic area.

## Pashto in Afghanistan and Pakistan

The Pashto-speaking population is almost equally divided between the two countries. Speakers in Pakistan are known as Pathans or Afghans.

Kandahar belongs to the South-western dialect group, Kabul to the North-western (sometimes called 'Eastern', forgetting the Pakistani speakers). The South-eastern dialects are those of Baluchistan, the North-eastern those of the North West Frontier Province of Pakistan, including Peshawar. The Kandahar and Peshawar dialects are the basis of the two standard written forms of Pashto.



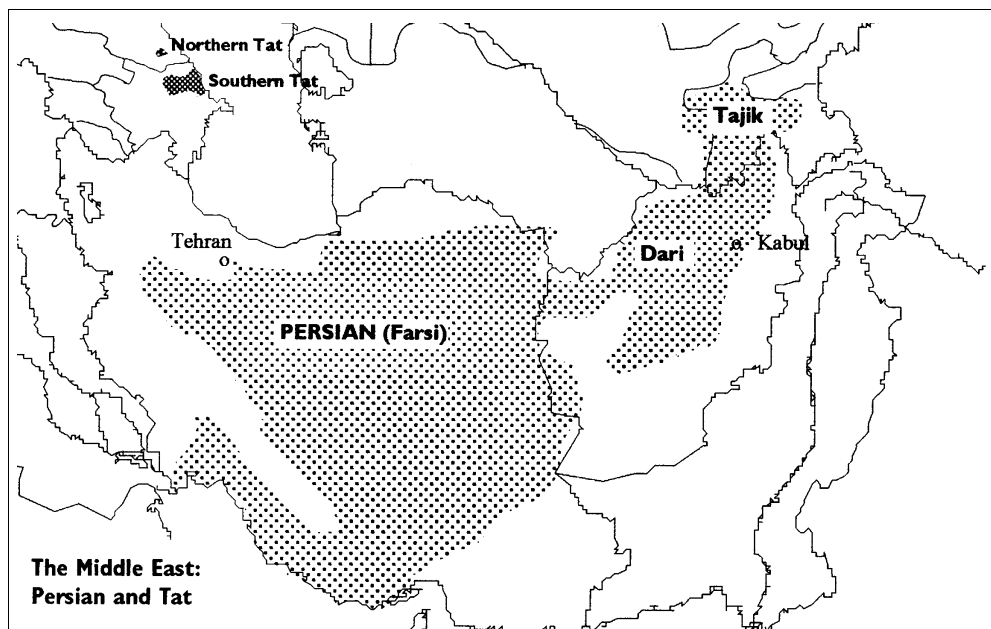
**Moral**

Wagâfī wāfa . kār na xpāl kā;  
mardān hayā day . čē kār da bāl kā.

All men work for their own good;  
The real man is he who works for others.

Lines 1–2 of a rubāʿī by Khushhāl Khān. In Pashto, as in Persian, this is a four-line verse form in which lines 1, 2 and 4 rhyme. Each half-line has five or six syllables.

Transliteration and translation: D. N. Mackenzie, 'Pashto verse' in *Bulletin of the School of Oriental and African studies* vol. 21 (1958) pp. 319–33



# PERSIAN, DARI AND TAJIK

**PERHAPS 31,000,000 SPEAKERS**

*Iran, Afghanistan, Tajikistan*

**P**ersian has a far longer recorded history than any of the other IRANIAN LANGUAGES – a time-span which almost equals that of a fellow Indo-European language, Greek. It is the national language of Iran; in the form of Tajik, of Tajikistan; in the form of Dari, one of the national languages of Afghanistan.

Within the wider geographical context of Iran, the region known to Greeks and Romans as *Persis* corresponds to the modern province still called *Fārs*. *Fārsī* or *Pārsī* is the natural name for the language of this region, homeland of the rulers of the ancient Persian and later Sassanian Empires. *Pārsī-yi Darī*, ‘Persian of the court’, spread as such to Persian-ruled countries such as Afghanistan, particularly to the cities: there it is now called *Darī* or Kabuli.

*Tājik* was originally (as *Tāzī*) a name given by Iranians to the invading Arabs. Turkic peoples, before their own conversion to Islam, adopted *Tāzīk* as a name for both Arabs and Muslim Persians.

ARAMAIC, not Persian, was the trade and civil service language of the Persian Empire. But its rulers, from Darius I (550–486 BC) onwards, were proud of their Persian origin. Old Persian springs to life as one of the three languages of the Behistun inscription in which Darius announced his achievements (see box).

As the Persian Empire fell to Alexander of Macedon and his Greek successors, Aramaic gave way to Greek: but Greek never spread

far in Iran. Soon the Parthian Empire, whose ruling class was Iranian, held sway from Iraq to Afghanistan: the language of its elite, Parthian, gradually replaced Greek in administration.

*Parthian*, an Iranian language close to Old Persian but not its lineal descendant, was used increasingly for literature – by Zoroastrians, Christians and Manichaeans. Most Parthian texts that survive today come not from Iran but far off to the east, at the old oases on the Silk Road in Chinese Xinjiang. Texts in Parthian, Middle Persian, Uighur, Khotanese, Tocharian, Sanskrit, Tibetan, Chinese and other languages testify to the multilingual culture of these remarkable cities.

Persis remained independent of the Parthians. From here, eventually, emerged the next ruling dynasty, that of the Sassanian Empire, which ruled the Middle East from the 3rd to the 7th century. Its language is called *Pehlevi* – the traditional name – or *Middle Persian* or (when written in the old Avestan script) *Pāzand*.

Zoroastrianism – in its later form, sometimes described as ‘fire worship’ – was the state religion of the Sassanians, as it had been of Darius and his successors long before. The holy books of Zarathustra’s teachings were now collected and written down. They were in the ancient AVESTAN language, which was becoming difficult for worshippers to understand. Middle Persian literature consists largely of commentaries on the Avestan texts alongside new legends and hymns.

The fact that Parthian, Middle Persian and Avestan were all written in Aramaic script testifies to the transmission of the written culture of the Persian Empire, through the years of Greek domination, to its Parthian and Sassanian successors.

The Islamic conquest of the Sassanian Empire begins a new phase in the history of the Persian language. At first ARABIC was the language of government and culture: in particular, Arabic had to be used in Islamic worship. Persians are outstanding among authors of medieval Arabic poetry and prose.

With political resurgence, literature in *Modern Persian* emerged. Now written in the Arabic script, the language is full of Arabic loanwords, spreading from religious and administrative terminology to many facets of everyday life. Persian poetry has borrowed metres and verse forms from Arabic, and has transmitted them in turn to Turkish and Urdu. But Persian literature is far from imitative. One of the earliest masterpieces of Modern Persian, the 10th-century *Shāh-nāma* of Firdausī, is a literary epic – a form unknown to Arabic literature – that gathers and retells the long, legendary history of Persia. Original to Persian is the four-line *rubāʿī* best known in the verses of Omar Khayyām (1048–1131) that were turned so elegantly into English by Edward Fitzgerald.

Once more a language of government, under Mongolian and Turkish rulers, Persian was brought to Turkey, to central Asia and to India. It was the language of administration in India for hundreds of years – in Kashmir until after 1900.

In the earlier period of British rule, Persian was still the most useful linguistic accomplishment for the British who were involved with the administration of India. Even in the 19th century the British in India learned Persian before, or in addition to, 'Hindustani' or Urdu. The Persian of India was by now a distinct variety, heavily influenced both by the Turkish of the old ruling class and by the Indian culture that surrounded it. Textbooks warned British students that this was not the language of Iran.

Now almost forgotten in India, Persian remains a language of power well beyond the

frontiers of Iran. Often called *Dari*, literally the 'court language', it is the second language of Afghanistan and acts as a lingua franca throughout the country. It was the language of government there until 1933. The nomadic Aimaq and the Hazaragi of Afghanistan, culturally distinct, are Persian speakers.

To the north again, Persian speakers were a considerable proportion of the population in central Asia when that region came under Russian rule in the 19th century. The Soviet Union attempted to divide central Asia on linguistic lines: Tajikistan was formed as the national republic for Persian speakers of the north. Now an independent state, its tortuous frontiers are evidence that linguistic boundaries could not be effectively drawn here. Speakers of 'Tajik' and UZBEK lived side by side, and still form sizeable minorities in each other's republics.

Old Persian retained the noun inflexions of proto-Iranian, but Modern Persian has lost them almost completely. The noun suffix *-i*, called *iẓāfeh*, which derives from the Old Persian *yo* 'who, which', is used so frequently to form noun-noun and noun-adjective phrases that it is a distinguishing mark of Persian – and is often found in loans in Turkish and Urdu.

Persian displays its long history in loanwords from many sources. From Akkadian *pīru* 'elephant' came Old Persian *pīru* 'ivory' (and, from this, Sanskrit *pīlu* and modern Persian *fīl*); other early loanwords from Semitic languages are *maskk*, name of the inflated skins that armies can use to cross rivers; *tandūr* 'clay oven'. Greek *diphthera* 'skin' is the origin of middle Persian *diftar* 'book' and so of modern Persian *diftar* 'exercise book' and of *daftar* 'office'. From Chinese comes *chang* 'harp'. But the majority of the vocabulary of modern written Persian comes from Arabic; other, more recent loans are drawn from Turkish (especially military terminology), French and English.

### Persian and its relatives: the map on page 493

Persian, Dari and Tajik are the most-used names for the varieties of Persian that are

standard in Iran, Afghanistan and Tajikistan. Although any national state tends gradually to develop its own linguistic standard, the differences between these three are as yet slight. However, ‘Colloquial Persian is not everywhere alike. The pronunciation may differ largely. So I found the pronunciation of the priest in the fire-temple at Yazd difficult to follow, after becoming accustomed to the Isfahānī pronunciation. Vocabulary also may differ. In Isfahān I used the word *gurīkht* “he ran away” in vain. They had *varmālid*. But *gurīkht* was known to the villagers of Taft near Yazd’ (H. W. Bailey, ‘Western Iranian dialects’ in *Transactions of the Philological Society* (1933) pp. 46–64).

*Northern Tat* or *Jewish Tat* is the language of the Mountain Jews, a population of perhaps 15,000 in Dagestan (Russia) and at Kuba and Vartashen on the northern edge of Azerbaijan. Their cultural centre is Derbent. For these speakers, Tat is a language of education and the press.

*Southern Tat* is spoken by Muslims in north-eastern Azerbaijan. Their official language is AZERĪ, in which most are bilingual.

Still further to the south, a quite separate group of Iranian languages, spoken in Khalkhal and Qazvin in Iran, is confusingly called *Azari*,

*Tati* or *Southern Tati*. These belong to the ‘Caspian dialects’ and are not mapped here (see GILAKI).

Numerals in Persian			
		Middle Persian (Pehlevi)	Modern Persian
1	۱	ēvak	yak
2	۲	dō	dō
3	۳	sē	sih
4	۴	chahār	chahār
5	۵	panj	panj
6	۶	shash	shish
7	۷	haft	haft
8	۸	hasht	hasht
9	۹	nō	nuh
10	۱۰	dah	dah

Persian in writing

The letters ط ; س ص sound like Persian ط ; س sounds like Persian ت. The three superfluous letters, along with ع ظ ض ذ ح are hardly ever used except in Arabic loanwords.

In Soviet Tajikistan Latin script was adopted in the 1920s, to be replaced by Cyrillic script around 1940.

The Persian alphabet	
ي ه و ن م ل گ ک ق ف غ ع ظ ط ض ص ش س ژ ز ر ذ د خ چ چ ث ت پ ب ا	
a b p t ʃ j č ħ kh d ʒ r z z s š ʒ ž ʔ ʔ gh f q k g l m n v h y	



# POLISH

40,000,000 SPEAKERS

## *Poland*

**P**olish is one of the West SLAVONIC LANGUAGES – a close relative of Czech, Slovak and Sorbian.

*Polak* is the name Poles have traditionally given to themselves, and as *Polack* it is sometimes used in US English for ‘Polish immigrant’. It may once have meant ‘people of the fields’, as does the related word *Poljane*. Several foreign names for the country – Latin *Polonia*, French *Pologne*, German *Polen*, English *Poland* – come directly from this. In other languages a quite different term is used instead of ‘Pole’ or ‘Polish’: Hungarian *Lengyel*, Lithuanian *Leñkas*.

An obsolete English word for a three-masted merchant ship, *polacre*, means simply ‘Polish’.

The River Vistula, which flows through central Poland from south to north, roughly marked the prehistoric western limit of Slavonic-speaking territory. One of the chain of northern migrations that brought on the collapse of the Roman Empire was a Slavonic expansion westwards, across the Vistula, into what is now western Poland and eastern Germany. In due course the pressure was reversed: throughout medieval times German speakers pushed eastwards again into Poland, conducting what could be described at times as a crusade or a war of extermination. The German defeat at Tannenberg in East Prussia, in 1410, marks the end of this long episode. Even after it, however, German was a language of culture and power in Poland.

Throughout this period Poland remained an independent state and at times a very powerful one, uniting with Lithuania in 1569 to rule the

regions now called Belarus and Ukraine. But Polish sovereignty was extinguished in 1795, and its territory was shared out between Prussia, Austria and Russia.

Poland reasserted its independence in 1918 after the Russian Revolution, gaining territory both from Germany in the west and north and from Belorussia and Ukraine in the east. Vilnius, which is now the capital of Lithuania, had a Polish majority and was Polish territory between 1920 and 1939. Poland suffered disastrously in the Second World War. In its aftermath the country’s frontiers were drastically redrawn. German speakers were driven westwards across the ‘Oder-Neisse line’, the new western border, while the Soviet Union also advanced westwards into territory that had formerly been Polish. For some decades Poland itself was a Soviet satellite state – but in the 1980s Polish resistance eventually undermined Russian rule over eastern Europe.

Poland has been a Christian state since AD 966, and remains staunchly Catholic. Christianity was introduced by missionaries from the Czech lands. Polish thus has many early Czech loanwords. Latin loanwords came later, less from religious than from humanist and scholarly influences. Because of the country’s later political history there are also loans from German, Belorussian and Ukrainian.

Polish literature begins with a 14th-century collection of sermons, *Kazania świętokrzyskie*, and a hymn, *Bogurodzica*, which may be older. At this early period much writing, both religious and secular, was in Latin. The greatest Polish poets are perhaps the Renaissance humanist Jan Kochanowski (1530–84) and the exiled ‘national

bard' Adam Mickiewicz (1798–1855), author of the epic *Pan Tadeusz*.

Among a later generation of émigrés from Poland was the English novelist Joseph Conrad (1857–1924; born Konrad Korzeniowski), who began his career as a sailor.

Emigré communities grew further during and after the Second World War, with Poland's continuing political vicissitudes. The three largest communities are in the United States, Canada and Britain. The core of the Polish-speaking community in Britain, numbering 90,000, is formed by the airmen and soldiers who fought with the allies after the German seizure of Poland. At least two-thirds of this exiled group was male, leading to much intermarriage – and so to rapid assimilation. A Polish daily newspaper is

still published in London, *Dziennik Polski*.

A strong Polish minority remains in Lithuania and in western Belarus. Until the 18th century Polish had been the language of the elite in these countries, as it was also in the Ukraine. The Belorussian coat of arms once bore the words 'Proletarians of all countries, unite!' in Belorussian, Russian, Polish and Yiddish.

Polish has an oddly complex system of grammatical gender. The singular has masculine, feminine and neuter genders – and, in the accusative case and masculine gender only, a further distinction between animates and inanimates. The plural has only a 'masculine personal' and an unmarked gender, the former used for groups including at least one human male. For a table of numerals see CZECH.

**Asserting Polish**

A niechaj narodowie wždy postronni znaja  
Iż Polacy nie geśi, że swój język mają!

Let other nations always know  
That Poles are not geese, but have their language too!

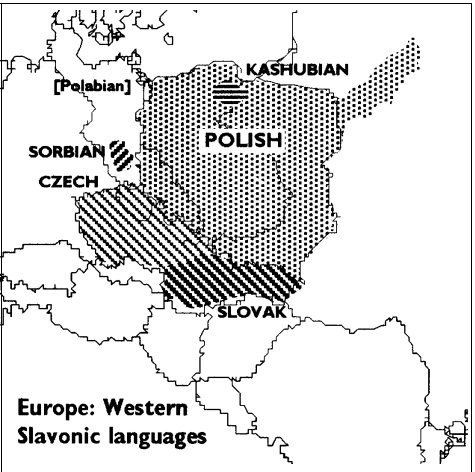
Mikolaj Rej (1505–69); from R. G. A. de Bray, *Guide to the Slavonic languages* (London: Dent, 1951) p. 591

**The Western Slavonic languages**

'Literary *Polish* has drawn upon the dialects of the three successive capitals of the land: the Great Polish dialect of Gniezno, the Little Polish dialect of Cracow and the Mazovian dialect of Warsaw. The Silesian dialects served it as a filter for Western borrowings during the Middle Ages and the Renaissance, while the dialects of the eastern borderlands influenced it during the Baroque and Romantic periods . . . This is why the debate on the origin of the Polish literary language, which raged from the beginning of this century until after World War II and pitted the Great Polish hypothesis against the Little Polish one, has predictably ended in a deadlock' (*The Slavic literary languages: formation and development* ed. Alexander M. Schenker, Edward Stankiewicz (New Haven, 1980) p. 209).

The Mazovian dialect of Polish shares some features with *Kashubian* (Cassubian or Pomeranian), the language of about 150,000 speakers in the countryside around Danzig. All are bilingual

in Polish: many can claim to be quadrilingual, speaking also the Low German of their former peasant neighbours and the High German they were taught at school. Kashubian is identifiable in documents from the 15th century onwards. There is some local literature in print, though no standard spelling has been fixed.



SORBIAN is the only remaining Slavonic language that is spoken west of the rivers Oder and Neisse in what is now Germany. *Lower Sorbian* is the speech of the countryside around Cottbus, south of Berlin. *Upper Sorbian* centres on Bautzen near the Czech border.

*Polabian* was once spoken to the north and north-west of Berlin. It became extinct around 1700.

There is a continuum of dialects from CZECH south-eastwards through *Moravian* to SLOVAK. Moravian dialects differ from south to north, the northern having much in common with neighbouring varieties of Polish. Slovak divides into Western, Central and Eastern dialects. The

Central dialect has similarities with South Slavonic languages such as Slovene, although geographically separated from them.

### The look of Polish

The Latin alphabet required drastic adjustment to Polish phonetics. Symbols were required for the nasal vowels *ą* and *ę*, the palatalised consonants *pi bi fi vi mi ki gi*, the alveo-palatal fricatives *cz dź sz ż rz* and the palatal fricatives *ć ci dź dzi ś si ź zi ń ni*. It is the frequent use of *z* (a rare letter in most other European spelling systems) that makes Polish look so unpronounceable.

# PORTUGUESE

**155,000,000 SPEAKERS**

*Portugal, Brazil, Mozambique, Angola, Guiné and Cape Verde*

**O**ne of the ROMANCE LANGUAGES, Portuguese is in origin a twin of GALICIAN (see map there). Through Portuguese discoveries and conquests it has become one of the major languages of the world.

The Iberian peninsula is the western extremity of the area in which Latin had become the everyday language of the Roman Empire. As the empire collapsed, the north-west corner of the peninsula – modern Galicia and northern Portugal – became the territory of a tribe of Germanic invaders, the Suebi. Little is known of the language they spoke, though it is possible that one Portuguese word, *britar* ‘break’, may come from Suebic. Latin remained the everyday language: this region is the cradle of early Galician-Portuguese, ancestor of the two modern languages.

Most of Portugal, like most of Spain, came under Arabic-speaking rule in 711–18. The Cantabrian mountains, south as far as the River Douro, remained unconquered: here the two fiefs of Galicia and Portugal were established. Portugal – south of the River Minho, and already including some new conquests as far south as the Mondego – became an independent kingdom in 1143. Portugal’s dialect spread southwards with continuing conquests, displacing the somewhat different Romance dialect, MOZARABIC, which had been spoken by the subject population in the Islamic period. But that is only half the story: like Spanish, Portuguese has borrowed from Arabic many terms for cultural concepts, government and administration, architecture and the arts, evidence that the high Islamic culture of the south exerted great influence on the conquering Christians of the north.

Galician-Portuguese is first recorded in the late 12th century, in legal documents and in the earliest examples of a lyric poetry which continued to flourish until the 14th (see box). Even in Spain, Portuguese was felt to be the proper language for lyric: the devotional *Cantigas de Santa Maria* are the work of Alfonso the Wise, King of Spain (reigned 1252–84). By the 14th century Galician and Portuguese were beginning to grow apart, Galicia remaining a backwater of Spain while Portugal was self-consciously independent, except for the sixty-year ‘captivity’ – the union with Philip II’s Spain – in 1578–1640. During that period, some Portuguese intellectuals took to writing Spanish.

The growth of the Portuguese kingdom and the progress of its discoveries and conquests were recorded in a series of chronicles, histories and memoirs, including the monumental *Décadas da Ásia* of João de Barros (named after the ‘Decades’, ten-book divisions of the Roman history written fifteen centuries earlier by Livy). The voyage of Vasco da Gama was also immortalised in one of the greatest of literary epics, *Os Lusíadas* by Luís de Camões.

By the 16th century Portuguese must have sounded very much as it does today. Writers and scholars liked to emphasise its differences from Spanish and its historical links with Latin: old-fashioned Portuguese spelling reinforced these, continually reverting to the Latin forms of words in defiance of real pronunciation. A reform (*nova ortografia*, ‘new spelling’) was adopted in Portugal in 1916: slightly modified, it was accepted in Brazil in 1931. It abandoned these ‘etymological’ spellings in favour of a relatively phonetic system. However, with nasal vowels, ‘whispered’

final vowels and frequent final *sh/zh* sounds Portuguese pronunciation is highly distinctive. While Portuguese speakers can understand Spanish relatively easily, Spanish speakers find spoken Portuguese much more difficult.

Besides its Arabic loanwords Portuguese has borrowed from Tupí, from Indian and south-east Asian languages, and from French. There are now many English loanwords. For a table of numerals see CATALAN.

### The world-wide range of Portuguese

With the settlement of the Azores in 1439, and later Madeira, Portuguese became one of the first European languages to spread outside Europe by colonisation. Soon afterwards Portuguese speakers began to settle in Brazil. Eventually Brazil came to outweigh by far the tiny mother country, not only in size but in terms of number of Portuguese speakers.

The Portuguese of Portugal and that of Brazil have a relationship comparable to British and American English. To European speakers of the language, Brazilian Portuguese has a distinctive accent, some important differences in vocabulary and some minor differences in spelling. Naturally, a considerable proportion of modern Portuguese literature is in fact Brazilian literature.

However, the range of varieties of Brazilian Portuguese may well be greater than those of American English – for good historical reasons. For a long period the lingua franca of Brazil was not Portuguese at all but *Lingua Geral* (see TUPÍ): this gave way to pidgin and creole varieties of Portuguese, as well as to the standard language. Africans brought to Brazil as slaves used a pidgin Portuguese as their language of communication. Even recently, numerous migrants to Brazil from European countries have developed their own community languages: *Fazendeiro* was the creole once spoken in São Paulo by a mixed Italian-African population.

A true Brazilian Creole form of Portuguese still exists in some rural black communities; it is known as *Tabarenho*, *Matutenho* or *Caipira*.

### Portuguese creoles and pidgin languages

On Portugal's Indian Ocean trade routes, pidgin Portuguese must have been in use from the 16th century onwards: arguably its origin goes back even to the earliest West African voyages in the 15th century. For a while, Portuguese pidgin was a lingua franca of trade between Europeans and Indians in India. In the Portuguese territories of Africa and Asia this pidgin must have become a mother tongue, a creole, quite early.

In its original locality it has living and thriving descendants: the CRIOULO of Cape Verde Islands and Guiné, the related creoles of São Tomé and Príncipe (where Portuguese remains the language of education), and (on an island that long ago came under Spanish rule) *Fá d'Ambó* 'Language of Annobón'.

In the larger mainland territories of Angola and Mozambique stable Portuguese creoles have not become established, though pidgin forms of the language have been used. In both of these, now independent states, Portuguese retains its importance as a language of administration and education. The distinctive Portuguese of Angola is used by several novelists.

As a household language and a lingua franca, pidgin Portuguese was already so well established in Sri Lanka, which the Portuguese lost

#### Sri Lanka Portuguese Creole in 1807

'There is still a large body of inhabitants at Columbo and the other settlements in Ceylon, known by the name of Portuguese. They probably number to the amount of five thousand; they are, however, completely degenerated, and exhibit complexions of a blacker hue than any of the original natives . . . The greater part of them were admitted by the Dutch to all the privileges of citizens, under the denomination of burghers. A corruption of their original language is still spoken over all the sea coasts. It is very easily learned, and proves of great utility to a traveller who has not time to study the more difficult dialects of the natives.'

Rev. James Cordiner, *A description of Ceylon*  
(London, 1807) pp. 88–9



to the Dutch between 1632 and 1658, that it not only survived but was adopted in Dutch households too. Portuguese Creole, influenced by both Tamil and Sinhala, is still spoken there by several thousand people of mixed European, Lankan and African origin, who are known as 'Burghers' and 'Kaffirs'. The larger groups are to be found in Colombo, Trincomalee and Batticaloa. There is also a community of Creole speakers in Kandy, though that city was never under Portuguese rule: it was the capital of an independent kingdom in Portuguese and Dutch times.

At Goa, the long-lasting base of Portuguese rule in India, the linguistic history is quite different. Continuing contacts between Portugal

and the colony led to gradual changes in the creole, which – by the time of the Indian takeover in 1961 – was better described as a strongly marked regional form of Portuguese.

Further east, in Malaysia, Malaccan Creole or *Papia Kristang* still has some thousands of speakers; Singapore Portuguese is still in religious use in the Catholic community. Several recorded Portuguese pidgins and creoles of coastal towns in south-east Asia and Indonesia are now extinct, including that of Jakarta – but in Macao, a Portuguese possession due to become Chinese in 1999, there are perhaps, 4,000 speakers of the *Macauenho* creole.

The early Portuguese pidgin of the West African coast has more descendants than this.



Portuguese as a world language

There seems to be no doubt that it is part of the ancestry of the Spanish, English and French creoles of the Atlantic and Indian Oceans. Each of them has a different history, and the Portuguese element is most easily recognisable in PAPIAMENTU (usually classed as a Spanish creole) and Saramaccan and Matuwari, which have become ENGLISH CREOLES.

Portuguese has not ceased to spread across the world. There are currently at least 4,000,000 Portuguese-speaking émigrés, including as many as 1,000,000 migrant workers in France. There are strong Portuguese settlements in California, Massachusetts and Toronto and a large community of Madeiran origin in Venezuela.

### The influence of Portuguese

The Portuguese discoveries and conquests were crucial in establishing contacts between European cultures and those of other continents. In west and east Africa, southern Asia, the Malay archipelago and south America, the presence of Portuguese mariners, traders, soldiers, missionaries and administrators ensured that Portuguese loanwords would be adopted in many of the languages of the world. In African languages the commonest Portuguese loanwords are *mesa* 'table', *sapata* 'shoe' (originally from Persian by way of Turkish and Italian), *chumbo* 'lead', *igreja* 'church', *ouro* 'gold', *prata* 'silver', *chapeu* 'hat' (originally from French) and *carreta* 'cart'. Portuguese provided names for newly imported American fruits and other foods: *mandioca* 'manioc, cassava', *goiaba* 'guavas', *anas* 'pineapple', *papaia* 'pawpaw', *milhos* 'maize'.

Examples from J. Knappert in D. Dalby and others, *Language and history in Africa* (London: Cass, 1970) p. 86

### The influence of Portuguese pidgin

Words from the 15th-century Portuguese West African Pidgin have become familiar across the world.

**Savvy:** *Savvy* – from Portuguese *saber* or *Lingua Franca* *sabir* – means 'know' in many of the creole languages of the world: *sibi* in Crioulo, *sabe* in Krio and Bislama, *save* in Tok Pisin, *sabiam* in Kamtok. By the 18th century, this word had been borrowed from the pidgins or creoles into English, where *savvy* colloquially means 'know-how' or 'skill'.

**Piccaninny:** Portuguese *pequenino* means 'little one, child'. By 1657 the word was in the pidgin or creole of Barbados. The form in Bislama and Tok Pisin is *pikinini*; in Sranan (the English creole of Suriname) it is *pikiē*; in Kamtok it is *pikin*, in Jamaican Creole *pikni*. In the English of colonial America, South Africa and Australasia *piccaninny* became a usual word for a 'black child'.

Examples from Frederic G. Cassidy in *Pidginization and creolization of languages* ed.

Dell Hymes (Cambridge: Cambridge University Press, 1971) pp. 207–9

**Portuguese lyric**

Proençaes soen mui ben trobar  
 e dizen eles que é con amor,  
 mays os que troban no tempo da frol  
 e non en outro, sey eu ben que non  
 am tam gran coyta no seu coração  
 qual m'eu por mia senhor vejo levar.

The Provençals can make very good songs  
 and they say that they do it with love,  
 but of people who make songs at the time of flowers  
 and not at other times, I know very well that they  
 have no such great care in their hearts  
 as I feel in mine for my lord.

In general, medieval Galician-Portuguese lyric poetry owes much to Provençal (OCCITAN) models. But the typical *Cantiga de amigo*, in which the speaker is a woman mourning an absent lover, is not so close to Provençal poetry in its feeling – as King Diniz (1279–1323), author of this example, rightly says. In fact it has more in common with the tone of Mozarabic lyrics, suggesting that it reflects something of the medieval popular songs of the Iberian peninsula.

The early lyric poetry of Portugal was forgotten for centuries. In the 19th century much of it came to light once more in two priceless manuscripts discovered in Italian libraries.



# QUECHUA

**PERHAPS 9,000,000 SPEAKERS**

*Bolivia, Peru, Ecuador, Argentina, Colombia*

**O**ne of the AMERIND LANGUAGES, Quechua was the ruling language of the powerful Inca empire destroyed by Spanish conquerors in the 16th century.

*Quechua* (*Quichua* in Ecuador and Argentina) is said to derive from a local term for the temperate Andean valleys and their inhabitants. By Peruvian speakers themselves the language is sometimes called *runa simi*, 'language of men', and some linguists have used that term to distinguish the old Quechua dialects of central Peru. By early Spanish authors it was called *lengua general*, or more fully *lengua general del Inga*, the 'common language of the Incas', or *lengua general del Perú*. Colombian dialects are called *Inga* or *Ingano*, 'Inca'.

The first printed work in Quechua was a dictionary by Domingo de Santo Tomás, published in 1560. Early texts of importance include Jurado Palomino's *Quechua catechism*, the hymns collected by Cristóbal de Molina, and the dramas *El pobre mas rico* (16th century), *Usca paucar* (18th century) and *Ollantay*. In spite of the large number of speakers, Quechua nowadays appears rather little in print and in the media.

The wide geographical range of Quechua reflects the reach of Inca government in its last years. The most distant regions of Quechua speech, in Ecuador, Colombia, southern Peru, Bolivia and northern Argentina, are unexpectedly alike in some dialect details. This suggests that they originate from relatively recent colonisation at a time when a standard form of speech formed a lingua franca throughout the vast empire – an empire that was surprisingly centralised

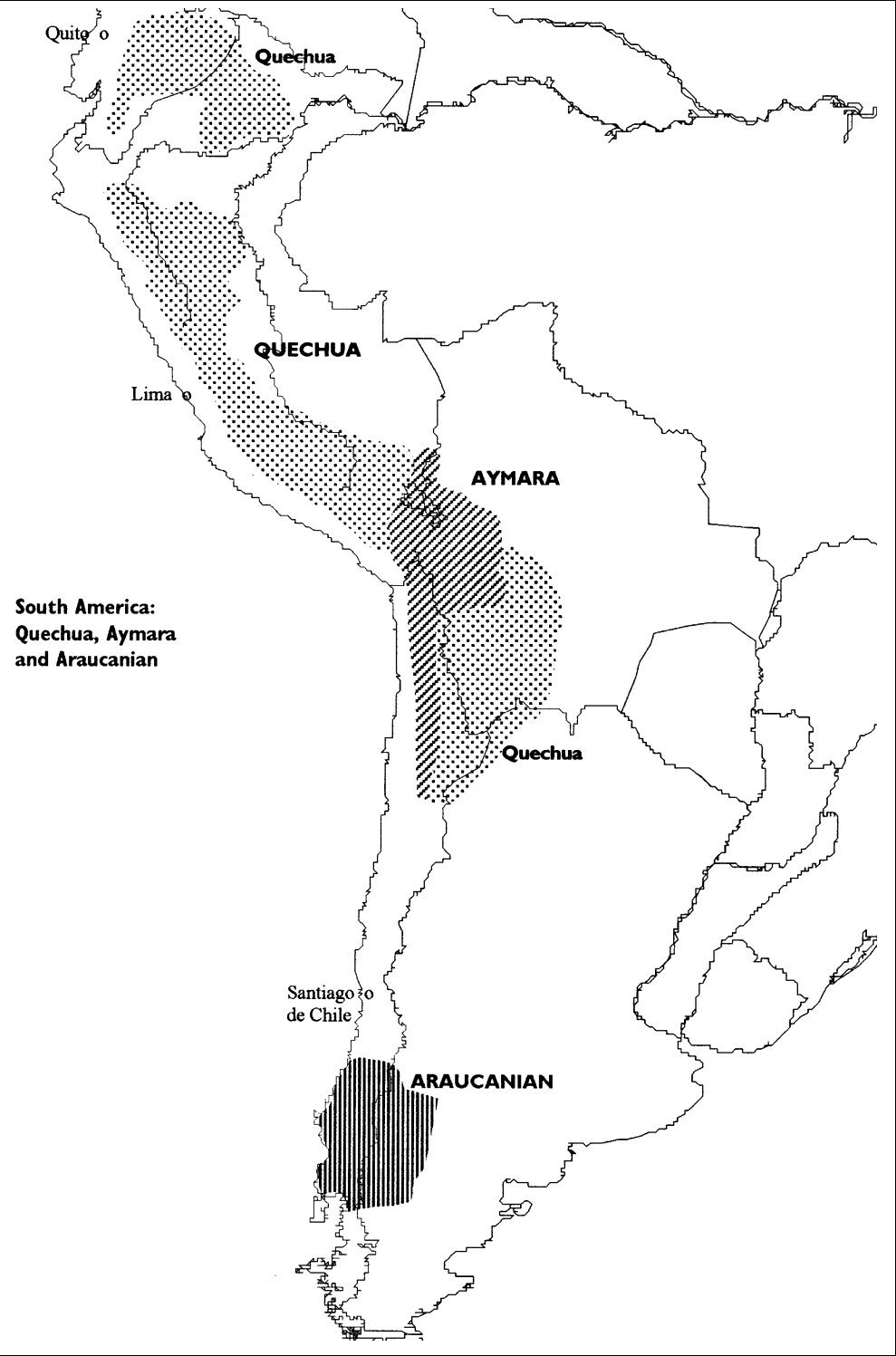
in its administration, with a highly organised army and much movement of population. By contrast, the dialects of central Peru are more strongly differentiated, as if originating from an earlier historical period.

Curiously, Quechua had not been the language of the empire's ruling stratum for very long. Until the Inca themselves came to power, in the 15th century, the rulers had been speakers of *Chimú*. This unrelated language survived among a few speakers until the 19th century, but it is now extinct. For an outline of *Chimú* see George L. Campbell, *Compendium of the world's languages* (London: Routledge, 1991) pp. 302–5.

Quechua has continued to spread, superseding local languages, under Spanish rule, though at the same time it continually gives way to Spanish. There was once a pidgin language, *Media lengua*, with Spanish grammar but largely Quechua vocabulary. Modern pidgins still exist: *Chaupi lengua* in Ecuador, *Llapui* in Bolivia, the latter with Quechua suffixes but Spanish words.

Nowadays Spanish, as the official language throughout the Andes, threatens the continued vitality of Quechua. In Peru Quechua was declared joint official language in 1975, but lost this status in the constitution of 1979.

Quechua 'documents' were transmitted across the Inca empire by means of quipu strings. The colour of the strings, the shape of the knots and the number of knots all contributed to fix a meaning which could be repeated by the messenger who carried the string, but, it is usually argued, could not be reconstructed by others. Those found by archaeologists could not be



interpreted unless they were essentially numerical, having to do with astronomy and the calendar. However, recent research by W. B. Glynn has suggested that quipu strings were a partly phonetic recording system.

The first ten numerals in Quechua are: *shuj*, *ishcai*, *quina*, *tahua*, *pichica*, *sucta*, *canchis*, *pusaj*, *iscun*, *chunja*.

Crucial for reconstructing Inca civilisation is a work written in Spanish by the son of a Spanish conquistador and an Inca lady. Garcilaso de la Vega 'el Inca' published his *Royal commentaries of the Incas and general history of Peru* in Córdoba, Spain, in 1609.

## Languages of the Andes

Quechua extends widely along the Andes chain, an indication of the former extent of Inca power and colonisation. The oldest dialects appear to

be those of central Peru (Ancash, Huánuco, Junín, Pasco and part of Lima department).

Quechua country is divided by the Aymara-speaking region in Bolivia. The Quechuan dialects from the Bolivian border north-westwards as far as Ayacucho have been said to show signs of stronger Aymara influence than the rest, as if Aymara had earlier been spoken there.

AYMARA, which itself shows strong Quechua influence, has about 2,000,000 speakers in the neighbourhood of Lake Titicaca.

ARAUCANIAN speakers, who number about 300,000, classify themselves as *Pikunche* 'northerners', *Huilliche* (or *Veliche*) 'southerners' and *Pehuenche* 'foresters'. Dialect differences among these are said to be considerable.

In early colonial times three *lenguas generales*, 'common languages', of the Andes were recognised: Quechua, Aymara and the now-extinct Puquina. Jesuit missionaries were expected to learn all three.

### The traditions of Huarochirí

Inca mythology and ritual are recorded in the *Traditions of Huarochirí*, written in Quechua for the Spanish missionary Francisco de Avila. Avila ordered the record to be made because 'diabolical practices are best combated by those who are fully informed of them'. But what did his Quechua assistant (known only as 'Tomás') think of the task? He begins:

Runa Yno ñiscap Machoncuna ñaupá pacha quillcacta yachanman carca, chayca hinantin causascuncunapas manam canancamapas chincaycuc canman, himanam Viracochapas sinchi cascanpas canancama ricurin, hinatacmi canman. Chayhina captinpas canancama mana quillcasca captinpas caypim churani cay huc yayayuc Guarocherí ñiscap Machoncunap causascanta, yma ffeenioccha carcan, yma yñach canancamapas causan, chay chaycunacta . . .

If the Ancients of the people called Indians (Yno) had known writing, then all the traditions of their former life, now doomed to fade away, would have been preserved. They would have shared the fortune of the Spaniards (*Viracocha*) whose traditions and past prowess are on record. Since it is not so, I shall write down the traditions of the Ancients of the land called Guarocherí, into whom one father breathed life: their *faith* and their customs as they are remembered to this day . . .

The 'one father' of the district of Guarocherí is the Inca god Pariacaca – and yet, for the 'faith' of the Ancients, the writer borrows a word from the Spanish missionaries, *fe*.

# QUICHÉ

850,000 SPEAKERS

*Guatemala*

**O**ne of the MAYAN LANGUAGES, Quiché was at the time of the Spanish ascendancy the language of an important people whose centre of government was at Umatlán in modern Guatemala. Quiché speakers now form the second largest linguistic community of Guatemala, but the language has no official status.

The masterpiece of traditional Quiché literature, *Popol Vuh* 'Book of Counsel', was written down in the Latin alphabet in the 16th century on the basis of an earlier text in Mayan hieroglyphics. Both versions are now lost, but a 17th-century copy of the alphabetic text survives. It tells the mythical and historical story of the Quiché. A second important work is *Rabinal Achí*, a drama, apparently pre-Columbian in origin but written down only in the 19th century. Its lengthy dialogues take place between a victor and a prisoner of war who is fated to be sacrificed. Quiché poetic literature, like that of NAHUATL, is typified not by metre but by the use of parallel expressions.

Quiché has many loanwords from Nahuatl. They clearly come from a dialect similar to

one now spoken on the Pacific coast, and, just as clearly, they result not from Aztec activities (which did not reach this part of central America) but from earlier Toltec influence. The words concerned are mainly religious or military: examples include 'altar', 'incense', 'demon', 'axe', 'palace' – but also 'cradle' and 'fishnet'.

## The creation of human beings

In their last, successful attempts to create human beings, the Gods found the necessary ingredients in the Mountain of Nourishment:

'Thus they were pleased with the provisions of the Good Mountain, full of sweets, stiff with yellow maize and white maize, thick with paxtate and cacao, stuffed with zapotes, anonas, jacotes, nances, matasanos and sweetmeats, rich foods filling the stronghold that is called Broken Place, Place of Bitter Water. All edible fruits were there, all cereals large and small, all vegetables large and small' (*Popol Vuh*).

# RAJASTHANI

PERHAPS 20,000,000 SPEAKERS

*India*

**R**ajasthani is the collective name for the INDO-ARYAN LANGUAGES or dialects of the Indian state of Rajasthan (see map at GUJARATI). Rajasthan was politically highly fragmented, consisting until the end of British rule of numerous small feudal states.

In a wide sense Rajasthan now belongs to the Hindi-speaking area. In censuses, many speakers will say that their language is Hindi. There is no strong literary tradition in any modern Rajasthani dialect: Hindi is the usual written language. However, Rajasthani is linguistically closer to Gujarati than to Hindi, and it makes most sense to regard it as an independent language. Old Gujarati, as recorded in texts of the 12th to 14th centuries, is the common ancestor of both Gujarati and the Rajasthani dialects of today.

Heroic literature, as well as religious poetry, was much cultivated at the courts of old Rajasthan. At most of them, the Braj form of HINDI was the usual medium. At the court of Marwar, however,

two literary languages were in use. *Pingal* was a mixture of Apabhramsa and Braj with local Rajasthani elements. *Dingal*, known in manuscripts from the 15th century onwards, was a literary form of Rajasthani itself: it is sometimes described as the caste dialect of the *caṛans* or bards.

## Numerals in two Rajasthani dialects

Mālvi		Mārwarī
ek	1	ek
do	2	doy
tīn	3	tīn
cār	4	ciyār
pāñc	5	pāñc
che	6	chau
sāt	7	sāt
āṭh	8	āṭh
nau	9	nau
das	10	das

## Pābūjī: brigand, local god and epic hero

Heroic and epic song is popular in modern Rajasthan. The story of Pābūjī, a 14th-century brigand or warlord, is performed by singer-priests, *bhopos*. This is the opening of an episode recorded from performance in 1976:

Baiṭā Pābūjī koḷū maṇḍa rai darbār:  
 baiṭā ṭhākār Pābūjī ek ghoṛām rī vātām kijai cālvai.  
 Cāmdā sāṃvat phir ryā āṃpe aṇ dhartī rai cāruṃ kijai mer:  
 Pābū rai caḍāi deval nī mīlai . . .

Pābūjī was seated in his court at Koḷ;  
 As he sat there lord Pābūjī raised the matter of horses.  
 'Chāndā my chieftain, we have travelled round the four borders of the earth,  
 We have not found a horse for Pābūjī to ride . . .'

John D. Smith, *The Epic of Pābūjī* (Cambridge: Cambridge University Press, 1991) pp. 116, 285

# REJANG

PERHAPS 1,000,000 SPEAKERS

*Indonesia*

One of the AUSTRONESIAN LANGUAGES, Rejang is spoken to the south of Minangkabau and to the north-west of its close relative Lampung along the south-western coastal strip of Sumatra. Contacts with Malay, the trade language of the archipelago, go back hundreds of years. Today there is much bilingualism in Indonesian, now the official form of Malay: massive migration from other parts of Indonesia threatens the stability of Rejang and Lampung culture. Rejang has a rich literature of love lyrics, tribal

histories and epics – which outsiders have traditionally not been allowed to hear. It is largely a literature of group performance, transmitted orally. However, texts are sometimes written down in the old Rejang alphabet ultimately descended from the Brahmi of ancient India. Only a small proportion of the population needs literacy in this alphabet, though it is also used for magical and medical incantations and spells. The local poetic language mixes Rejang words with older Malay. See map at LAMPUNG.

### The Rejang alphabet

‘The Rejang writing is so simple, uniform, and perspicuous, both in regard to the form of the characters and disposition of the syllables, that from this evidence alone I should not hesitate to pronounce it the design and execution of one head and hand.’

William Marsden, ‘Remarks on the Sumatran languages’ in  
*Philosophical transactions of the Royal Society of London* vol. 71 (1781)

ha ga nga ta da na pa ba ma cha ja nya sa ra la ya wa ha mba ngga nda nja a

Barat lawut  
tunggu maring gunung  
meteri keilangan  
sumeui maring gunung  
meteri kekasi  
ruma ketunun  
bulan purnama

Lover’s poem of grief, written on a bamboo tile. Translation by Mervyn Jaspan,  
*Redjang ka-ga-nga texts* (Canberra: Australian National University, 1964) pp. 36–7, 68

As the sea  
On a mountain  
The princess is lost  
To an ancestral tiger on the mountain,  
The princess my love.  
A house of weaving;  
A full moon.

# RIFIA

**PERHAPS 1,500,000 SPEAKERS**

*Morocco, Algeria*

**R**ifia is one of the BERBER LANGUAGES (see map at TAMAZIGHT). It is spoken along the northern coast of Morocco and eastwards into western Algeria as far as the neighbourhood of Alhucemas. There are scattered groups of Rifia speakers even further east, as far as Arzeu, the result of recent migrations.

The Berber languages of Morocco are sometimes grouped under the name Shilha, and Rifia may be called Northern Shilha. It is also called locally *Tarifit*, ‘language of the Riff’, the name of the people who speak it. It has no official status in Morocco or Algeria. Male speakers are usually bilingual in Arabic, but women are not.

# ROMANCE LANGUAGES

**T**he last Roman Emperor was deposed and the western empire, shaken by Germanic invasions and internal chaos, collapsed in AD 476.

For five centuries LATIN had been the language of government and education, and the everyday language of a growing number of people, throughout the western provinces, from Spain all the way to the Balkans. Easy travel and communications had kept the spoken Latin of the empire surprisingly uniform: there were no big dialect differences.

Latin was far too well established to disappear: but with political fragmentation, there was no longer anything to prevent local accents and dialects emerging. Some became the languages of independent governments. These, overshadowing their rivals, achieved the status of standard languages.

And so, where it has not been displaced by later invasions, Latin still survives – in the form of its descendants, the Romance languages. In fifteen hundred years, they have grown widely apart; yet all can be clearly seen to originate in the language of the Roman Empire.

The Romance languages all differ from Latin in one interesting way. They have a definite article, equivalent of English *the*: French *le, la, les*; Portuguese *o, a, os, as* and so on. How did this feature originate? It has been suggested that the idea is borrowed from Greek. Admittedly such basic features are not usually borrowed from language to language. But Christian texts, in Greek and in literal Latin translation, had a huge influence on the way that later Latin was spoken and written. In these texts Latin *ille* ‘that’ is frequently used as an equivalent to Greek *ho* ‘the’; and this Latin word is indeed the origin of the Romance definite article.

## Lingua Franca

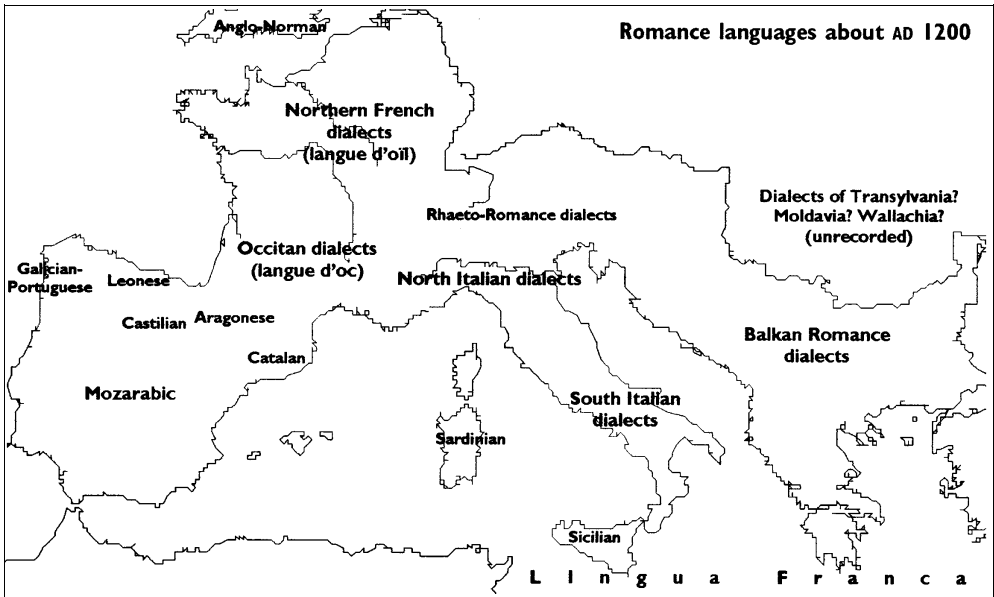
Ever since people began to sail the Mediterranean they have had to communicate with speakers of other, multiple languages. Indirectly we know of the resulting language mixtures even in prehistoric times, thanks to the word-borrowing to which early Mediterranean languages such as Latin, Greek and Egyptian offer clues.

In the Middle Ages, for the first time, a few wordlists and phrases of the Mediterranean trade language were written down. It was called *Lingua Franca*, ‘Frankish language’, after the usual Arab term for West Europeans. It also has the name *Sabir* – which meant ‘know’ in *Lingua Franca* – because a common opening gambit was ‘Do you know . . . ?’ This same word, origin of English *savvy*, turns up in many of the modern pidgins and creoles of the world (see box at PORTUGUESE). The words of *Lingua Franca* came from Italian, Occitan, Catalan, Spanish, Portuguese and Arabic (*taibo* ‘good’ is the one Arabic word in Encina’s poem, see box).

This trade language or pidgin of Mediterranean seaports had also another use. The Christian slaves acquired in large numbers by the Barbary pirates of Islamic North Africa needed to communicate with one another and with their owners: *Lingua Franca* served their purpose, gradually shifting in vocabulary depending on the current origins of slaves.

*Lingua Franca* no longer exists: or does it? In North Africa in the 19th century the slaves’ *Lingua Franca* eventually merged into French. But from its use on the Mediterranean seaways and coasts a form of *Lingua Franca* came to be spoken among British homosexual groups, especially sailors and theatre people, under the name *Polari* (meaning ‘speak’, Italian *parlare*). And according to one theory, championed by the linguist Keith Whinnom, the French and Eng-





lish creoles of the world all descend from medieval Lingua Franca because this was the basis of the 16th-century Portuguese traders' pidgin.

Lingua Franca is the oldest recorded trade language or pidgin. So the phrase *lingua franca* has become a standard term for a medium of communication among speakers of multiple languages.

### The reach of Rome

The first map shows Latin in retreat. The invasions of the Lombards, Goths and Franks had no long-term impact on the developing Latin/Romance speech of the old empire. Anglo-Saxon, Slavonic, Hungarian and Arabic settlements had a stronger effect. In these regions, by the 12th century, Romance speech had disappeared. MOZARABIC, the Romance language of Islamic Spain, was to be replaced by Spanish in the course of the *Reconquista*. Dalmatian, the language of the Adriatic coast, died out in the 19th century. Throughout the Middle Ages, Lingua Franca aided communication in the multilingual ports of the Mediterranean.

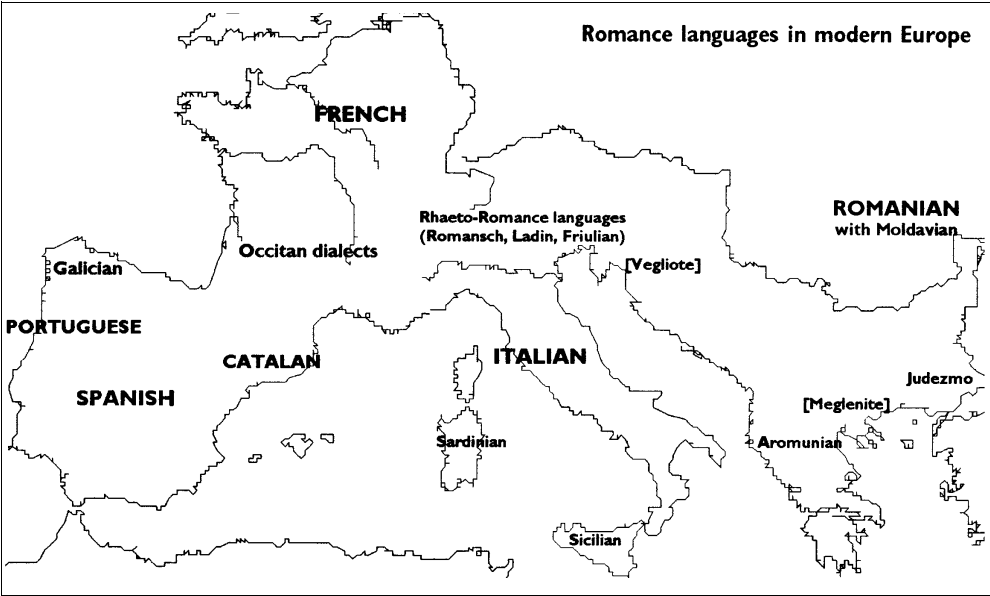
In the second map the Romance languages of modern Europe are shown. The 20th-century Romance languages are PORTUGUESE, GALICIAN, SPANISH, CATALAN, OCCITAN, FRENCH, ITALIAN,

### Reconstruction confirmed

A 'proto-Romance' language can be reconstructed from the modern Romance languages, just as with other language groups. There is one difference: a real ancestral language, Latin, is on record. So what do the slight differences between reconstructed proto-Romance and actual Latin mean? In general, they show the difference between the high culture of the educated elite – the writers and audience of classical Latin literature – and the everyday speech of the masses. It was the latter that gave rise to the various Romance languages. Also they remind us that even in a rich corpus of literature some rare words may go unrecorded.

On occasion, scholars reconstruct forms in proto-Romance, and these forms are later discovered in previously unknown documents or inscriptions. Thus on the basis of Romanian *struț*, Italian *struzzo*, Provençal and Catalan *estrus* and Old Spanish *estruc*, 'ostrich', we would reconstruct a proto-Romance form *strūtju*; this form has now been discovered, written *struthius*, in the late Latin translations of the Greek medical writer Oribasius.

Robert A. Hall, Jr., *External history of the Romance languages* (New York: American Elsevier, 1974) p. 74, adapted



**The earliest record of Romance speech**

In AD 842 an alliance between Charles the Bald and Louis the German was sealed by the taking of mutual oaths by the French and German monarchs and by their armies. So that everyone could understand what was going on, the oaths were taken in the everyday languages of the two countries, and not in Latin. The official chronicler Nithard recorded the proceedings word for word. Louis took the oath in French:

Pro Deo amur et pro christian poblo et nostro commun salvament, dist di in avant, in quant Deus savir et podir me dunat, si salvarai eo cist meon fradre Karlo et in aïudha er in cadhuna cosa, si cum om per dreit son fradra salvar dift, in o quid il mi altre si fazet et ab Ludher nul plaid nunquam prindrai, qui, meon vol, cist meon fradre Karle in damno sit.

For God’s love and for the Christian people and our common salvation, from this day onward, so far as God gives me knowledge and power, I will defend this my brother Charles and will be of help in everything, as a man in justice ought to defend his brother, so long as he does the same for me, and I will accept no plea from Lothar which might, by my will, damage this my brother Charles.

For the German text as sworn by Charles, see box at GERMAN.

Nithard, *Histoire des fils de Louis le Pieux* ed. Ph. Lauer (Paris: Les Belles Lettres, 1926) p. 104

SARDINIAN, ROMANSCH, Ladin, Friulian, ROMANIAN and AROMUNIAN. Three of these have been carried far across the world as the languages of

colonial empires. They now have descendants of their own. For more details see FRENCH CREOLES, PORTUGUESE and SPANISH.

**The Mediterranean pilgrimage**

Peregrin taibo cristian,  
Si kerer andar Jordan  
Pilya per tis jorns pan,  
Ke no trobar pan ne vin.

Good Christian pilgrim,  
If you want to reach the Jordan  
Steal bread for your days of travel,  
For you will find neither bread nor wine.

Lingua Franca song of 1521: after L. P. Harvey and others, ‘Lingua Franca in a villancico by Encina’ in *Revue de littérature comparée* vol. 41 (1967) pp. 572–9

# ROMANI

**POSSIBLY 3,000,000 OR MORE SPEAKERS**

**T**here seem to have been three separate westward emigrations of nomadic peoples from India in medieval times. Through the centuries, the travelling musicians and dancers belonging to these peoples have had a powerful influence on the poetry and art of Europe and western Asia.

From one of these peoples descend the Lomavren, nomadic groups of Azerbaijan and Armenia, who once spoke a quite distinct language. They now speak ARMENIAN in a special dialect that retains many Indo-Aryan words.

From another came the Domari (Nawari or Nuri), the still numerous Gypsies of Syria, Lebanon, Israel, Egypt and other Middle Eastern countries. The Karachi, a related group, live in Iran and central Asia. The Miturp are a smaller group living in south-eastern Turkey.

The third migration is that of the Roma, with whose language this article is concerned.

The *Roma*'s name for themselves dates from the period during which they lived in the Byzantine or 'Roman' Empire. They reached Europe from the East at a time when awareness was growing of the Islamic challenge to European Christendom: hence the tendency to confuse them with Muslims and to call them 'Egyptians' (English *Gypsies*), 'Tatars', 'Saracens'. They have been called 'Bohemians', too, because Bohemia was an important Romani centre from the 15th century to the mid 20th. The oldest recorded Greek name for the Roma is *athinganoi*, 'avoiding the touch', a reminder of their rules of ritual purity.

The term *Vlach*, used for one major Romani dialect group, can cause confusion. It has also been used in the past for the Romanian and Aromanian languages and their speakers.

Like Lomavren and Domari, Romani is one of the INDO-ARYAN LANGUAGES and it seems to be most closely related to PANJABI. It is far from clear when the earliest speakers of Romani left northern India. Dates as early as the 5th century and the 8th century have been suggested. Their westward migration had reached south-eastern Europe by the 13th century, perhaps earlier.

The Balkans and Transylvania are major centres of Romani population, and Romani shared in some of the phenomena of language convergence that characterise Balkan languages (see ALBANIAN). Romani dialects divide easily into Vlach and non-Vlach. The ancestors of the Vlach dialect speakers underwent four hundred years of serfdom in what is now Romania, and after the 1860s spread across the southern Balkans and across the world. Their language is naturally strongly influenced by ROMANIAN and has in turn influenced Romanian. Non-Vlach dialects descend from the Romani who bypassed Romania and reached central Europe by way of Serbia and Croatia from the early 15th century on, spreading from there. In these dialects there are no Hungarian and few Romanian loanwords.

Roma were subject to execution in Britain in the 17th century. Well over a million were killed under German rule between 1933 and 1945, though German satellite states, such as Slovakia, did not take part in this genocide. Later, in communist eastern Europe, many Roma were forcibly settled in villages to obliterate their traditional way of life.

Like the Indian peoples to whom they are related, Roma subdivide into stable castes with differing traditions, trades and dialects. Most of these groups retain a nomadic way of life, though others settle in towns and villages. The Roma population of many countries is thus complicated in its make-up and difficult to map.

## Romani as a standard language

The usual external attitudes to Romani language and culture have been lack of comprehension alternating with dangerous hostility. However, in a few countries where Romani speakers are still to be found in quite large numbers, the language is beginning to be recognised as a medium for education and literature.

In Macedonia it has constitutional status and is taught in some schools. Work has been done there on the standardisation of language and spelling, based on the majority Arlija dialect with some modifications to accommodate Džambaz.

In former Czechoslovakia, too, a standard written form was developed after the liberalisation of 1969. This is based on the Carpathian-type dialect of the majority Slovak Roma, and is used (particularly in Slovakia) in poetry, plays and journalism.

## Offshoots of Romani

There is a big difference between the dialects that still have a recognisably Romani structure, and those that have adopted the grammar of the surrounding language while still using some Romani words. Linguists call these para-Romani languages.

Roma communities may come to share part of their lifestyle, and thus need to communicate, with other elements: in Britain, for example, they have interacted with local outlaws and 'drop-outs', and with Irish, Scottish and Welsh

travellers. An Anglo-Romani pidgin may have first developed, as long ago as the 16th century, to allow communication among these groups while maintaining the privacy associated with the traditional Romani culture, with its long-inherited rules of ritual purity. In Britain Romani itself has now almost died as a mother tongue, but Anglo-Romani, learnt by boys at the age of nine or ten as they begin to work with their fathers, is still much used. There may be 90,000 speakers in Britain and twice that number in the United States and Australia.

In a similar way, para-Romani languages are known that are based on the grammars of German, Catalan, Portuguese, Basque, Spanish, Greek, Swedish and Norwegian.

Anglo-Romani (the linguists' term) is called by its speakers *Pogadi-jib* 'broken language', *Posh 'n' posh* 'half and half' and other similar names, showing clearly its origin as a mixture of Romani and English. The Spanish equivalent, *Caló*, gets its name from the Romani word *kalo* 'black' because of its speakers' typical complexion: it is also called *Romano* (a mixed Romani-Spanish word).

*Caló* is a mixture of the extinct Romani dialect of Spain (apparently closely related to Balkan and north European Romani) and of the Andalusian dialect of Spanish. All speakers of *Caló* are bilingual, but even when their Spanish is a quite different dialect, their *Caló* has the same Andalusian features. *Caló* has a history of at least two

## Domari: language of Middle Eastern Gypsies

The little-known language of the Gypsies of the Levant belongs, like Romani, to the Indo-Aryan group. Scholars dispute whether Domari and Romani had the same early history – whether they result from the same migration out of India, or from two migrations at different times. This is the beginning of a fairy tale in Domari:

Asta yikaki tilla-tmalik, potres des u shtar u shtarne. Qolde goresan u minde khalesan u gare.  
Laherde eqasri atre ghulaskaki, u dires ghulaski das u shtar u shtarne . . .

There was one king who had eighteen sons. They mounted their horses and went away.  
They saw a castle that belonged to a ghoul, and the ghoul had eighteen daughters . . .

R. A. Stewart Macalister, *The language of the Nawar or Zutt, the nomad smiths of Palestine* (London: Bernard Quaritch, 1914)

hundred years, and probably much longer, but most of the half-million Spanish Gypsies now use it less and less. It is also spoken – still with Andalusian Spanish features – in Portugal and in some South American countries.

Various forms of Turkish slang, argot and secret languages incorporate Romani loan-

words. Clearly there was once an Anatolian dialect of Romani: it exists no longer, though, as a result of recent migrations, some speakers of Balkan Romani dialects now live in Anatolia. The Gypsy language *Boşa*, spoken in eastern Turkey near Lake Van, is a variety of ARMENIAN.

### The secrets of Roma

Romani and para-Romani languages have a special advantage to their speakers: most of those with whom they do business will never learn to understand them. Even when Romani adopts the grammar of neighbouring languages, it maintains a distinct, and thus secret, vocabulary.

**Bobaní:** Why is La Habana, capital of Cuba, called *Bobaní* in Caló, the language of the Spanish Romani? The cant name comes from a simple pun. In the original name *Haba*, which happens to be the Spanish for 'bean', is followed by *-n-* and the Spanish feminine ending *-a*. Look for equivalents in Caló and you will find *bobi* 'bean' + *-n-* + *-í* the feminine ending, hence *bobaní*.

After Peter Bakker in *Romani in contact: the history, structure and sociology of a language* ed. Yaron Matras (Amsterdam: Benjamins, 1995) p. 133

### Roma and the wider world

**Gorgio:** *Gadžó*, feminine *gadží*, is the Romani term for a non-Gypsy. The Roma recognise their own distinctiveness with the saying *Rom Romeha, gadžo, gadžeha* 'Rom goes with Rom, foreigner with foreigner'. This word has become known in several languages. In English it is usually spelt *gorgio*. In Bulgarian *gádže* now means 'boy friend, girl friend, lover'. In colloquial Romanian *gagiu* means 'man, he-man' and the feminine *gagică* means 'girl friend, pretty girl'. In the Romanian translation of the Popeye cartoons, *Gagica* is the name of Popeye's girl friend Olive Oyl. The words also have a pejorative meaning: *gagiu* 'gigolo', *gagică* 'easy lay'.

After Corinna Leschber in *Romani in contact: the history, structure and sociology of a language* ed. Yaron Matras (Amsterdam: Benjamins, 1995) p. 162

# ROMANIAN

**24,000,000 SPEAKERS**

*Romania, Moldova, Ukraine*

One of the ROMANCE LANGUAGES descended from Latin, Romanian is the majority language of Romania and of Moldova, and is spoken by minorities in Ukraine, the United States, Australia and other countries. Its closest relatives are AROMUNIAN and Italian.

Romanian is known to its own speakers as *limba română*. Until recently the spellings *Roumanian*, *Rumanian* were standard. Traditionally, speakers of Romanian were known to their neighbours as Vlachs (Hungarian *Olah*, Greek *Vlákhos*), hence the name of *Wallachia*: the Romanians themselves call this province *Țara Românească*, 'Romanian land'. For ideological reasons Romanian was called *limba moldovenească*, 'Moldavian', in Soviet Moldova.

Romanian was the last of the major Romance languages to be recorded in writing, in documents and religious texts of the mid 16th century.

Where were the speakers of Romanian in medieval times? Hungarians and Romanians dispute this endlessly. Did Latin speakers maintain their language in the mineral-rich mountains of Transylvania, after the Roman legions retreated from there in AD 271? Or did Latin speakers from further south – modern Yugoslavia and Bulgaria – gradually migrate northwards in medieval times? In that case, might the Hungarians have reached Transylvania before the future Romanians did? Both sides have relied on the thousand-year-old puzzle in their arguments about who should rule modern Transylvania.

Three Romanian-speaking provinces emerge from the mists of medieval Balkan history: Wallachia, the lower Danube plains; Moldavia, reaching north from the Danube delta towards Ukraine and Poland; and, tucked in the fold of the Carpathians, Transylvania (known to Germans as *Siebenbürgen*, the seven castles, and to Hungarians as *Erdély*, the woodland). For centuries these provinces were divided by politics. Moldavia and Wallachia were Christian principalities subject to the Turkish Empire; Transylvania was generally under Hungarian and Austrian rule, and still has large Hungarian and German minorities. Fleetinglly united under Michael the Brave in 1599–1601, the three provinces were brought together again in 1918. Eastern Moldavia was seized by the Soviet Union in 1944: as Moldova it is now an independent country.

## Romanian and its neighbours

Romanian has in some ways remained especially close to Latin. It is the only Romance language that has a 'case' distinction in nouns, formed with the help of the definite article (Latin *ille*, 'that') which is now suffixed to the noun: *drac*, 'devil' (earlier 'snake'); *dracul*, 'the Devil'; *dracului*, 'of the Devil'.

Romanian also shows very strong structural influence from an early form of Slavonic – as if, at the time of the Slav invasions of the Balkans, large numbers of Slavonic speakers rapidly adopted the then standard language of the region. The numbers '1' to '10' are purely Latin in origin: *unu*, *doi*, *trei*, *patru*, *cinci*, *șase*, *șapte*, *opt*, *noi*, *zece*. Numbers from eleven on-

wards are made up of Latin roots arranged in the Slavonic way. Hyphens are inserted in these examples to clarify the structure: *un-spre-zece* 'eleven', literally 'one on ten', just like Russian *odin-na-desat'*; *trei-zece* 'thirty', literally 'three tens', like Russian *tri-desat'*.

Old Slavonic has continued to exert a dominating effect on Romanian as the traditional language of the Orthodox Church. More recently French has had intellectual prestige. The basic vocabulary remains Romance, but with many shifts of meaning: *bărbat* 'man' (earlier 'bearded'); *inimă* 'heart' (earlier 'spirit'); *pământ* 'world' (earlier 'pavement'). There are numerous loanwords from Old Slavonic, German, modern Greek, Turkish and French: *izvor* 'spring'; *halbă* 'beer mug'; *drum* 'road'; *tutun* 'tobacco'; *bagaj* 'luggage'. Colloquial Romanian borrows liberally from ROMANI, though these loanwords tend to be left out of dictionaries.

A sequence of historical chronicles comes out of the earliest period of Romanian writing. But this was a multilingual culture, and Moldavia's greatest 18th-century writer, Prince Dimitrie Cantemir, wrote in Latin, German and Turkish as well as Romanian. His monumental *History of the Growth and Decline of the Ottoman Empire* was first published (in 1734) in English translation. Folk song in Romania has drawn vigour from its varied origins: Gypsies have been among its chief exponents. The French composer Berlioz was inspired by it, as were native poets such as Vasile Alecsandri.

Romanian was first written in a local variant of the old Cyrillic alphabet, as used for Old Slavonic. The Latin alphabet was adopted in the 19th century. As elsewhere in the Soviet Union, a modern Cyrillic alphabet was used in Moldova between 1944 and 1991. Independent Moldova now refers to its language as 'Romanian' and uses the Latin alphabet.

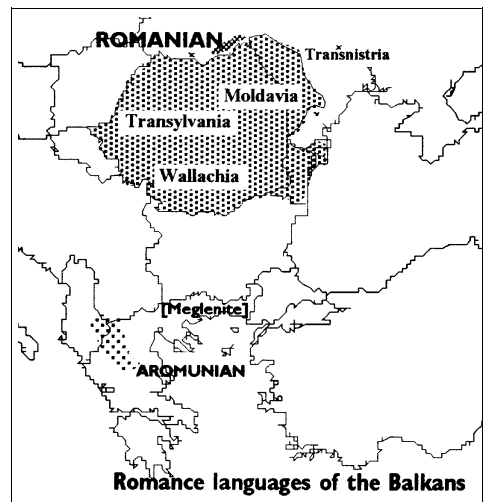
## Romance languages of the Balkans

Romanian and its relatives have interesting features in common with other Balkan languages (see also ALBANIAN).

*Romanian* is the language of the three contiguous provinces, Wallachia, Moldavia (including Bessarabia, now Moldova) and Transylvania. To the east of Moldavia, a Romanian-speaking minority on the left bank of the Dniester formed the basis of a Moldavian republic, part of Ukraine, in early Soviet times. When the Soviet Union annexed Bessarabia in 1940, the newly created Moldavian Soviet Republic (independent Moldova from 1991) incorporated part of this early Moldavian SSR – the district now known as Transnistria. Romanian speakers are the largest single population group there, but they make up only about 40 per cent of the total population, the remainder being Ukrainian and Russian speakers.

AROMUNIAN is spoken in the neighbourhood of lakes Ohrid and Prespa, in north-western Greece, eastern Albania and south-western Macedonia.

*Meglenite* is the language of a community of *Vlashi* who have lived in a remote mountainous district of Greece north of Saloniki. They had become Muslims in the 18th century. As such they were liable to expulsion from Greece after 1923, and most of them settled in Turkey, where their language may now be extinct. A few are said to remain in Greece.



# ROMANSCH

**65,000 SPEAKERS**

*Switzerland*

One of the less known of the ROMANCE LANGUAGES, Romansch was recognised in 1938 as the fourth national language of Switzerland, alongside German, French and Italian.

Romansch is spoken in the canton of *Grisons* or *Graubünden*, 'the Grey League', which preserves the name of a self-defence organisation of Romance speakers set up in the 15th century. It became part of Switzerland in 1803. Germans once called this language *Chur-Wälsch*, '“Welsh” or foreign speech of Chur', for Chur was the capital of Roman Rhaetia and was once the centre of Romansch. Chur, even its cross-river suburb *Wälschdörfli* '“Welsh” village', now speaks German: Romansch survives only in the upper valleys of the Rhine and the Inn.

Its origin is as a dialect of the provincial Latin of the central Alps, which were incorporated in the Roman Empire during the reign of Augustus. Before Roman times the area appears to have been Celtic-speaking; by the end of the Roman Empire there must have been an unbroken region of distinctive Romance speech here, gradually driven into the high valleys by the encroachment of German from the north and of Italian from the south.

Romansch itself, with its long history of local linguistic pride and a high rate of literacy, has so far survived and even prospered. The Engadine dialect was first printed as early as 1552 in Jacob Bifrun's *Christiauna fuorma*, a catechism; a translation of the New Testament followed in 1560. Both Romansch and Engadine, in several varieties, are used as school languages and in the local media.

Two other 'Rhaeto-Romance' languages, Ladin and Friulian, survive in the Italian Alps.

Ladin is spoken in four valleys radiating from the mountain core of the Dolomites: those of the Avisio and Noce, traditionally Venetian territory, and those of the Gader and Gardena, which were Austrian until 1919. Ladin has some use in primary schools and is locally recognised in the region of Trentino-Alto Adige.

Friulian covers a much larger territory and has as many as 600,000 speakers. Trieste was a Friulian-speaking city in medieval times, but Venetian long ago replaced Friulian as the everyday speech of the head of the Adriatic. In spite of its large number of speakers Friulian has no official status nationally or regionally: it has been historically overshadowed by Venetian and is now in thrall to Italian. Yet it remains an everyday medium of communication, and many local German and Slovene speakers are bilingual in Friulian. It has some literature, too: Pier Paolo Pasolini's first book of poetry was in Friulian, and he founded the now defunct *Academiuta di Lenga Furlana* 'Academy of the Friulian Language'.

## The Rhaeto-Romance languages on the map

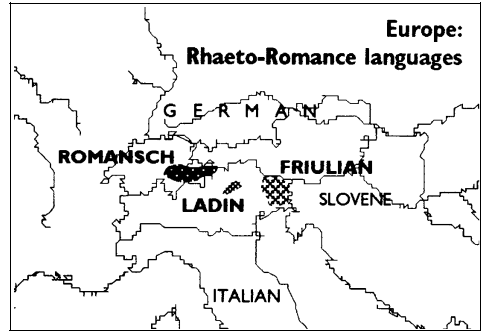
*Romansch* is the language of the upper Rhine valley, spoken by a largely Catholic population. The *Sursekan* 'above the woods' dialect, centring on Disentis, is the standard form. It is divided by the forest of Flims from *Subsekan*.



**Rhaeto-Romance numerals**

<b>Surselvan</b>		<b>Friulian</b>
in, ina	1	un, une
dus, duas	2	doi, dos
treis	3	tre
quater	4	cuatri
tschun	5	cinc
sis	6	sis
siat	7	siet
otg	8	vot
nov	9	nuv
diesch	10	dis

*Engadine*, close to Romansch, is spoken in the upper Inn by a Protestant population. The chief town here is Samaden.



*Ladin* is spoken in four valleys of the Dolomites. It may have about 30,000 speakers.

*Friulian* is the local language of the Italian autonomous region of Friuli-Venezia Giulia. There are over half a million speakers there, and a few in neighbouring districts of Slovenia.

**Fair Catine**

Indulà vastu, Catine biele?  
 O voi a moris de barâz, o missâr pari.  
 Mostrimi lis moris di barâz, Catine biele –  
 La ciavra lis à mangiadis, o missâr pari.  
 Mostrimi la ciavra, Catine biele –  
 Il beciâr la à copade, o missâr pari.  
 Mostrimi il beciâr, Catine biele –  
 Il beciâr al è sotiara, o missâr pari.  
 Mostrimi la tiara, Catine biele –  
 La nêf la à taponada, o missâr pari . . .

Where are you going, fair Catine?  
 I'm going blackberrying, father.  
 Show me the blackberries, fair Catine –  
 The goat ate them, father.  
 Show me the goat, fair Catine –  
 The butcher killed it, father.  
 Show me the butcher, fair Catine –  
 He's under the earth, father.  
 Show me the earth, fair Catine –  
 The snow has covered it, father . . .

After D. B. Gregor, *Friulan: language and literature*  
 (New York: Oleander Press, 1975) pp. 284–5

*Arâz* 'bramble' is one of the local words that seems to descend from an early Celtic language, cf. Irish and Breton *barr* 'branch'.

# RUNDI, RWANDA AND HA

**12,000,000 SPEAKERS**

*Burundi, Rwanda, Tanzania, Congo (Kinshasa) and Uganda*

**T**hese, with two smaller dialect groups, belong to the BANTU LANGUAGES: they make up the Southern Lacustrine group. Speakers of any one of the five can easily understand the others, so it is logical to treat them as one language here.

The full local names of these languages are *kinyaRwanda* 'language of Rwanda', *kiRundi* and *kiHa* or *giHa*; the traditional states were *Rwanda*, *buRundi* and *buHa*.

Burundi and Rwanda, in spite of fearsome ethnic divisions, have no language problem or linguistic minorities. Like neighbouring peoples to the east (see NKORE), they consist of political units in which two main castes have existed side by side – the tall, cattle-owning Tutsi, historically supreme, and the peasant Hutu. Below these again are the Twa, hunters and gatherers, potters and musicians. It is said that the dialect of the Twa is easily distinguishable from that of the other two castes.

Early Bantu speakers, apparently represented by the modern Hutu, have probably occupied this region for over two thousand years. The arrival of the Tutsi from the north-east is an event of a few hundred years ago, according to oral tradition. They were probably speakers of a NILO-SAHARAN language (not 'Hamites' or Afro-asiatic language speakers, as anthropologists used to think). Whatever they spoke on their arrival, they soon adopted the Bantu language of the ancestors of the Hutu.

The three kingdoms were annexed by Germany at the end of the 19th century and occu-

pied by Belgian troops in 1916. Buha was acquired by Britain in 1921 and is now part of Tanzania. Rwanda and Burundi remained under Belgian rule until 1962 when they regained their status as independent states.

Rwanda oral literature was composed partly by the Hutu, whose speciality was court literature: historical prose, dynastic poems, poems in praise of cattle, and warrior poems of self-praise. This literature, carefully transmitted by admired and skilled performers, is in a complex and archaic language – for full understanding of the verse, prose commentaries are said to be needed. Rundi, too, has its poetry in praise of one's ruler, one's cows and oneself. Popular poetry, tales and songs, some traditionally performed by Twa singers and clowns, included songs to console a newly married girl.

Rundi and Rwanda are now the official languages, alongside French, of Burundi and Rwanda. Although knowledge of Rwanda or Rundi by itself ensures communication over a wide area, Swahili, the regional language of East Africa, remains important in trade and everyday life. French, the language of the Belgian protectorate which governed 'Ruanda-Urundi' until 1962, is taught in secondary schools and is still a medium of culture in the two independent countries. There are French loanwords in Rwanda and Rundi, such as *zuzi* from French *juge* 'judge'.

Rwanda has a distinction of vowel length, four tones and seven distinct sentence-level intonation patterns. For a table of numerals see GANDA.

### Rundi, Rwanda and their relatives

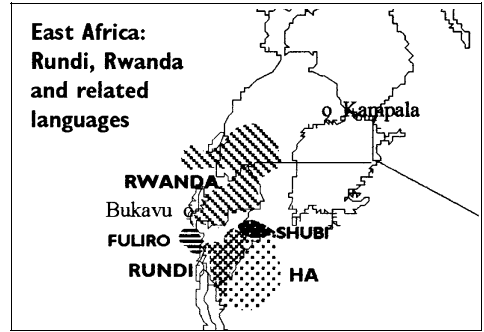
*Ha*, with 725,000 speakers, is a language of Kigoma Province, Tanzania. Here the two castes are Tusi and Ha.

*Rundi* has about 5,000,000 speakers and is the national language of Burundi. It extends into the district of Bugufi in Tanzania.

*Rwanda* or Runyarwanda has rather over 6,000,000 speakers. It is spoken by both sides in the Rwandan fighting, and by minorities in Uganda (Bufumbira in the Kigezi district) and Congo (Kinshasa). Recently a huge number of Rwanda-speaking refugees have fled to Congo.

*Shubi* has 150,000 speakers.

*Fuliru*, with 275,000 speakers in Congo (Kin-



shasa), was classed by Guthrie alongside Rundi, Rwanda and Ha; other linguists separate them.

These languages form a single dialect continuum with NKORE and its relatives.

# RUSSIAN

**PERHAPS 175,000,000 SPEAKERS**

*Russia, Ukraine, Kazakhstan, Belarus,  
Estonia and other Soviet successor states*

**O**ne of the Eastern SLAVONIC LANGUAGES, Russian was spoken in the northern marches of the expanding Slavonic-speaking territory in the first millennium AD.

*Rus'* was once the name of the principality ruled from Kiev – modern Ukraine. In medieval times the cultural and political centre of gravity shifted to Moscow, and thus the language of Moscow came to be called *Russian*.

The crystallisation of Russian as a separate language can be traced to the Mongolian-Tatar conquests of the mid 13th century and their destruction of Kiev in 1242. This was soon followed by Lithuanian and Polish annexation of Ukraine. Kiev, capital of Ukraine, had previously given religious and political leadership. Constantinople, metropolis of Orthodox Christianity, itself faltered and was to fall.

Moscow had once been a northern outpost, and was still to be subject to the Golden Horde of the Mongols and Tatars for another two and a half centuries. For all that, Moscow now came to be seen as the 'Third Rome', inheritor of the spiritual mantle of the Roman and Byzantine empires. Russian missionaries were eventually to spread Christianity to many of the peoples of greater Russia.

The language of this religious flowering was the Ukrainian-Russian variety of OLD SLAVONIC. Spoken Russian of the period – *Middle Russian* – shines through it in the adjustments that writers made, consciously or unconsciously, in the old sacred language, and in the everyday words that were introduced into documents and historical writings in discussing everyday matters.

Spoken Russian became, for the first time, a

fully written language with the reforms of Peter the Great (1672–1725). Not only did he introduce a revised alphabet (see box): he encouraged authors to write as they spoke. Mikhail Lomonosov (1711–65) reasserted the suitability of the transitional Moscow dialect as the Russian standard (though the Court was then at St Petersburg). It is this historic Moscow accent that contributes to the spoken Russian of today some of its characteristic sound profile, notably the pronunciation of unstressed *o* as *a*.

The first great period of Russian literature was the 19th century – the period of Tolstoi's and Dostoyevskii's novels, of Pushkin's limpid verse, and of some great collections of folk poetry and folk tales. The Soviet period was both good and bad for Russian language and literature: good because literacy is far more widespread and Russian is extensively known as a second language; bad in that publication was restricted and ideologically steered, more effectively than under the monarchy.

In spite of modern developments, Russian retains – by comparison with Ukrainian – a great number of loanwords from Church Slavonic, inherited from the centuries during which this was the only written language of Russia. Examples of other loanwords are *zhemchug* 'pearl' from Tatar; *rynok* 'market' from Polish; masculine *sportsmen*, feminine *sportsmenka* 'athlete' from English. Russian also forms new words by calquing: *neboskryob* has the same meaning, and the same structure, as English 'skyscraper'.

Russian is generally thought one of the more difficult of Indo-European languages for an English learner. Word stress is heavy and affects the pronunciation of surrounding sounds, but it is

not easily predictable and not marked in the script. In noun and adjective declension there are three genders, six cases and two numbers. Verbs generally have two 'aspects', perfective and imperfective (a frequentative aspect is also found). The basic forms of these cannot be predicted from one another by any simple rule, and a range of forms to mark tense, mood, number and person is derived from each.

Russian is the source of some English words, including *samovar* and *cosmonaut*.

### The Russian diaspora

After throwing off Tatar domination, Russian expansion from the 15th century first encompassed Kazan and Astrakhan (two of the three khanates once owing allegiance to the Golden Horde), then Siberia, where trading ventures were followed by Russian administration and Christian missionaries. In the north Russian influence moved outwards from the seventeen historic northern cities that were early centres of settled life: Apatiti, Bilibino, Vorkuta, Vuktyl, Dudinka, Igarka, Inta, Kovdor, Murmansk, Nickel, Nadym, Norilsk, Salekhard, Severomorsk, Urengoy, Usinsk, Yakutsk. Even Alaska was in the Russian sphere until 1860, and Russian was becoming a lingua franca for Amerindians there.

Meanwhile there came expansion into Karelia, Belarus and Ukraine, then the Crimea and the Caucasus, which Russia struggled to subjugate in a 'Hundred Years War' from 1760 to 1860. Finally, in the 19th century, central Asia was annexed. All these regions submitted to Russian settlement, Russian-led administration and education, and the increasing dominance of Russian in public life.

In many regions, autonomous armed Cossack settlements (see KAZAKH) marked the front line of Russian expansion. Cossacks are now once again a significant political force, with a measure of self-government. Mass migration and resettlement, including that of minority peoples, spread the use of Russian as the only available lingua franca (see also box at AZERI). Russian political exiles and criminals have, for generations, been dispatched to prisons and labour camps in northern European Russia, Siberia

and central Asia: if they survived incarceration, they sometimes married and settled locally. The social and political outcasts of Russia swelled the numbers of the Russian diaspora, creating fearful problems for the Soviet Union's successor states.

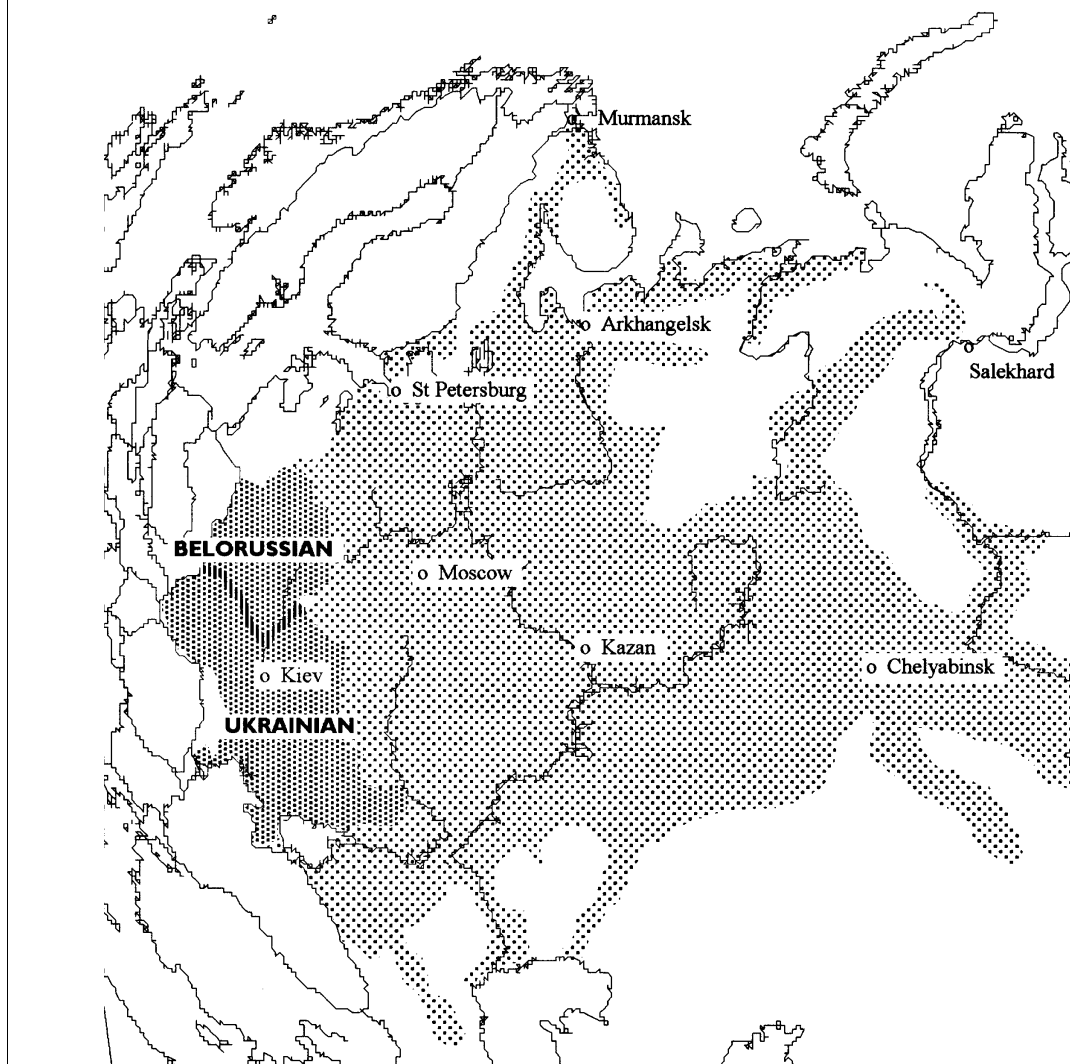
During the 19th century Russia also encompassed Finland, Estonia, Latvia, Lithuania, Bessarabia (Moldova) and a large part of Poland. At the 1918 Revolution these territories were lost. Of the minority language regions of the old Russian Empire that remained, the larger ones became constituent republics of the new Soviet Union. Russians continued to settle in these regions in ever larger numbers. Meanwhile Tuva and Mongolia became Soviet satellites. In the course of the Second World War the Soviet Union annexed Tuva and reconquered Estonia, Latvia, Lithuania and Bessarabia along with the easternmost provinces of Poland, Czechoslovakia and Hungary.

At the end of the war the German population was expelled from East Prussia, which was divided between Poland and the Soviet Union. The Soviet sector, centred on Kaliningrad (formerly Königsberg), was repopulated mostly with Russians, who remain there in a heavily armed enclave now bordered by Poland and independent Lithuania.

Poland, East Germany, Czechoslovakia, Hungary, Bulgaria and Cuba were Soviet satellite states from the end of the war until the collapse under Gorbachev. East Germany was especially heavily garrisoned, and many Russian military families settled there. In the same period Soviet influence over Yugoslavia, Albania, Romania, China, North Vietnam and North Korea was significant though less consistent. Several other countries, including Angola, Mozambique, Libya, Ethiopia and Afghanistan, experienced heavy Russian economic and military activity. Russian soldiers, technical advisers and labourers lived in some of these countries in considerable numbers and for long periods, and their own peoples went to work and study in Russia.

Russian has influenced the language of all these countries, notably the vocabularies of science, warfare, education and politics. For the present, Russian is much used in international communication among former Soviet countries, where it

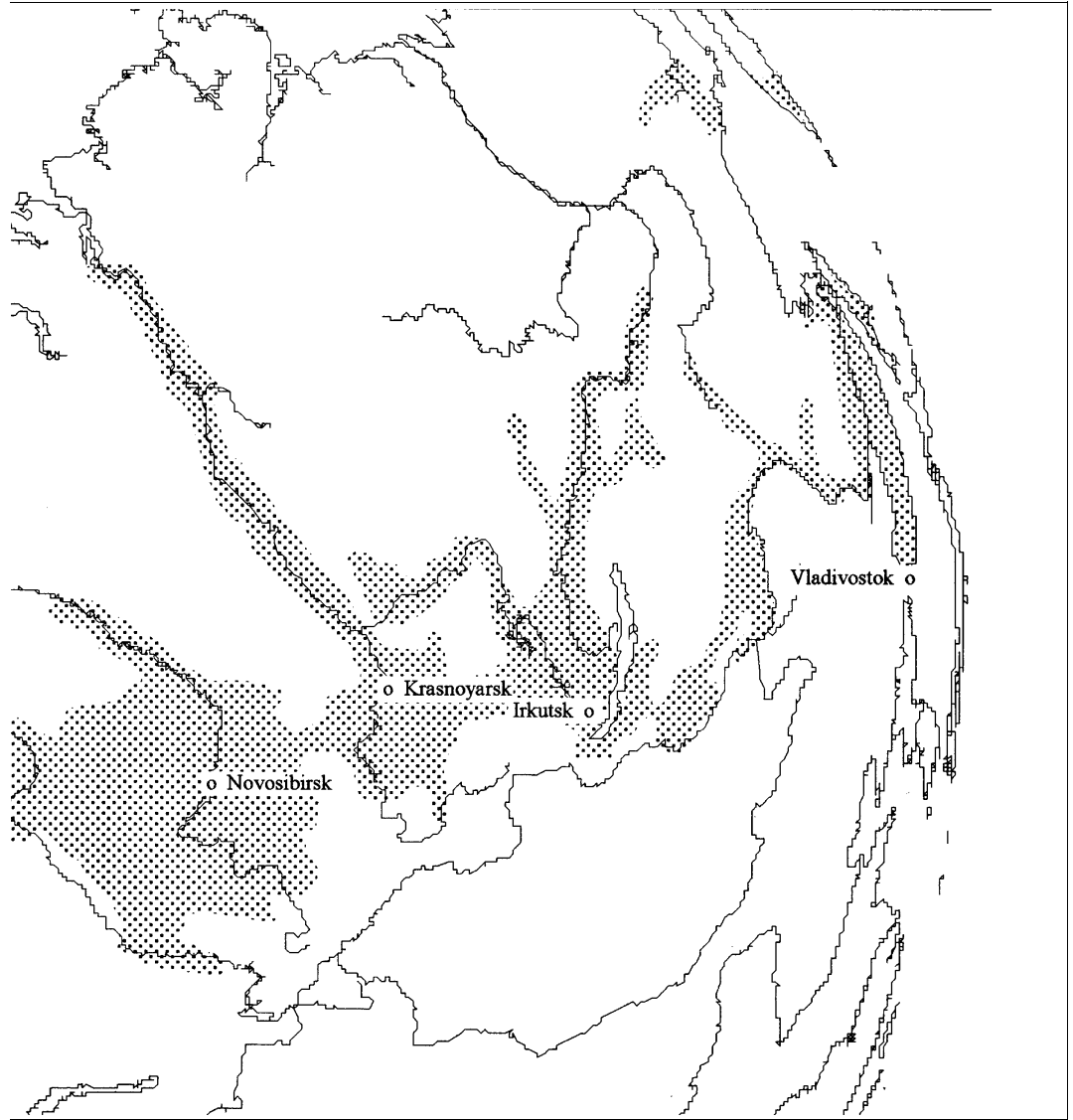
## Russian and the Eastern Slavonic languages



is at least a second language for millions of people. But almost everywhere this position is being eroded, generally in favour of English.

For, in the last decade, the Soviet Union has disintegrated. Led by Lithuania, all the Soviet republics have declared their independence. Self-government has been granted to numerous minority peoples that had not had Soviet republic status: even some of these, such as the Tatars and the Chechens, are pushing more or less violently for full independence.

This has left some very large minorities of Russian speakers, once confident of their dominant status, now cut off from power and struggling for equality. Russia's television service still broadcasts to these minorities, and a local Russian press exists in nearly all the former Soviet countries. From some territories, including Chechnya and Tuva, many Russian settlers have fled; in Azerbaijan, Uzbekistan, Tajikistan, Turkmenistan, Kazakhstan and Kyrgyzstan, a decline in the Russian population already noticeable in the



1970s has gradually speeded up as Russian speakers face unpopularity and discrimination. The very large Russian minority settled in northern Kazakhstan is the most secure. In the west there are powerful Russian minorities in Estonia, Latvia, Belarus and Ukraine.

Russian émigré communities, already sizeable in the 19th century, grew markedly after the 1917 Revolution and during the troubled 1930s and 1940s. There has been an important Russian émigré press in Paris, London, New York and

other cities. Much of the best Russian literature of the 20th century was published abroad, in a language that was becoming somewhat old-fashioned, to find few Russian readers – because ‘subversive’ literature could not be imported into the Soviet Union.

Some Northern dialects of Russian have the unusual feature of a suffixed definite article, like Bulgarian and other Balkan languages: *dom* ‘house’, *domot* ‘the house’.

## The Eastern Slavonic languages

The three languages form a dialect continuum. In particular, the North-eastern dialects of *Belorussian* resemble neighbouring Russian dialects; the South-western dialects resemble northern Ukrainian.

Standard *Ukrainian* is based on the Eastern or Left-Bank dialects of Kiev and Poltava. The Dnieper makes a clear boundary between these and the Western dialects, which have some resemblance to Polish. Speakers of the dialect of Transcarpathian Ukraine, which until 1944 belonged to Czechoslovakia, sometimes now claim for it the status of a separate language, *Transcarpathian Ruthenian*.

The Northern and Southern dialect groups of *Russian* are separated by a belt of transitional 'Central' dialects. Moscow's own dialect is one of these: its vowels are like those of the South, its consonants like those of the North.

### The Russian alphabet

А Б В Г Д Е Ё Ж З И Й К Л М Н О П Р С Т У  
Ф Х Ц Ч Ш Щ Ъ Ы Ь Э Ю Я

а б в г д е ё ж з и й к л м н о п р с т у ф х ц ч щ ъ  
ы э ю я

a b v g d y e y o z h z i i k l m n o p r s t u f k h t s c h  
sh shch " i ' e y u ya

ы, transliterated *i* in this book, usually appears as *y* when Russian words and names are included in English texts. This *grazhdanka* or secular alphabet was introduced by Peter the Great in 1708 to take the place of the Old Slavonic form of the Cyrillic script. Four more letters initially carried over from Old Slavonic – IVЪӨ *ivъө i i ye f* – were eliminated in the reform of 1917/18.

### The tale of Igor's campaign

Чи ли въспѣти было, вѣщей Бояне,  
Велесовъ внуче:  
'Комони пжуть за Сулою, –  
звенить слава въ Києве;  
трбы трубять въ Новѣградѣ, –  
стоять стязи въ Путивлѣ.'

Or you might have begun your song thus,  
wizard Boyan, grandson of Veles:  
'Horses neigh beyond the Sula,  
glory rings out in Kiev;  
trumpets sound in Novgorod,  
standards are raised in Putivl'!

From *The tale of Igor's campaign*

*Slovo o polku Igoreve* is the story of young Igor's disastrous campaign in the year 1185 – undertaken without the permission of his sovereign, the Prince of Kiev – against the wild Polovtsy ('Cumans'). It is the tale of his defeat, captivity and escape. More than that, this medieval narrative in poetic prose is 'a moving record of the last decades of Kievan Russia, on the eve of the Mongol invasion, and of her struggle to survive in an age of chivalry and disaster, of heroism and folly, of civil strife and barbarian invasions' (Dimitri Obolensky, *The Penguin book of Russian verse* (Harmondsworth, 1962) p. xxxiii). The only known manuscript of the tale was discovered around 1790 and was destroyed in the fire of Moscow in 1812, but not before an edition had been printed. Borodin's opera *Prince Igor* is based on the story.

The *byliny*, ballads collected in recent times from peasant singers in the Russian countryside, seem to recreate essentially the same world, though in them the enemy is usually (anachronistically) the Tatars. They centre on Kiev (and 'Prince Vladimir') or on the trading city of Novgorod. One of the Novgorod cycle is the story of Sadko, the minstrel who became a rich merchant.



### The first politically correct language?

Official, party-approved Russian was a dull language, spoken with the 'wooden tongue' that Stalin recommended. Concepts and institutions had to be named in such a way as to enjoin a correct point of view. This produced long, boring phrases which (because life is too short) had to be abridged in everyday use.

Official form	Everyday form	Formal translation	Everyday equivalent
Vsesoyuznii Leninskii <b>Kommunisticheskii Soyuz</b> <b>Molodyozhi</b>	Komsomol	Lenin's All-Union Communist League of Young People	Scouts
<b>Soyuz Sovetskikh</b> <b>Sotsiyalisticheskikh</b> <b>Respublik</b>	SSSR; Sovetskii Soyuz	Union of Soviet Socialist Republics	USSR; Soviet Union
<b>Vyssheye Uchebnoye</b> <b>Zavedeniye</b>	Vuz	Higher Education Institution	College

# SAMI

**PERHAPS 30,000 SPEAKERS**

*Norway, Sweden, Finland*

**S**ami, often called Lapp, one of the URALIC LANGUAGES, is spoken by gradually shrinking nomadic communities who traditionally lived by reindeer herding and fishing in the inhospitable northern regions of Norway, Sweden, Finland and Karelia.

In early times the Sami were known to outsiders as *Fenni* (Latin), *Finnar* (Old Norse: see box). They were *Finni mitissimi* 'Sami, the mildest of people' to a Latin historian of the Goths. The word possibly derives from proto-Scandinavian *finna* 'to find', because Sami speakers were foragers and hunters. The more usual modern name *Lapp*, applied to them by the Finns, is first recorded in the 13th century. *Sami* or *Saami*, now the preferred term, is based on the speakers' own name for themselves, *sápm̃e* singular, *sápm̃eht* plural. It is the same name as that of one of the historic tribes of the Finns, *Häme*.

Nothing was written in Sami till Nicolaus Andreas, a Swedish priest, printed an ABC and *Missal* in 1619. But the early history of the language can still be traced – through loanwords.

By 1500 BC, speakers of 'Early Proto-Finnic', the proto-language that would eventually split into Sami, Finnish and Estonian, were in contact with speakers of BALTIC and GERMANIC LANGUAGES: thus, from Baltic dialects, Lithuanian *lāšis* 'salmon' reappears as Sami *luossâ*; Latvian *dagla* 'touchwood, tinder' as Sami *duowle*; from proto-Germanic, Anglo-Saxon *hos*, Gothic *hansa* 'troop', German *Hansa* 'guild' reappears as Sami *guosse* 'guest'. They must have been settled around the shores of the southern Baltic at that period. Finnish and Estonian continued to bor-

row from Baltic languages after 1000 BC, but Sami did not. So, by then, early Sami speakers must have been moving north, away from these other groups.

At a later time, perhaps around AD 500, they were interacting with speakers of an early form of OLD NORSE: thus Swedish *sår* 'wound' reappears as Sami *saire*, *sarje*; Old Norse *fló* 'flea' as Sami *lawkes*; Swedish *ko* 'cow' as Sami *gussâ*. It is probable that by that time Sami was spoken across its present very wide geographical range, from central Norway to Karelia and the White Sea. It began to split into dialects: thus, from about AD 800 onwards, loanwords from early forms of Norwegian, Swedish and Finnish occur in the individual Sami dialects rather than in the language as a whole.

Its close relationship with Estonian and Finnish is only half the story of Sami. Its speakers are very different, in way of life, customs, beliefs, and also physically, from most Finns. It is clear that as Sami spread across the wide north of Scandinavia it also became the language of an already existing, unrelated people, whose way of life was already similar to that of the modern Sami. They, and their now forgotten language, explain the striking differences between Sami and Finnish.

Fewer than half of the people who consider themselves Sami or Lapp still speak the language. Only about 2,500 live in northern Finland (commonly thought of as the Lapp homeland) and all of them are bilingual in Finnish. The largest number of Sami speakers, about 10,000, now live in Norway, and the Northern Sami dialect is most often used in writing. It is also called *Norwegian Lapp*, but is spoken in parts of

Sweden and Finland as well as in Norway.

There has been a succession of spelling rules, from the first Norwegian Lapp publication in 1728 to the common orthography for all three countries adopted in 1978. There are three periodicals in Sami and occasional radio broadcasts, but not much publishing except for schoolbooks and religious books.

Sami is a 'synthetic' or 'agglutinating' lan-

guage of long compound words and of complicated phonetics. It is notable for its rhythmical stress pattern. Words are stressed on the first syllable, with secondary stresses on odd syllables. Even syllables are unstressed. (For the numerals see ESTONIAN.)

Examples from Mikko Korhonen in *The Uralic languages: description, history and foreign influences* ed. Denis Sinor (Leiden: Brill, 1988)

### Empty country

An early record of the Sami – and of the Permians (see KOMI) further east – is in King Alfred's Anglo-Saxon version of Orosius' Latin *Histories*. It includes a note about a Norwegian merchant, Ohthere, who had reported to Alfred in person on the voyage eastwards from northern Norway to the White Sea:

Ac him wæs ealne weg weste land on þæt steorbord, butan fiscerum & fuglerum & huntum,  
& þæt wæron eall Finnas; & him wæs á widsæ on ðæt bæcbord. þa Beormas hæfdon swiþe  
wel gebúd hira land; ac hie ne dorston þær on cuman. Ac þara Terfinna land wæs eal weste buton  
ðær huntan gewicodon, oþþe fisceras, oþþe fugeleras . . . þa Finnas, him þuhte, & þa Beormas  
spræcon neah an geþeode.

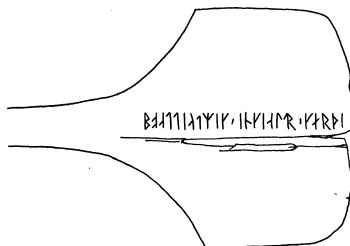
To starboard there was empty country all the way, except for fishers, birdcatchers and hunters.

They were all *Finnas* [Sami]. The *Beormas* [Permians] farmed their land well, but one dare not land there; the land of the *Terfinnas* was all empty, except where hunters or fishers or birdcatchers lived . . . The Sami and the Permians, it seemed to him [Ohthere], spoke nearly the same language.

King Alfred's Orosius ed. H. Sweet (London, 1883)

### The power of Sami

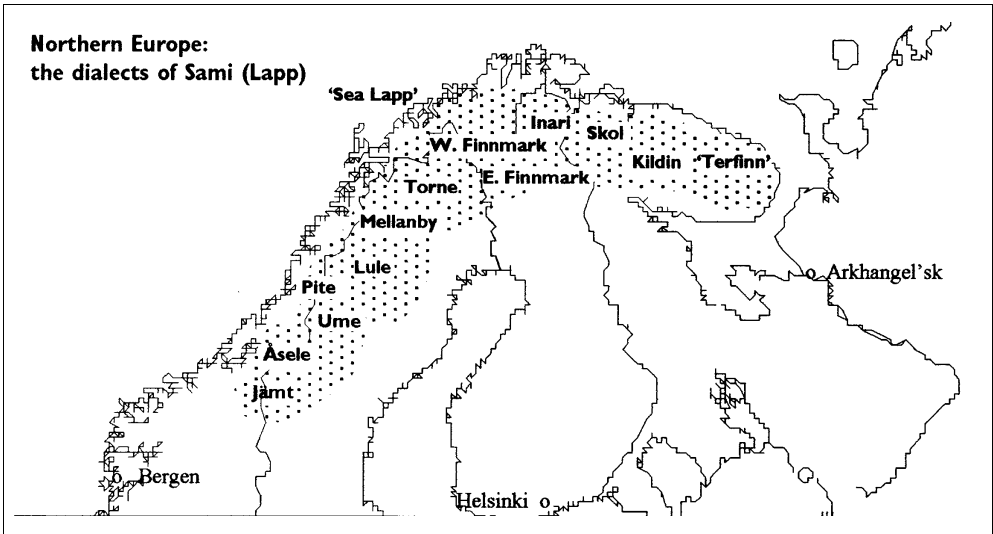
The oldest recorded text in Sami is a single word, 'Come to me' – part of a three-word message or spell (the other two words are in Old Norse) scratched in runes on a medieval wooden shovel, which was discovered in a bog at Indriðastaðir in Skorradal, Iceland, hundreds of miles of rough sea from the lands where Sami is spoken. The message – which scholars date to about 1200 – reads:



Boattiatmik Inkialtr kærþi

Come to me, Ingjald, please!

From Magnus Olsen, Knut Bergsland, *Lappisk i en islandsk runeinnskrift* (Oslo: Dybwad, 1943)  
(*Avhandlingar utgitt av Det Norske Videnskaps-Akademi i Oslo, hist.-filos. Klasse*, 1943 no. 2)



### Sami and its dialects

The line between Jämt Sami and Åsele Sami is the southern border of ancient *Finnmark*, the Sami territory that paralleled the northern Norwegian coastal realm of Halogaland. Later the Jämt and Åsele Sami were known as 'Mountain Lapps', paying tribute both to Norway, where they spent the summer, and to Sweden where they wintered in the lowland forests. The Ume Sami to their north were 'Forest Lapps', settled entirely in Sweden.

The Sami nomadic lifestyle has had to contend with such interruptions as the delimitation

of the Norway–Sweden frontier, in 1751, and the freezing of the Soviet Union's borders in the 20th century.

The dialect of the Sea Lapps, fishers settled on the far northern shores of Norway, has almost died out, replaced by Norwegian.

Based on Pekka Sammallahti, 'Die Definition von Sprachgrenzen in einem Kontinuum von Dialekten: die Lappischen Sprachen und einige Grundfragen der Dialektologie' in *Dialectologia Uralica* ed. Wolfgang Veenker (Wiesbaden: Harrassowitz, 1985) pp. 149–58 and other sources

# SAMOAN

**200,000 SPEAKERS**

*Western Samoa, American Samoa*

**S**amoan is a Polynesian language, belonging to the Oceanic branch of AUSTRONESIAN LANGUAGES. It is the language of the politically divided island group of Samoa in the western Pacific (see map at TONGAN). There are migrant communities in Hawaii, New Zealand, Australia and the United States.

Western Samoa has about 160,000 inhabitants, American Samoa about 30,000. In spite of the political frontier Samoan has no important geographical dialects. But the standard language does differ considerably from the colloquial. Colloquial Samoan is called *Tautala leaga*, 'bad language'. *Tautala lelei* 'good language' is used formally – in broadcasting, in church and school,

in writing, in poetry and in talking to Europeans. A polite register, with some hundreds of special words, is used when talking to people of high status.

*Tautala lelei* has thirteen consonants; *tautala leaga* has only ten, lacking *t n r*. There are no consonant clusters or final consonants, so 'ice cream' becomes *aisakulimi* in Samoan.

The first ten numerals in Samoan are: *tasi, lua, tolu, fā, lima, ono, fitu, valu, iʻva, sefulu*.

Based on Ulrike Mosel, 'Samoan' in *Comparative Austronesian dictionary* ed.

Darrell P. Tryon (Berlin: Mouton De Gruyter, 1995– ) pt 1 pp. 943–6 and other sources

# SAMOYEDIC LANGUAGES

**T**hese form a distinct branch of the URALIC LANGUAGES, spoken by very small numbers of forest and tundra dwellers and traders in western Siberia. The three northern Samoyedic languages are Nenets, Enets and Nganasan. Several southern Samoyedic languages are known from reports by 19th-century travellers, but only one – Selkup – survives.

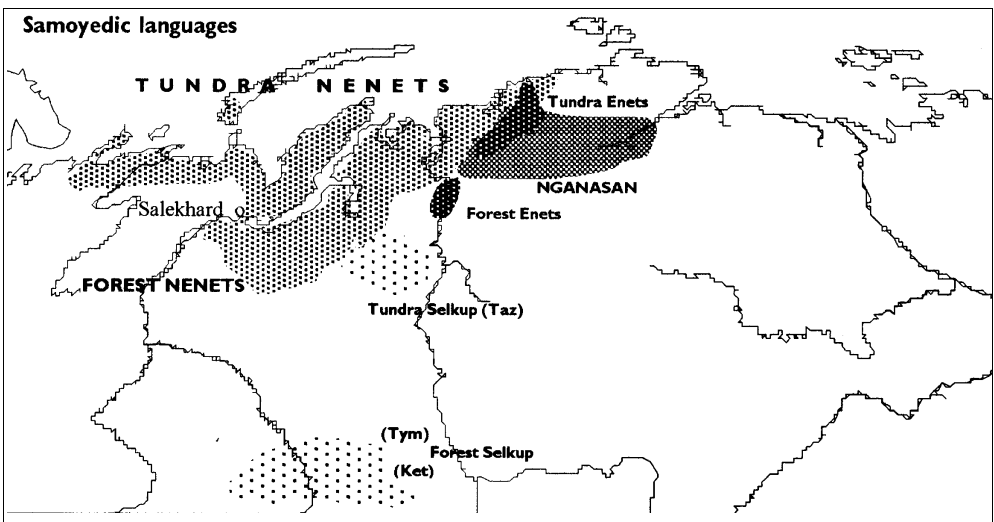
*Nenets, Enets and Nganasan* are words of the same origin: they mean 'real person' in the respective languages. *Söl'qup*, in the Selkup language, means 'earth man', a reminder, perhaps, that the Selkup used to live in underground houses excavated in steep river banks. *Yurak*, the traditional outsiders' name for Nenets, seems to be a Russian version of the Nenets word for 'friend' – which is how the Nenets would have described themselves when negotiating trading links with the Russians.

In Nenets nouns have seven cases and four declensions, known to linguists as absolute, pos-

sessive, predestinate-possessive and predicative. In the two possessive declensions, nouns are inflected for three persons and three numbers (singular, dual and plural) as well as for case. The total number of possible forms for each noun runs into hundreds: *we* 'dog', *wengkayunyi* 'my two dogs', *wengkayud* 'your two dogs' . . .

Scattered and fragmented as the Samoyedic languages and dialects are, their speakers are linguistically sophisticated. Most are multilingual: they regularly come into contact not only with other Samoyedic languages but also Khanty and Mansi (see HUNGARIAN), YAKUT and Dolgan, TUNGUSIC LANGUAGES such as Evenki, and, ever increasingly, RUSSIAN.

The most widespread language of the group is *Nenets* or *Yurak*, spoken by 27,000 people. There are two dialects. About 1,000, spread over an area of nearly fifty thousand square kilometres, speak Forest Nenets. They hunt, fish and keep small herds of reindeer. The remainder range the tundra between the White Sea and the Yenisei



delta, and speak Tundra Nenets. With their huge herds of thousands of reindeer they go north to the Arctic coast in summer, south to the forest edge in winter. The two dialects sound rather different, but most vocabulary is shared. Standard written Nenets, used in local primary schools, is based on the Tundra dialect of Bolshaya Zemlya.

*Enets* or Yenisei Samoyed, a closely related language, is spoken by an ethnic group of about 300, subdivided into four clans. Two-thirds of them, 'Tundra Enets', live as nomads on the tundra east of the Yenisei estuary. The rest, 'Forest Enets', occupy a taiga region close to the right bank of the Yenisei.

*Nganasan* or Tavgi has about 900 speakers in the vast Arctic wastes of the Taimyr Peninsula. They are traditionally hunters of wild reindeer. An eastern group, *Aja* or 'younger brothers', are former Evenki speakers whose language nowadays is a distinct form of Nganasan.

*Selkup* is properly the name for one localised northern dialect, which is the basis for the written form of Selkup (or Ostyak Samoyed) used in schools. Fishers, hunters and foragers, its speakers live on the banks of the Taz, Turuhan, Baïkha and Yelogui rivers. Northern (Taz), Central (Tym) and Southern (Ket) dialects, each with sub-dialects, are easily distinguished – yet the total number of speakers is only about 2,000.

### Founder of Samoyed linguistics

Mathias Alexander Castrén (1813–52) travelled at the request of the Russian Imperial Academy of Sciences. A colleague of Elias Lönnrot (see FINNISH), he made two epic journeys in Arctic Russia, preparing reports, ethnographic studies, grammars and wordlists of many languages. Castrén was appointed Professor of Finnish at Helsinki in 1851, and married in the same year. His health broken by tireless exploration in inhospitable climates and primitive conditions, he died a few months later.

### Numerals in two Samoyedic languages (from Castrén's notes)

Forest Nenets		Selkup
ngop	1	oker
shtye	2	sitte
nyahar	3	naagur
tyeat	4	teetta
samblyang	5	somblea
mat'	6	muktet
sheu	7	seeldyu
shenttyeat	8	sitte tyaadin göt
kashem yut	9	oker tyaadin göt
yut	10	kööt

The origin of the numerals '8' and '9' – 'two less than ten', 'one less than ten' – is more obvious in Selkup than in Nenets. The Latin numerals '18' and '19', *duodeviginti*, *undeviginti*, are built in just the same way.

From *Samojedische Sprachmaterialien*  
gesammelt von M. A. Castrén und T. Lehtisalo  
(Helsinki: Suomalais-Ugrilainen Seura, 1960)

# SANGO

## PERHAPS 5,000,000 SPEAKERS OF SANGO AS A LINGUA FRANCA

### *Central African Republic*

**S**ango is a creole based on a dialect of Ngbandi, which is one of the Ubangi group of NIGER-CONGO LANGUAGES, with a very long history in the Ubangi valley. Sango is now the 'national language' of the Central African Republic.

Sango was the dialect of a fishing people living around Mobaye on the River Ubangi. In the 19th century Sango began to be known as a language of trade westwards and downriver among the Gbanziri and Buraka, but it was after 1889, when the French founded their fort Bangui (now capital of the Central African Republic), that the use of Sango (by now a simplified, pidginised offshoot of the original dialect) grew more rapidly. French traders, soldiers and administrators used the river for communications, and took many of the Gbanziri, Buraka and other river people into their service. Where French was as yet unknown, Sango had to be the common language in which communication took place.

This language of trade and urban employment is known in French as *Sango Commercial*, and its own users call it *Sango tí Salawísi* 'Working Sango' – *salawísi* comes from French *service* 'employment'. It developed by way of *Sango tí tulúgu*, 'Soldiers' Sango'.

When colonial development turned inland, Sango was already the favoured language in local government, the army and trade. It was also easy for most people to learn, deriving as it does from a member of the Ubangi group. Thus it now serves as a second language almost throughout the Central African Republic, for speakers whose mother tongues are Banda, GBAYA, Gubu, Langbasi, Manja, Mono, Togbo – all of them Ubangi languages – and for speakers of Kare, a Bantu

language. Historically, to travel, to trade, to find work outside one's home village, to be sophisticated, one had to know Sango.

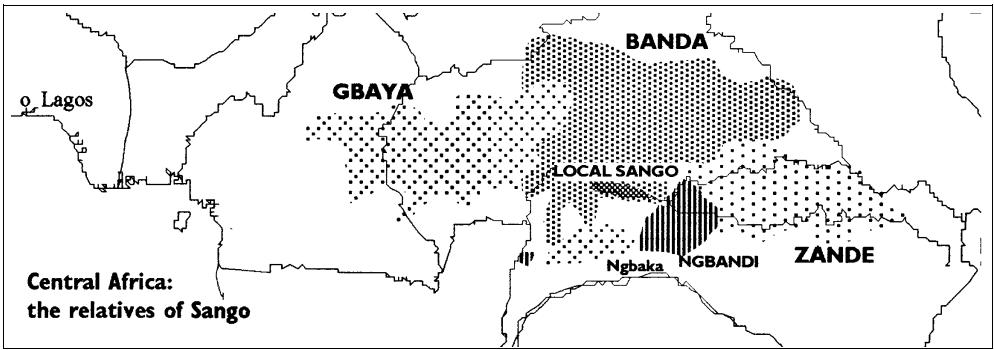
Christian missionaries found Sango essential to their purposes. It was they who developed a written form of the language for textbooks and Bible translations, which began to appear in the 1920s. However, the Roman Catholics and the Protestants did not work together. The Protestants use a standard form of the language, and a spelling, based on the dialect of Bossangoa in the north, while the Roman Catholic standard language and spelling originate from Mbaiki near Bangui.

The true standard, however, from which regional dialects borrow and to which they adapt more or less, is not that of the priests but of the traders and truck drivers, whose Sango is fairly homogeneous all over the country. The Sango spoken by children also tends to be much more uniform than that of their parents. While traders' and children's Sango borrows readily from French, old people's Sango is much more influenced by local languages.

In 1965 Sango officially became the national language of the Central African Republic, though French is the language used for official publications. Sango is still, for most people, a second language rather than a mother tongue, but in the big towns the number who speak only Sango and French, and never learn a regional language, is growing all the time.

The first ten numerals in Sango are: *ókò, óse, otá, osió, ukú, omaná, mbásámbará, miombe, ngombáya, bale ókò*.





### Banda, Zande, Sango, Ngbandi and Gbaya

Banda, Zande, Ngbandi and Gbaya are four Ubangian languages whose early speakers, at least three thousand years ago, migrated from the west into the area that is now divided between southern Sudan and the Central African Republic.

ZANDE speakers lived mostly in Sudan until, in the 18th century, they invaded the country between the Uele and Mbomu rivers, now in Congo (Kinshasa). Five Zande dialects are usually recognised.

*Ngbandi* speakers had earlier moved south-westwards to the land between the Ubangi

and the Zaire. There are now 200,000 of them, mostly in Congo (Kinshasa). Most are bilingual in LINGALA, the major lingua franca of northern Congo. But meanwhile, through its own use as a lingua franca of the Ubangi fishing people and traders, Ngbandi has given rise to the much more important and widespread *Sango*.

BANDA and GBAYA are the hinterland languages of the Central African Republic. While Zande and Ngbandi speakers migrated in turn eastwards and south-westwards, Banda and Gbaya probably remained roughly where they were throughout the last two millennia, and have gradually split into numerous dialects, now so distinct that some are not mutually intelligible.

# SANSKRIT

## CLASSICAL LANGUAGE OF INDIA

**S**anskrit is the linguistic parent of all the INDO-ARYAN LANGUAGES. It was the inspiration to the remarkable linguistic researches of ancient Indian grammarians – and it catalysed the development of linguistics in Europe in the 19th and 20th centuries.

*Sam̐skṛta*, so named by later Indians, was ‘the perfected language’, by contrast with *Prākṛta*, ‘the common language’.

No contemporary records in Sanskrit survive from before the 2nd century AD (when Indian rulers began to use Sanskrit, instead of everyday Prakrit, for their inscriptions). But the language is far older than that.

### Vedic and classical Sanskrit

The oldest form of Sanskrit now known is in the venerable poems of the *Ṛgveda*, a collection of religious lyrics usually thought to have been composed around 1000 BC. They are the sacred texts of Hindu worship, transmitted orally for many centuries and still recited, word perfect, in modern rituals. The dialects in which they are composed must have been current somewhere in north-western India – not a ‘perfected language’ but a real spoken language of three thousand years ago.

From that point onwards, the history of Sanskrit diverges from that of the colloquial languages of India. In everyday speech, the early Vedic dialects developed and diverged, by way of the regional Prakrits of the first millennia BC and AD, into the Indo-Aryan languages of today.

Meanwhile, in ritual and in learned memory, around the *Ṛgveda* poems other texts gathered, in gradually developing forms of ‘Vedic Sanskrit’. Successively the later Vedas, the Brahmanas, the Upaniṣads became canonical in their

turn. There soon followed a vast scholarly apparatus of commentaries, paraphrases and glossaries whose purpose was to ensure the true pronunciation and correct understanding of the ancient collections.

This apparatus is so comprehensive that the *Ṛgveda* must certainly have been transmitted accurately from the mid first millennium BC until it was eventually written down. The linguistic work also led – about 400 BC – to the writing of a grammar, not of Vedic but of the somewhat later language of the scholars and commentators, which was evidently itself no longer the speech of every day. This grammar by Pāṇini, so concise and formulaic that commentaries on it have to be many times its length, is often admired (less often read) by modern theoretical linguists.

What Pāṇini fixed, and later authors wrote on his pattern, was ‘classical Sanskrit’. It is a beautiful, flexible, musical language, the vehicle of remarkably luxuriant poetry and astonishingly concise prose. It was never anyone’s mother tongue.

Classical Sanskrit literature flourished for a whole millennium after Pāṇini’s time: in fact it has never ceased to be cultivated. Many of the earlier texts are now lost. Aśvaghoṣa’s *Buddhacarita*, the story of the Buddha, written in the 1st century AD and known from later manuscripts, is one of the oldest that survives. Kālidāsa, who wrote plays and poetry in the 5th century, is the acknowledged master. Later writings stretched almost infinitely the flexibility of Pāṇini’s linguistic design, with compound words of enormous length and ever more extravagant rhetorical figures.

The volume of classical Sanskrit literature is vast. The language is still used for poetry, technical writing and doctoral dissertations: modern Sanskrit works are regularly published. The

readership is limited, but by no means vanishing. In the Indian census of 1971, 200,000 people claimed to speak the language. Sanskrit, and not only the Vedic Sanskrit of the *R̥gveda*, can still be heard in the religious rituals of Hindu sects.

### Sanskrit as a world language

In the early centuries AD Sanskrit gradually gained a role as the 'official' and learned language of all India – and of the many countries then under Indian political and religious influence. By now it was far removed from the everyday language of north India. It had to be learnt. Once learnt, it evidently served for communication very widely.

As a sign of its wide familiarity, a popular form of Sanskrit was used for oral epic poetry not long after Pāṇini's time. Eventually written down at vast length, the *Mahābhārata* became India's national epic. No less popular, the *Rāmāyaṇa* of Vālmīki, perhaps composed about AD 200, told the story of Rama using the same form. Their language is far simpler than that of contemporary classical Sanskrit poetry. They have inspired translations, imitations and retellings in all the modern languages of India.

There are Sanskrit inscriptions of the first millennium AD not only from India itself but also from Sri Lanka, Burma, Thailand, Cambodia, Vietnam, Malaya, Sumatra, Java and Borneo. While the very oldest Sanskrit inscription comes from India, almost as old is the stele of Vo-can in Vietnam, about AD 200: by this time Sanskrit had become an elite language of the states of Fu-

nan and Champa in Indochina, as it was soon to be in the Malay archipelago. Sanskrit literature gave both themes and poetic forms to the literature, sculpture and painting of south-east Asia and Indonesia. All the languages of culture in the region have borrowed heavily from Sanskrit.

The holy texts of Buddhism, first composed in PALI, were translated into Sanskrit, fleshed out with extensive new writings and commentaries, and studied in China, Japan and Tibet. At first it was a very strange kind of Sanskrit, Prakrit words put back artificially into Sanskrit shapes, 'Buddhist Hybrid Sanskrit'. Likewise the Jain holy texts, originally in ARDHAMAGADHI, accrued Sanskrit translations and commentaries. Not only inscriptions but versified state histories, such as the Kashmiri Kalhana's *Rājataranginī*, show the importance of Sanskrit to Indian royal courts.

### The nature of Sanskrit

It was in the 18th century that Europeans began seriously to explore the literary culture of India. From the beginning, those who learnt the classical language, Sanskrit, were impressed by its similarity to the Greek and Latin of their own school and university studies. Some thought it was actually the parent language of Greek and Latin: for the rather different view of Sir William Jones, essentially a view that linguists would still accept, see box at INDO-EUROPEAN LANGUAGES. The 'perfection' of Sanskrit – the complexity of its forms and the impressive way that the tradition of Pāṇini had reduced them to rule – continued to enthrall linguists: 19th-century at-

#### A list of Hindu divinities

बभ्रुरेको विष्णुः सूनरो युवाञ्ज्यङ्क्ते हिरण्ययम्  
[Soma]  
योनिमक आ ससाद द्योतनो ऽन्तर्देवेषु मेधिरः  
[Agni]  
वाशीमेको बिभर्ति हस्त आयसीमन्तर्देवेषु निधुविः  
[Tvāṣṭr]  
वज्रमेको बिभर्ति हस्त आहितं तन वृत्राणि जिघ्नते  
[Indra]  
तिग्ममेको बिभर्ति हस्त आयधं शुचिरुगो जलापभेषजः  
[Rudra]

One – brown, bitextile, cheerful, young –  
gilds him with sweet gold.  
One, nestling, flames in his furrow,  
wise amid the gods.  
One's hand holds an adze of iron,  
working amid the gods.  
One's hand holds a skybolt fast:  
with it he kills his foes.  
One's hand holds a piercing blade –  
bright, angry, drenching, calming . . .

*R̥gveda* 8.29, attributed to Kasyapa or Vaivasvata Manu

**Indo-Aryan horse races**

Older than the oldest surviving Sanskrit texts, there is indirect evidence that a mysterious Indo-Aryan language was known, far away to the west, in the Hurrite-speaking kingdom of the Mitanni in eastern Anatolia. An émigré or captive Mitannian named Kikkuli wrote a manual of horsemanship in HITTITE – in which he incorporated technical terms, for the length of a race as measured in ‘turns’ of the course, that are unmistakably Indo-Aryan:

eka-vartanna	one-lap
tera-vartanna	three-lap
panza-vartanna	five-lap
satta-vartanna	seven-lap
navartanna	nine-lap

No Indo-Aryan language texts survive in Near Eastern cuneiform. How did these exotic words get into their author’s vocabulary? The Mitanni must have learnt their horsemanship skills from Indo-Aryan teachers, by way of a diplomatic link, a migration or a conquest that is otherwise quite unrecorded.

The words listed above can be compared with the Sanskrit numerals in the next box.

tempts at reconstructing proto-Indo-European still look too much like Sanskrit and not enough like the European offspring of the family.

Sanskrit is indeed in many ways our closest single source for proto-Indo-European. It was spoken, however, a world away – in geography and in

**Numerals in Sanskrit and Pali**

Sanskrit		Pali
ekas	1	eko
dvau	2	dve
trayas	3	ti
catvāras	4	catu
pañca	5	pañca
ṣaṭ	6	cha
sapta	7	satta
aṣṭau	8	aṭṭha
nava	9	nava
daśa	10	dasa

climate – from any country where proto-Indo-European may have been spoken. Words for plants, animals and products specific to India were borrowed or newly coined. *Tāla* ‘palm’, *khala* ‘threshing-floor’ came from DRAVIDIAN LANGUAGES; *marica* ‘pepper’ came from an Austroasiatic language; *hastin* ‘elephant’ was an invented word, literally ‘the one with a hand [on its trunk]’.

**Sanskrit in writing**

Sanskrit can be written and printed in almost any of the scripts of south and south-east Asia – and it still is. Most often used nowadays is the Devanagari of north Indian languages. The standard Latin transliteration given in the box was adopted by a congress of Orientalists a century ago and is almost universally accepted among modern scholars.

Additional symbols for  $\bar{r}$  and  $\bar{l}$  scarcely occur in practice.

**Sanskrit in Devanagari and Roman**

प्र प्रा इ ई उ ऊ ऋ ॠ ए ऐ ओ औ क ख ग घ ङ च छ ज झ ञ ट ठ ड ढ ण त थ द ध न प फ ब भ म य र ल व श ष स ह  
a ā ī ū ṛ ̄ e ai o au k kh g gh ṇ c ch j jh ṇ ṭ ṭh ḍ ḍh ṇ t th d dh n p ph b bh m y r l v ś ṣ h

# SANTALI

**4,000,000 SPEAKERS**

*India, Bangladesh*

**S**antali is one of the Munda languages of rural eastern India (see map at MUNDARI). These are now known to belong to the family of AUSTRASIATIC LANGUAGES.

Santali is spoken in the western districts of West Bengal; there are about 100,000 speakers in Bangladesh. It is traditionally an unwritten language; speakers who learn to read and write do so in Bengali or Oriya, and on the occasions when Santali is written, Bengali or Oriya scripts are generally used. Christian missionaries and their followers have preferred Latin script.

As with its close relative Mundari, Santali and its speakers have little recorded history before the 19th century. While Mundari are typically hill peoples, many Santali speakers are lowland farmers (see box at MUNDARI): many, too, have migrated to Assam to find work in the tea plantations there.

The most important scholarly work on Santali is the great five-volume dictionary by the missionary P. W. Boddington, published in Oslo in 1929–36. Boddington's Santali grammar remained unfinished at his death.

Numerals in MUNDARI and Santali		
Mundari		Santali
mid'	1	mit'
barya	2	bar
apia	3	pɛ
upunia	4	pɔn
mōrea	5	mɔrɛ
turwia	6	turui
ēa	7	eae
iryia	8	iril
ārea	9	arɛ
gelea	10	gel

The traditional counting system continues in twenties up to '400'. '20' in Santali is *mit' isi*. In both Mundari and Santali, however, the native numerals are used less than before: Hindi (or, in the case of Santali, Bengali) numerals are taking their place.

The checked *t'* in Santali *mit'* 'one' was the same sound which 'has received wide publicity in Harry Lauder's pronunciation of *Sat'urday night*', according to the missionary linguist R. M. Macphail.

## In search of a tradition

'The Santali *Oi Cemet'* script was devised about fifty years ago by a Santal, Pandit Raghunath Murmu, as part of his extensive program for culturally upgrading the Santali community. As he saw it, every respectable high-culture language in India had its own script and an old (written) literature. He provided Santali with an ingenious, indigenous, and custom made script, and was also responsible for providing a Santali classical epic as the Santali analogue to the [Sanskrit] *Mahābhārata*. The script is not based on Devanagari principles; it is an alphabet with all vowels indicated by separate letters ordered linearly on the Roman model.'

Norman H. Zide in *Current trends in linguistics* vol. 5 (The Hague: Mouton, 1969) pp. 425–6

# SARA

**900,000 SPEAKERS**

*Chad*

**S**ara, a group of dialects belonging to the NILO-SAHARAN LANGUAGES, is spoken in the southwestern corner of Chad, south of N'Djamena, between the Logone and Chari rivers.

Sara speakers were particularly mobile during the French suzerainty in central Africa. Many of them worked on the Congo–Ocean railway from Brazzaville to Pointe-Noire, others at the port cities of Congo (Brazzaville), Gabon and Camer-

oun, others again at desert outposts in Chad and Niger, where they served in the militia. In colonial times, therefore, Sara was heard in many of the larger towns of French central Africa.

Sara has three tones, which are important in verb conjugation. It has the unusual feature of a doubled plural form: *de* 'person', *dege* 'people', *degege* 'parties of people'.

# SARDINIAN

**PERHAPS 1,500,000 SPEAKERS**

## *Sardinia*

**T**he Sardinian dialects are widely recognised as a language separate from ITALIAN (see map there). Of all the ROMANCE LANGUAGES, Sardinian is agreed to be the closest to classical Latin. Why?

The answer is that Sardinia was one of the earliest Roman conquests, in 238 BC. Latin must soon have been widely used there, as a *lingua franca* of the army, administration and trade. And this was early Latin, pre-classical even, before the rapid changes that were soon to affect the language. The last population upheaval was in 177, when tens of thousands of Sardinians were sold in Italy as slaves. After the vast conquests of the 1st century, when population interchanges and troop movements were spreading the ‘vulgar’ Latin of the empire, Sardinia, no longer rebellious, had become a backwater.

In medieval times Catalan and Spanish were the languages of government in Sardinia. Italian was little used there until 1714, but since that time it has been the official and literary language of the island. Long despised and rejected by Italian authorities, Sardinian is a language almost without literature, although occasional records of the language go back as far as AD 1080. It has now been recognised as a regional language of Sardinia, in second place to Italian, a change of status which may help to keep Sardinian alive. It is still widely used on the island, though all younger people are bilingual in Italian.

Sardinian was already seen by medieval Italians as quite a different language from their own: Dante, who wrote on the Italian dialects, saw Sardinian as something distinct, ‘imitating Latin as apes imitate men’. In a troubadour verse by Raimbaut de Vacqueyras an imagined Genoese lady finds her suitor’s Provençal as difficult as the speech of ‘a German or Sardinian or Berber’: *no t’entend pluï d’un Toesco, o Sardo o Barbari*.

While all the other Romance languages developed a seven-vowel system, Sardinian retains the five vowels of classical Latin. There are four dialects. Logudorese, spoken inland, is least affected by Spanish, Catalan and Italian, and gives the strongest impression of closeness to Latin.

Corsican is spoken by about 170,000 inhabitants of the island of Corsica, which has been French territory almost continuously since 1769 – coincidentally the year in which the most famous Corsican, Napoleon Bonaparte, was born. The language shows natural similarities to the Tuscan dialects of Italian, with independent features deriving from the curious history of Corsica, which has seldom accepted external rule without resistance. French has been the language of education and administration for well over a century: Corsican survives, so far, in local and household contexts.

Numerals in Italian, Sardinian and Corsican			
	Italian	Sardinian	Corsican
1	uno, una	unu, una	unu
2	due	duos, duas	dui
3	tre	tres	tre
4	quattro	battor	quàtru
5	cinque	kimbe	cinque
6	sei	ses	sei
7	sette	sette	sette
8	otto	otto	ôttu
9	nove	nove	novu
10	dieci	deke	dèce



# SASAK

**2,100,000 SPEAKERS**

*Indonesia*

**S**asak, one of the AUSTRONESIAN LANGUAGES, is spoken on the island of Lombok, east of Bali.

The island was politically and culturally under Balinese influence from the 17th century until recently but, unlike Bali, it is largely Muslim. Sasak has thus been influenced both by Balinese and by Arabic, as well as by Malay and its newer official form, Indonesian. The same *aksara* 'script' used for JAVANESE (see table there) and Balinese has also been used in the past for Sasak, but most of those who write the language now adopt the Latin script familiar from Indonesian.

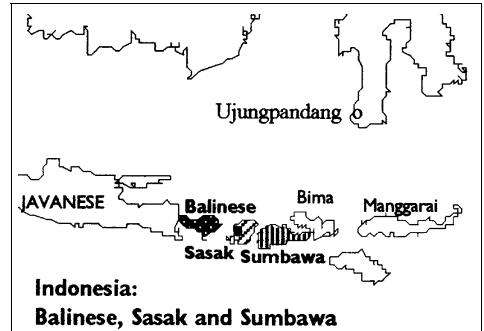
The first ten numerals in Sasak are: *səkek*, *duə*, *təlu*, *əmpat*, *limə*, *nəm*, *pitu*, *balu*, *siwa*, *sə-pulu*.

## Balinese, Sasak and Sumbawa

The three languages are close to one another linguistically.

*Sasak*, spoken in Lombok, is traditionally divided into five dialect groups: *ngeno-ngene*, *meno-mene*, *mriak-mriku*, *kuto-kute*, *ngeto-ngete*.

BALINESE is spoken in Bali and Nusa Penida,



with an enclave in Lombok. The conservative Mountain dialects are to be distinguished from the Lowland group; within this, the north-eastern dialects form the standard of the press and media, while the dialect of Denpasar, in the south, is quite distinctive. Balinese dominance of Sasak radiated from the area of Mataram and Cakranegara, where there are still about 80,000 speakers of Balinese.

*Sumbawa* is the language of the western extremity of Sumbawa island. It has about 300,000 speakers.

# SEMITIC LANGUAGES

**T**his group of languages of the Near East shares a similar grammatical structure and easily recognised relationships among words: it is clear that they descend from a single, lost, 'proto-Semitic' language, perhaps spoken about four thousand years ago. The Semitic group is more distantly related to several others in a family spanning much of north Africa: together they are usually known as the **AFROASIATIC LANGUAGES**.

The *Semitic* group is so called because A. L. Schl  zer, in 1781, looking for a term to denote the group including **ARABIC**, **HEBREW** and **ARAMAIC**, saw that the speakers of these languages were descendants of Noah's son Shem or Sem, according to the biblical story in Genesis 10.21–31 and 11.10–26.

Later it was observed that the South Arabian languages, classical **ETHIOPIC** and many of the modern languages of Ethiopia (see **AMHARIC** and map there) belong to the Semitic group too. And as cuneiform tablets and inscriptions of the ancient Near East began to be deciphered, in the 19th century, some of these languages, notably **AKKADIAN**, turned out to be Semitic as well.

In the earliest written records, speakers of Semitic languages – Akkadian, Hebrew, Phoenician and others – lived in what are now Iraq, Syria, Lebanon, Jordan and Israel.

*Ugaritic* is the name given to the language of the ancient city of Ugarit (modern R  s S  mar). A rich collection of texts in Ugaritic dates from the 14th and 13th centuries BC. It was a North-west Semitic language, with a unique feature – its script is cuneiform (formed of wedge-shaped strokes for impressing on clay tablets) like Akkadian, yet it is a true alphabet, each symbol representing a single consonant as in Hebrew and Arabic.

Phoenician inscriptions come from the cities along the eastern coast of the Mediterranean and are dated between the 10th and 1st centuries BC. They also are in an alphabetic script, and one that was not designed for writing on clay tablets. These early Phoenician scripts are again crucial in the history of the alphabet: they were the source from which the first known Aramaic, Hebrew, Moabite and Samaritan alphabets developed. Inscriptions in Punic, a variety of Phoenician, are known from the neighbourhood of ancient Carthage, the powerful Phoenician colony in what is now Tunisia. An Etruscan–Phoenician bilingual inscription, unfortunately too short to help much with the decipherment of Etruscan, was found at Pyrgi in Italy.

Semitic languages share an important structural peculiarity. Roots, embodying a basic meaning, consist of consonants alone: vowels are inserted, and prefixes and suffixes added, in fixed patterns, to make complete noun and verb forms and other derivatives.

It may well be because of this feature (also true of ancient **EGYPTIAN** and other Afroasiatic languages) that early alphabets, developed for Egyptian and the early Semitic languages, provide symbols for consonants only: the reader was able to insert vowels in reading based on a knowledge of the general structure of the language. This feature made for quicker writing. However, when applied to languages with quite different structures, such writing conventions were inadequate. The later history of alphabets is partly a history of the solutions to this early design fault.

For maps of the modern Semitic languages see **ARABIC**, **ARAMAIC**, **AMHARIC**.

# SENA

**1,200,000 SPEAKERS**

*Mozambique*

**S***ena* is the name for one of the BANTU LANGUAGES, one that is closely related to the now better-known NYANJA (see map there). It is spoken south of the Nyanja-speaking region, in central Mozambique. In Portuguese colonial times Sena was regarded as one of the three major African languages of the country.

Historically it is the language of an ancient town, *Sena* – a nexus of trade on the lower Zambezi – and of the surrounding peoples. When first described in Portuguese sources, in 1571, Sena had probably already been in existence for some centuries, used by Muslim – Arabic- and Swahili-speaking – traders who dealt

in the cotton cloth of this region and in the gold that came from the old-established annual fairs of the Manyika goldfields far inland (see SHONA). The Portuguese built a settlement nearby, with a church dedicated to St Catherine of Siena, and until the 1760s Sena was a centre of Portuguese rule in central Mozambique. Tete later took its place.

The first grammar of Sena was written by an anonymous Portuguese around 1680 – a reminder of its early importance in European trade. He called it *Arte da lingua de Cafre*, ‘Grammar of the language of Kaffir’. In later sources the language appears as *Cafreal de Sena*, *Chisena* or *Senna*.

# SERBIAN, CROATIAN AND BOSNIAN

**18,500,000 SPEAKERS**

*Serbia, Croatia, Bosnia, Montenegro*

**T**hese are the current names for one of the South SLAVONIC LANGUAGES, best known as Serbo-Croat, the majority language of the former Federal Republic of Yugoslavia.

Illyrian dialects (possibly related to modern Albanian) had been spoken in the western Balkans at the time of the Roman conquest, and had gradually been supplanted by Latin. Romans called the region *Illyria* and *Dalmatia* (both names have occasionally been used for Serbo-Croat). Slavonic speakers probably settled here in the 6th and 7th centuries AD.

The division between Croatians and Serbians is in origin a religious one. After conversion to Christianity, in the 9th century, both groups used the OLD SLAVONIC liturgy. The Croatians, gradually aligning themselves to Rome, retained the Glagolitic alphabet of the earliest Slavonic texts – and eventually adopted Latin script for everyday use. The Serbians maintained their links with Constantinople, and Serbia became one of the main centres of literacy in Old Slavonic in the widely used Cyrillic script. A third religious element entered the scene when, after the Turkish conquest of Serbia and Bosnia, Islam gradually spread in the cities and towns of Bosnia, and there the language was sometimes written in Arabic script.

Croatia remained largely outside the Turkish Empire: here the languages of power until the end of the 19th century were the German of Austria and the Italian of Venice, the latter particularly along the Adriatic coast. Montenegro (*Crnagora*, the 'Black Mountain') maintained an

obscure independence.

In spite of political and religious divides and in spite of the different scripts, the language spoken over this whole region was relatively uniform. Until the 18th century the languages of literature, education and government were very varied – German, Hungarian, Italian, Latin, Old Slavonic, Slavenoserbian (a mixture of Russian Church Slavonic and Serbian), Turkish. There had been some writing in the local language, mostly in Croatia, with varying spellings.

Vuk Karadžić (1787–1864), a Serbian who lived for most of his life in Vienna, was inspired by Bosnian–Serbian folk poetry to spread the use of the popular language. He received support from Jernej Kopitar, the Slovene linguistic reformer. Vuk's collections of songs, beginning 1814, his grammar of 1814 and his dictionary of 1818 began the process of formulating a new Serbian national language. It was firmly based on a Hercegovinan dialect: Vuk had family links with the region, and much of the greatest folk poetry came from there. This choice did not please all Serbs – but it happened to place the new Serbian language very close to the dialects of Croatia, and by the Literary Accord of 1850 it was accepted as the literary language of Croatia too, replacing the local standards in which some Croatian literature had already been written.

Serbia and Montenegro continued to use the Cyrillic script. Croatia used Latin: so did Bosnia as the Arabic script fell out of use. Spelling reforms promulgated by Vuk and others made Latin and Cyrillic easily interchangeable.

In 1918, with the formation of Yugoslavia, for the first time nearly all the speakers of Serbo-Croat were under the same government. Their continuing cultural differences were marked by the division into Croatia, Bosnia, Serbia and Montenegro when Yugoslavia became a Federal Republic in 1945 – and were disastrously highlighted when it collapsed into warring states at the end of the 1980s.

The two standards, Croatian and Serbian, have now been joined and will grow further apart. They already differed in some features. Serbian has tended to borrow international terms from other languages, while Croatian has tended to invent words, often ‘calques’: Serbian *biblioteka*, Croatian *knjižnica* ‘library’. Besides this, Serbian is richer in Turkish and Greek loans while there are more numerous Latin and German words in Croatian.

### Beyond the Balkans

Many Serbians fled northwards to Hungary after a failed revolt against Turkish rule in 1690. The Serbian population of Vojvodina, then in southern Hungary, dates from this migration. Vojvodina was incorporated in Yugoslavia in 1918. Under the Austrian Empire there were important communities of Serbo-Croat speakers in Vienna and Budapest.

Today, émigré Serbo-Croat speakers are numerous. There are at least 250,000 in the United States. Some Muslim speakers of Serbo-Croat have escaped religious persecution by fleeing to Turkey: as many as 100,000 had done so by 1980, and others have followed since the recent warfare in Bosnia.

A Croat community, now numbering 20,000, settled around Oberpullendorf, in Austria, in the 16th century. They were sometimes called Wasserkroaten (water Croats). Until 1921 the official language that surrounded them was Hungarian: now it is the German of Austria. There is some publishing in Croat in this district: the spoken language, however, has changed so much that people from Croatia can no longer understand it without help.

### The names of Serbo-Croat

The term “Serbo-Croatian”, including names of both groups, has a very recent origin. In Serbo-Croatian the language is, or has been, called by the following names: *srpskohrvatski* and *hrvatskosrpski* “Serbo-Croatian, Croato-Serbian”, *srpski ili hrvatski* and *hrvatski ili srpski* “Serbian or Croatian, Croatian or Serbian”, *srpski* “Serbian”, *hrvatski* “Croatian”, *ilirski* “Illyrian”, and occasionally even *naš jezik* or *naški* “our language”. In the 18th century, regional terms like *slavonski* “Slavonian” were also common. These terms have been used with a number of meanings and connotations. Their exact significance depends as much on the writer’s dialect as on the period involved. One of the striking features of the *Književni dogovor* “Literary Accord” of 1850, which established the principle of a unified literary language for the Serbs and Croats, is its failure to name the language. One wonders if this omission was deliberate or simply an oversight on the part of the signers.’

Kenneth E. Naylor, ‘Serbo-Croatian’ in *The Slavic literary languages: formation and development* ed. Alexander M. Schenker, Edward Stankiewicz (New Haven, 1980) pp. 65–83, abridged

### The South Slavonic languages

Literary *Slovene* is based on the Lower Carniola (*Dolenjsko*) dialect with some features from *Gorenjsko*, which includes Ljubljana, to the north. The Slovene-speaking districts of Carinthia (*Korosko*) in Austria have dialect features linking them to Slovak and an archaic stress pattern.

*Serbo-Croat* dialects are customarily classified by the word used for ‘what’: *što*, *kaj* or *ča*. *Štokavian* covers all of Serbia, Montenegro and Bosnia and parts of Croatia. All the dialects are mutually intelligible, though the *Čakavian* of the northern Adriatic coast and islands is difficult for others because of its accentuation, which is more like that of Slovene and Russian. A division into

**The Serbian alphabet and  
the Croatian/Bosnian equivalents**

А Б В Г Д Ђ Е Ж З И Ј К Л М Н  
Њ О П Р С Т Ћ У Ф Х Ц Ч Ш

а б в г д ђ е ж з и ј к л м н  
њ о п р с т ћ у ф х ц ч ш

ABVGDĐEŽZIJKLMN  
NjOPRSTĆUFHCČDŽŠ

abvgdđežzijklmn  
njoprstćufhcčdžš

sub-dialects is made according to a typical vowel or diphthong developing from earlier Slavonic *Ѣ*. In most of Serbia the sub-dialect is *Ekavian* while in Croatia, Bosnia and Montenegro *Ijekavian* is most widespread.

*Macedonian* dialects can be classified by word stress: on the third syllable from the end in the widespread western dialects, on the second from the end in the Castoria and Florina areas of Greece, and variable in other northern Greek



districts and in the Pirin region of Bulgaria.

Standard *Bulgarian* is based on the eastern dialects but with a mixture of western features. In the western dialects, as in Serbia, earlier *Ѣ* develops into *e*, while in the east it most often becomes *ya*. The western dialect of Sofia and its neighbourhood is known as *Shopski*.

# SEYCHELLOIS

**75,000 SPEAKERS**

*Seychelles*

**O**ne of the FRENCH CREOLES, Seychellois or Seselwa or Seychelles Creole is the mother tongue of 95 per cent of the inhabitants of the Seychelles.

French was the language of the original colonisers. Under English rule, from 1811 to 1976, English was the language of administration, while French remained the language of the educated elite. Through this period the Creole of the majority was not used in writing or in any

official context. The first book in Seychelles Creole, a translation of St Mark's Gospel, appeared only in 1974.

Shortly after independence, in 1981–2, Creole was declared the first national language (English and French being the other two) and became the language of primary education. There was a flood of reading books, textbooks and government publications in Creole.

# SHAN

**PERHAPS 3,000,000 SPEAKERS**

*Burma, China, Thailand*

**S**han belongs to the TAILANGUAGES. Spoken by several million people in inner south-east Asia, Shan was once the ruling language of most of the 'Shan states' that used to maintain a precarious autonomy in the highland valleys between Burma, China and Thailand. The land routes linking Burma with China pass through this region, and Shans have traditionally been prominent in long distance trade.

*Shan* is a Burmese name. Speakers call themselves *Tai*, distinguishing southerners as *Tai Taii* and northerners as *Tayok* and *Tai Neua*.

Shan chronicles trace the history of some states as far back as the early 13th century: this includes the old principality of Mogaung, now Burmese-speaking, and the AHOM domain in the valley of Assam. In the vacuum left by the collapse of the Pagan kingdom in Burma, Shan-speaking rulers extended their power over much of the Burmese plains, but linguistically they were soon absorbed by the surrounding Burmese. In most of the remaining Shan states, Shan speakers were numerically dominant in the valleys, surrounded by numerous hill peoples, including speakers of WA, Palaung, HANI, LAHU and LISU. In the northern Shan states there is a complex cultural interchange among speakers of Shan and of JINGHPAW.

Shan is the most widely spoken language of two adjacent administrative districts. One is the Shan State in eastern Burma, over which the Burmese Army and local warlords have fought for decades. North of this is the Dehong Dai autonomous region of Yunnan, China. Shan is also the dominant language in the Machongson district in north-western Thailand.

Shan is a written language with its own alphabet, which, like that of Burmese, is an adaptation of the MON script. The traditional culture is Theravada Buddhist, originally acquired by way of Burma. Buddhist 'birth stories' and other local legends are written in Shan; the scriptures are studied in PALI, which is written not in Shan but in Burmese script.

The best-known dialect has been the Burmese Shan standard of the old city of Möngnai, once a state capital and the seat of a Burmese viceroy. Chinese Shan or Tai Nüa is the dialect of the Shan-speaking districts of Yunnan. Between these two falls the dialect of Hsenwi, a state that used to maintain close relations with the Chinese administration. Northern Burmese Shan is perceptibly different from the southern dialect, with six tones instead of five. For a table of numerals see TAI LANGUAGES.

## **A Shan city in decay**

'From Möngnai a fine view of the whole valley. On the west are abrupt jungle-covered rocks crowned with pagodas. The slope is finely wooded, and for the great part overgrown with jungle. A long road, which might be made into a magnificent drive, leads up to the town. The brick walls are about ten feet high, machicolated and falling to ruins. The town has every sign of having been at one time fine and spacious, now a pitiful wreck. Pitched [camp] in the centre of the town by the side of the road under a pipal tree. Shortly after arrival the Prince sent me four watermelons and two water-jars.'

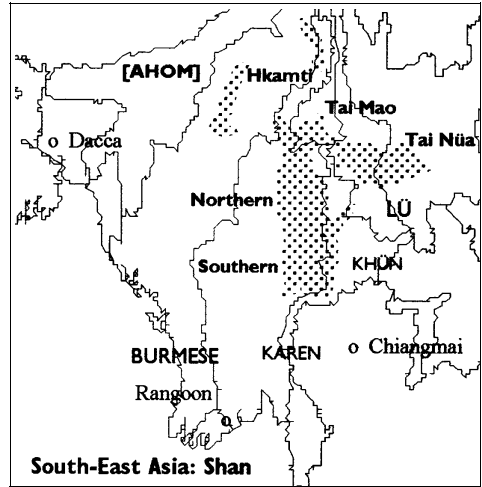
Sir George Scott: Diary, May 1887 (India Office Library and Archives)



## Shan and Ahom

There are three major subdivisions of Shan, in terms both of dialect and of culture. *Southern Burmese Shan* roughly matches the area administered from Möngnai (Burmese *Monè*) before the British conquest, and the western part of the 'Southern Shan States' of British rule. *Northern Burmese Shan* is spoken in the area administered from Hsenwi (Burmese *Theinni*) in the 18th and 19th centuries, the 'Northern Shan States' of British rule. In earlier centuries this area acknowledged Chinese suzerainty. *Chinese Shan* (or Tai Nüa, Tai Mao) is the language of the states to the north and east of these, which have generally been subject to China and are now grouped in the Dehong Dai autonomous region.

There are smaller, more scattered remnants of Shan states in north-western Burma, notably in the valley of Hkamti Long, where there are about 3,000 remaining speakers of Hkamti Shan. Shan rule once stretched as far as the valley of Assam, where until the 18th century AHOM was the ruling language.



## Shan scripts

The traditional alphabet, shown here, has not enough symbols to express the sounds of Shan fully. There are two standard styles of writing, known traditionally as Burmese Shan (with rounded characters) and Chinese Shan (a cursive script). Based on these different styles, new extended alphabets, with symbols indicating the tones, have now been adopted in printing in Burma and China. The Hkamti script, a third variety, is now almost forgotten.

### The Shan alphabet

က ခ င ဂ သ ဂု တ ထ ဃ ပ ဟ် မ ယ ရ လ ဝ သ ဌ န

ka kha nga sa hsa nya ta hta na pa hpa ma ya ra la wa sha ha a

# SHONA

**8,000,000 SPEAKERS**

*Zimbabwe, Mozambique*

**O**ne of the BANTU LANGUAGES, Shona is the overall standard form of the dialects that are spoken by most inhabitants of the country once called Southern Rhodesia. Shona is now the national language and lingua franca of independent Zimbabwe. There are also many speakers of Shona dialects in Mozambique.

No local traditional name is known for the whole Shona group as now defined: people generally used to identify themselves with smaller ethnic groupings. *Shona* is in origin a name applied by NDEBELE speakers to the conquered population, but it is now widely accepted. From it come the older English terms *Mashona* for the people, *Mashonaland* for their territory, and *Chishona* for the Zezuru dialect.

In stark contrast with the Bantu languages that lie to the south, from NDEBELE to XHOSA, Shona comprises a series of distinct, related dialects each of which has a fairly compact territory. This testifies to a long period of relatively peaceful development *in situ*, and makes it almost certain that early Shona speakers were the builders of the 8th- and 9th-century stone *zimbabwe*, the towns and burial places whose ruins are the most striking archaeological feature of the country, and after which it has now been named.

Portuguese reports of the 16th and 17th centuries describe the Empire of Monomotapa and name several peoples who can be identified with still-existing Shona ethnic groups and dialects, notably *Manica* and *Mocaranga* (Manyika and Karanga), the latter being the most powerful.

The first recorded foreign invasion is that of the NDEBELE, in 1838: they conquered some

southern Shona districts and established supremacy over the rest. It was not long after this that David Livingstone travelled through 'Mashonaland', visiting and naming the Victoria Falls in 1855. English-speaking settlers, at first under the auspices of Cecil Rhodes's British South Africa Company, held power for a hundred years until independence in 1980.

Shona dialects were first used in writing by Christian missionaries, who began to work among Shona speakers in 1859 and produced religious books and Bible translations in Karanga, Zezuru, Manyika and Nda. u.

The dialects were so close to one another that the competition of several written forms within Southern Rhodesia was seen as undesirable by the British colonial administration, both for its own reasons and for the sake of the nascent publishing industry. In 1931 the linguist Clement Doke recommended that Shona and NDEBELE be the two official African languages of the territory, and that the official literary form of Shona be based on the Karanga and Zezuru dialects. These were grammatically very much alike, they were the largest two dialects in terms of number of speakers, and Zezuru was the dialect of the capital (then called Salisbury).

Foreign loanwords in Shona include *mufarinya* 'cassava' from Portuguese *farinha* 'flour'; *mutšara* 'line' from Swahili *mstari* and originally Arabic *mistara* 'ruled line on paper'. Shona also has numerous Afrikaans loanwords.

The first ten numerals are: *poshi*, *piri*, *tatu*, *cina*, *shanu*, *tanhatu*, *cinomnge*, *rusere*, *pfumbamnge*, *gumi*.

### Shona literature and music

'Traditional literature is expressed in clan histories, folktales and proverbs. In the folk tales the principal character is the Hare (Sulo, Shulwe, Tsuro), who assumes many qualities, but there are also other important animals – lion, zebra, baboon – and some well-known human [characters]. Stories are told at night by one of the family elders, and while a few of the tales point a moral, the main emphasis is on entertainment.

'There are war songs, hunting songs, love songs, work songs and ritual songs. Most songs among the Shona are associated with musical accompaniment and frequently with action. Karanga of the Victoria District are outstanding composers of the whole country, and many of their musical compositions have spread, with variations, from one tribe to another.

'Dancing is also highly developed, and in association with song and musical accompaniment marks such social events as marriages, funeral ceremonies (at which the women sing) and exorcism.'

Hilda Kuper, 'The Shona' in *Ethnographic survey of Africa. Southern Africa* part 4 (London: International African Institute, 1954) p. 14, abridged

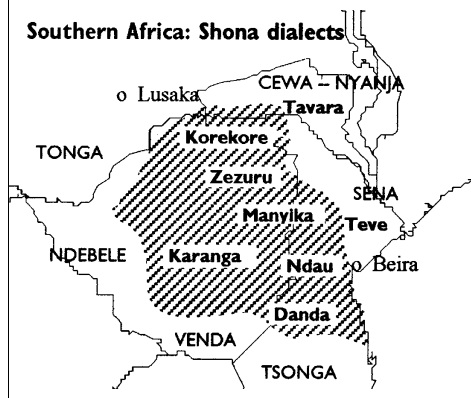
### The standardisation of literary Shona

Recommendation 5. That a dictionary of Shona be prepared, to be as inclusive as possible of words from Zezuru, Karanga, Manyika, and Ndau. That for the present Korekore words be admitted sparingly and that the use of colloquial words from the dialects of Budya, Tavara, Karombe, Danda, Teve, Eangwe etc. be discouraged.

Clement C. Doke, *Report on the unification of Shona dialects*, Salisbury, 1931.

Doke's standard Shona has gradually come into being: the *Standard Shona dictionary* appeared at last in 1966. Doke's recommendations on spelling were less successful. Like many English-speaking linguists of his time, he liked the Africa Alphabet of the International Institute for African Languages and Cultures. He recommended a 33-letter version for Shona, with several characters not found on normal typewriters or in normal printing fonts. These were dropped in 1955, but it was only in 1967 that an adequate and generally acceptable orthography was introduced.

### Southern Africa: Shona dialects



### Shona lullaby

Rú-u, rú-u, rú-u, rú-u,  
Harúruhwe, wáchema mwaná –  
Harúruhwe, wáchema mwaná:  
E, é, é, é, tazvínzwa, babá!  
maítweiko, babá, wangu?  
matadzírweiko némwana?

Rú-u, rú-u, rú-u, rú-u,  
Harúruhwe, the child is crying,  
Harúruhwe, the child is crying:  
E, é, é, é, I heard your complaint, father!  
What has been done to you, father?  
How has your child offended you?

A. C. Hodza, *Shona praise poetry* (Oxford: Clarendon Press, 1979) p. 19

# SINDHI

**9,000,000 SPEAKERS**

*Pakistan, India*

**O**ne of the INDO-ARYAN LANGUAGES, Sindhi is spoken in south-eastern Pakistan, centring on the city of Hyderabad. Within Sindhi territory, Karachi forms an enclave with a majority of Urdu speakers. There are also about a million Sindhi speakers in India – some settled close to the Pakistan border, others in trading communities in many of the big cities.

*Sind*, the lower Indus region, was the first part of India to be captured by the Muslims, in 712. Its name was already ancient – a variant of the river name, *Indus* in Latin but *Sindhu* in Sanskrit.

At the time of the Muslim conquest Indo-Aryan languages had already been spoken here for over a thousand years. From this point onwards, however, Sind was often to be politically distinct from neighbouring parts of India, and its language, too, gradually became distinct. Arabic and then Persian were the languages of government for a thousand years. Sindhi has, naturally, numerous loanwords from these languages and relatively few from Sanskrit.

There are a few early records of the Indo-Aryan speech of Sind, from the 8th century onwards, and evidence exists that literature in the local language was being written down. It is lost, however: as far as surviving texts are concerned, a native Sindhi literature begins at the end of the 15th century. Its main tradition is of lyric poetry, most often showing Sufi inspiration. The greatest Sindhi poet is perhaps Shah Abdul

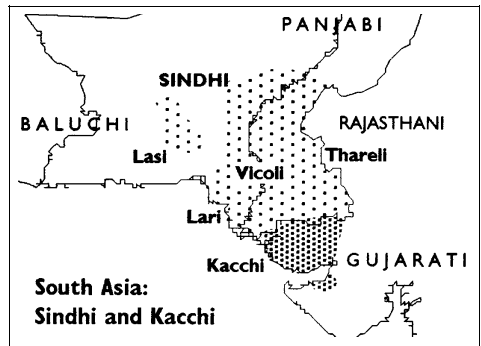
Latif of Bhit (1689–1752).

Sindhi is usually written in its own script, adapted in 1852, under British influence, from the Perso-Arabic script used for Urdu. Unlike Urdu, printed Sindhi is often typeset rather than calligraphed.

The first ten numerals in Sindhi are *hiku*, *bba*, *ṭī*, *cāre*, *pañja*, *cha*, *sata*, *aṭha*, *nava*, *ḍḍaha*.

## Sindhi and Kacchi

*Vicolī* is the ‘standard’ dialect of Sindhi, on which the modern literary language is based. *Lāsī* is the dialect of Las Bela, *Lārī* that of the lower Indus. *Tharellī*, spoken in the eastern desert region, is close to Rajasthani. In the desert region of Cutch, in north-western Gujarat state, India, *Kacchī* has about 500,000 speakers. Linguistically it is best regarded as a Sindhi dialect, but its speakers use Gujarati as their literary language. When written, Kacchi uses the Gujarati alphabet.



**Kacchī balladry**

Kuṃārī kandh namāyā, goṭeṃ ghā kerī:  
coṭā jhalyā cosarā, vaḍhy ūṃ kandh vicā . . .

Virgins bowed their necks, bridegrooms struck the blows:  
They seized their four-braided hair, and struck them across the neck.

The disastrous aftermath of the battle of Jhārā in 1762, when Ghulām Shāh Kalhorā of Sind conquered Cutch.

*Linguistic Survey of India* vol. 8 part 1 ed. Sir G. A. Grierson (Calcutta, 1919) pp. 201–6

# SINHALA

12,000,000 SPEAKERS

*Sri Lanka*

**S**inhala or Sinhalese is the majority language of Sri Lanka (Ceylon), spoken by at least two-thirds of the population. It is one of the INDO-ARYAN LANGUAGES, separated by hundreds of miles from its relatives, for in south India DRAVIDIAN LANGUAGES are spoken.

The island figures as *Laṅkā* in classical Sanskrit literature, notably the epic *Rāmāyaṇa*, which tells the story of Rāma's invasion of the island to recover his captive wife Sitā. *Sri Lanka* (the honorific prefix *Śrī* is a modern embellishment) is thus a revival of an ancient name. But the island was better known as *Sinhāladvīpa*, locally *Sihaladīpa* (a name already found in a Greek source of the 6th century): this is the origin of the language name *Sinhala* and of the traditional medieval and modern names for the island, Arabic *Sarandīb*, Portuguese *Ceilão*, English *Ceylon*.

How did Sinhala come to be spoken where it is? The Pali history of Sri Lanka, *Mahāvamsa*, tells that the language was brought to Sri Lanka by settlers under the leadership of Vijaya in the very year of the *nibbāna* or 'extinguishing' of the Buddha in distant northern India, a date given as 504 BC. Linguists have concluded that this cannot be far from the truth: for Sinhala branched off from the other Indo-Aryan languages at a very early stage, long before dialect divisions among them had hardened into language boundaries. An Indo-Aryan colonisation must indeed have taken place around the 6th century. And by the 3rd century BC inscriptions from Sri Lanka show that the ancestor of Sinhala had already developed some of the special features of modern Sinhala – notably the absence of

an aspirated/unaspirated distinction in consonants, one of the many ways in which Sinhala shows the influence of its Dravidian neighbour TAMIL.

The inscriptions are all that we have to trace the history of the language until the 9th century AD, the date of the earliest surviving Sinhala literature. The older poetic form of Sinhala, the vehicle of medieval poetry, is *Eḷu*. This was already an archaic tongue, one that had to be formally learnt, in the 13th century, for that was when the grammar *Sidatsaṅgarāva* was compiled, on the model of classical Tamil grammars. This venerable work now serves as a source of information on lost early literature. The dictionary tradition begins with *Piyummala*, a guide to poetic synonyms, dating from the 15th century – and, even earlier, with a Sinhala translation of the Pali dictionary *Abhidhānappadīpikā*.

Sri Lanka was an early stronghold of Buddhism and has remained one of the principal centres of this religion. Thus, throughout its history, Sinhala has coexisted with Pali, the classical language of Theravada Buddhism and the language of religious education. The two languages must at the outset have differed only slightly. Commentaries on Pali texts have been a traditional genre of Sinhala literature, and the language is full of Pali loanwords. For centuries Buddhist monks and scholars have travelled between Sri Lanka, Burma, Thailand and Cambodia.

The island was divided among seven kingdoms when first visited by Portuguese explorers in 1505. The final result of long warfare by Portuguese, Dutch and British invaders was that the whole island came under British rule from

1803 with the conquest of the kingdom of Kandy. Portuguese missionaries had made many converts to Christianity. Although Catholics were persecuted under Dutch rule, which lasted in Colombo from 1656 to 1795, as many as 10 per cent of the population are Catholics today.

A Sinhala–Dutch dictionary, the first involving a European language, appeared in 1759, and a Sinhala–English dictionary in 1821.

English was the medium of education till 1948. In independent Sri Lanka Sinhala was proclaimed the sole official language in 1956. A destructive civil war, concentrated in the north of the country where there is a Tamil-speaking majority, led to the emigration of large numbers of Tamil speakers, but eventually to the recognition in 1987 of official status for English and Tamil; the war, however, continues. English is often used informally in administrative contexts, but Sinhala remains the official national language and civil servants must be prepared and able to use it.

There is a considerable émigré population of Sinhala speakers in Singapore, Thailand, the United Arab Emirates and Canada.

Spoken as it is at a crossroads of international trade and travel, Sinhala has borrowed from many languages. It has been influenced by Dravidian languages, especially Tamil, in phonology, vocabulary and syntax. Portuguese became a powerful influence in the 16th century; Dutch and English followed. Loanwords include *kamise* ‘shirt’ and *keju* ‘cheese’ from Portuguese; *kokis* ‘cake’ and *almariya* ‘wardrobe’ from Dutch; *telefon* ‘telephone’, with many other modern terms, from English.

Sinhala has a suffixed definite article in singular nouns: *miniha* ‘man’, *minihek* ‘the man’, *minissu* ‘men’. The first ten numerals are: *eka*, *deka*, *tuna*, *hatara*, *paha*, *haya*, *hata*, *aṭa*, *namaya*, *dahaya*.

### Literary and colloquial Sinhala

What was said above of loanwords leads to the crucial distinction between literary or ‘high’ and colloquial or ‘low’ Sinhala: for while colloquial Sinhala adopts modern loanwords freely, the

literary variety draws on Sanskrit, Pali and *E/ū* for its abstract and technical vocabulary. The attraction of English is so powerful that it is noticeable in both varieties. In literary Sinhala one finds Sanskrit-like new compounds that turn out to be modelled (‘calqued’) on English compounds or expressions, while colloquial Sinhala has numerous pure borrowings from English.

Literary Sinhala is associated with Buddhism and education; it is the language of newspapers, official documents and formal oratory. It is used in media news reports and government forms. It is the language of most printed books, including poetry and novels (though passages of dialogue may be written in the colloquial language).

The difference between the two forms is so great that they may legitimately be regarded as two languages. It involves not only the vocabulary but also the basic structural words of the language and the grammar of noun and verb forms: in literary Sinhala verb forms are marked for person, in spoken they are not. Even the personal pronouns differ.

All speakers learn to understand literary Sinhala to some extent: its everyday uses make this ability almost indispensable. But not all learn it thoroughly. The vocabulary is the greatest barrier to the unpractised hearer or reader. It requires sound education and practised skills to produce literary Sinhala extempore for extended periods: relatively few speakers can do this without difficulty. Since 1947 there have been calls for the use of colloquial Sinhala as the official and literary standard, and this controversy continues. Meanwhile, a formal variety of colloquial Sinhala, with some of the grammatical features of the literary language, is now used in lectures, political speeches, sermons and pop songs.

Among the dialects of Sinhala, Kandyan shows more Tamil influence, partly because it lies nearer to areas of Tamil speech in Sri Lanka and India, partly because Kandy remained independent of European rule until 1803. The low-country dialect has adopted more loans from Portuguese, Dutch and English. The dialect of

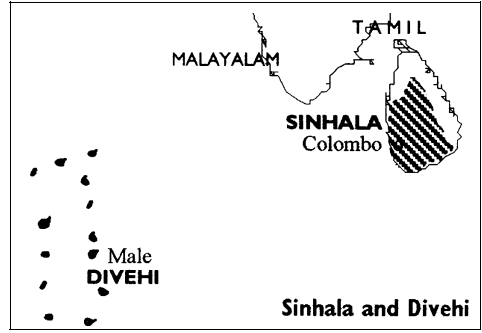
Colombo and the south-west is becoming the national standard. Cutting across geographical borders, the dialects of *Rodiyas* or untouchables, and of *Ahikuntakayas* or travellers, are distinctive, with some of the features of a 'secret language' intended to exclude outsiders. Farmers and foragers are said to have a secret vocabulary of auspicious words to counteract the effect of evil spirits and to ward off the dangers of the forest.

### Sinhala and Divehi

*Sinhala* is the majority language everywhere in Sri Lanka except in the north-east, where the majority speaks Tamil. Many Tamil speakers are bilingual in Sinhala.

The distinctive *Vedda* dialect is spoken by a small aboriginal group whose ancestors may be surmised to have spoken an unrelated language, now forgotten.

DIVEHI is the language of the Maldives, also spoken on Minicoy Island, India, where there are about 5,000 speakers.



### The Sinhala script

The Sinhala alphabet is one of the descendants of Brahmi (see INDO-ARYAN LANGUAGES). Like other Indic scripts it is written from left to right.

In its development through the centuries the script has gradually become more rounded: it is perfectly adapted for writing on palm-leaves with a stylus. The usual printed fonts have exaggerated heavy and light strokes, which is not a feature of the handwritten script. The tiny loops and circles typical of Sinhala are easily confused in poorly printed text.



# SINO-TIBETAN LANGUAGES

**T**he Sino-Tibetan language family includes three main groups: CHINESE, the KAREN LANGUAGES and the Tibeto-Burman languages.

The fact of a relationship among the languages of the Tibeto-Burman branch has been recognised for many years. Karen is very different from these. Chinese differs even more extensively, and must have separated from the rest many thousands of years ago, but the phonetics and some of the vocabulary of proto-Sino-Tibetan have been successfully reconstructed and the ultimate relationship among these languages is not in doubt.

Their prehistory has not been fully worked out. It seems possible that proto-Sino-Tibetan was spoken, perhaps ten thousand years ago or more, in south-eastern China, not far from where the dialects ancestral to the AUSTRO-TAI LANGUAGES were also spoken. From here the speakers of early Karen and Tibeto-Burman dialects will have moved gradually westwards, though there is no evidence for the migration beyond the presence of those languages where they are now on the map of south-east Asia. Meanwhile speakers of early Chinese dialects will have spread generally north-westwards from their evident nucleus near the south-eastern coast.

Pyu, the earliest known written language of Burma, extinct for almost a thousand years, was also a Sino-Tibetan language: its closer relationships are uncertain. The Pyu stone inscriptions were deciphered by C. O. Blagden, in 1917, on the basis of the four-language Myazedi inscription at Pagan in central Burma.

It used to be assumed that the Miao-Yao languages (see MIAO and YAO) and the TAI LANGUAGES would turn out to be related to Sino-Tibetan, but that was a false lead. As monosyl-

labic, tonal languages they are superficially very like Chinese; but then Chinese has changed drastically from its proto-Sino-Tibetan ancestor – and the changes may partly have been the result of interaction with early forms of Miao, Yao, Tai and other languages now believed to belong to the Austro-Tai family.

Proto-Sino-Tibetan was not as thoroughly monosyllabic as so many of its descendants are. Nor can proto-Sino-Tibetan tones be reconstructed, and it may not have been tonal at all, though again the great majority of modern Sino-Tibetan languages do make use of tones.

## Reconstructing proto-Sino-Tibetan from the 'monosyllabic' languages of today

'If a high-powered racing car is driven at terrific speed into a cement wall, the results on the car will somewhat parallel those on polysyllabic Sino-Tibetan words. The front part will be greatly compressed, parts will have dropped out, and there will be considerable distortion; but the body will probably remain fairly intact.'

Robert Shafer, *Introduction to Sino-Tibetan* (Wiesbaden: Harrassavitzgh 1974) p. 21

## The controversial subgrouping of Sino-Tibetan

The CHINESE languages make up one main division of Sino-Tibetan. Clearly related to the rest, they have been developing separately from them for many thousands of years.

The KAREN LANGUAGES are sufficiently distinct for most linguists to agree that they too form a main division, separate from both Chinese and Tibeto-Burman.

The Tibeto-Burman languages, the third division, have spread and divided far more widely than the other two. They pose a difficult problem of subgrouping – a solution to which would help to elucidate the prehistory of south-east Asia and southern China. The most likely subdivisions of Tibeto-Burman seem to be those numbered 1 to 5 below.

1. JINGHPAW or Kachin.

2. Rung. Languages that may belong to this group include Rawang, with 100,000 speakers in Kachin State, Burma, and Qiang or Ch'iang or Dzorgaish, with 130,000 speakers in Sichuan, China. Some scholars think that Tangut (cf. 4 below) belongs here.

3. Kado, a language with 200,000 speakers in Burma, China and Laos.

4. The Burmese-Lolo languages, including AKHA, Burmese, LAHU, LISU, Maru, Tujia and YI or Lolo: for all these see map at BURMESE. The

extinct TANGUT or Hsihsia was possibly also a member. BAI and NAXI may be distant relatives of this group, but their status is uncertain.

5. The Bodic languages, so called by linguists from the native name of Tibet, *Bod*. This includes TIBETAN itself and its close relatives Gurung (150,000 speakers in Nepal), Tamang (500,000 speakers) and Jiarong (100,000 speakers in Sichuan); also NEWARI, Limbu and Magar.

6. The Baric group, the most complex of all, which seems to have several subdivisions. The BODO-GARO LANGUAGES include Bodo, Garo and Tripuri. The KUKI-CHIN AND NAGA LANGUAGES include many languages of relatively small communities; they are close to the Mikir-Meithei group, consisting of MANIPURI and Mikir. The Mirish or Abor-Miri-Dafla or North Assam languages include Adi or Miri (500,000 speakers in Assam and Tibet) and Lhoba (200,000 speakers in Arunachal Pradesh and China).

# SIRAIKI

**PERHAPS 15,000,000 SPEAKERS**

*Pakistan*

**O**ne of the INDO-ARYAN LANGUAGES, Siraiki has not until recently been regarded as a separate language. It may be considered a dialect or a continuation of Western PANJABI (see map there), spoken in the lower Indus valley in a region where SINDHI is also spoken.

Siraiki has been the vehicle of religious poetry, collected in manuscripts and occasionally

published. But – in a local backlash against an influx of Muslim speakers of ‘eastern’ Panjabi, religious exiles from India after the 1947 partition – there has in recent years been agitation among local intellectuals and politicians for ‘language status’ for Siraiki in Pakistan.

The first ten numerals in Siraiki are: *hik, du, tre, cār, pañj, chi, sat, aṭh, naō, dah*.

Aukhā bbaṃdhī rakhaṇ yārīdā  
jivem kachā dhāggā.

Friendship is as hard to rebind  
as a half-untwisted thread . . .

Opening of a Siraiki poem, from *Linguistic survey of India* vol. 8 part 1

# SLAVONIC LANGUAGES

**A** branch of the family of INDO-EUROPEAN LANGUAGES, the Slavonic languages are spoken in Russia and eastern Europe by a total of 300,000,000 speakers.

As far back as they can be traced historically, Slavonic speakers called themselves by a name like *Slověne*. As *Sklavenoi* they are mentioned by Greek historians in the 6th century AD. For the language family as a whole many linguists now prefer the term *Slavic* to *Slavonic*.

At the battle of the river Lech, in 955, so many Slavonic prisoners were captured by the Germans that their ethnic name developed a new meaning – still seen in English *slave*, French *esclave*, Italian *schiavo*. The Venetian dialect form *ciao* ‘goodbye’, now popular in Italian and occasionally in fashion in English, is the same word in origin. Thus it means literally ‘I am your slave’.

Among the early Indo-European dialects we can discern a group – those that eventually became Baltic, Slavonic, Iranian and Indo-Aryan languages – in which a particular sound change took place, the change from *k* to *s*. In these languages, therefore, the word for ‘hundred’ begins with *s*. These dialects shared much more than the single sound change. A range of similarities, in sounds, in grammar, in vocabulary, links the Slavonic and Baltic languages. Slavonic languages, and sometimes the Balto-Slavic group as a whole, share a great deal with the Indo-Iranian languages too. For many hundreds of years they must have been spoken in adjacent regions, probably of the southern Ukraine, before the speakers of Indo-Iranian began the long, slow movement south-eastwards that has led their linguistic descendants to Iran and northern India.

Slavonic speakers, meanwhile, in all probability stayed exactly where they were: but their territory must gradually have expanded, reaching the Oder, and even the Elbe, to the west. There was certainly long term interaction between Slavonic and Germanic dialects.

Finally, in the early centuries of the Byzantine Empire, a startling series of migrations and conquests took Slavonic speech south into the Balkans. The Latin dialects of south-eastern Europe were almost engulfed, though ROMANIAN and AROMUNIAN remain, heavily influenced by Slavonic. The language of the Bulgars (see CHUVASH) was soon forgotten. Documents suggest that Slavonic speakers once formed a majority in mainland Greece. The modern languages of Bulgaria and the countries of former Yugoslavia are the legacy of these medieval Slavonic invasions.

The first recorded form of Slavonic speech is in the Christian texts in OLD SLAVONIC, many of them translations from Greek, which began to be written during the mission of Saints Constantine and Methodius to the Principality of Great Moravia in 862–3.

The missionaries had learnt their Slavonic much further to the south, in Macedonia, but at that date the spoken Slavonic languages cannot have differentiated very greatly and the texts they wrote could probably be understood by all Slavs.

Meanwhile about ten modern Slavonic languages have developed a written and literary form. They are still rather closer to one another than (for instance) Germanic or Romance languages. For some examples see the tables of Slavonic numerals at BELORUSSIAN, CZECH and SLOVENE.

## Members of the group

SLOVENE, SERBIAN and CROATIAN, MACEDONIAN and BULGARIAN are grouped as 'South Slavonic'. Geographically cut off from the rest, these languages result from Slavonic incursions into the Byzantine Empire in the 7th to 9th centuries.

CZECH and SLOVAK, POLISH and SORBIAN form the West Slavonic group. There is strong dialect differentiation in Poland: some would count Kashubian as a separate language.

The Eastern Slavonic languages (BELORUSSIAN, UKRAINIAN and RUSSIAN) were the last to differentiate, as the medieval Russian principalities became tributary to Lithuania, Poland and the Golden Horde. Ruthenian, spoken to the south-west of the Carpathians, may be considered a dialect of Ukrainian – but some claim for it the status of a language.

For maps see SERBIAN, POLISH and RUSSIAN.

## The trees of central Europe

For the woodland trees that are well known in the Ukraine and southern Russia, the Slavonic languages have inherited names that come from proto-Indo-European or from a very early period of proto-Slavonic. But for trees that are native only to the west of the Vistula and the Dniester, beech, sycamore maple, larch, yew and sorb, they use names borrowed from other languages. This seems to show that the Slavonic dialects developed east of those two rivers – and it confirms that the migrations leading to the differentiation of Western and Southern Slavonic languages, from Sorbian to Bulgarian, came later.

**Beech:** *buk* in Russian and Polish, *bukva* in Serbian and Croatian: the Slavonic word was borrowed from a Germanic form such as Gothic *boka*. There was a proto-Indo-European word for 'beech': it was *\*bhāgos*. It survived, with its original meaning, in Celtic Latin as well as in Germanic. Early Greek and Slavonic speakers, long settled in regions where the beech was unknown, used this word to name other trees. Greek *phēgós* means 'Valonia oak', *Quercus macrolepis*, and Russian *buzina* (with dialect forms like *bas*, *baznyk*) means 'elder'.

**Maple, sycamore:** *jawor* in Polish, *javor* in Russian and Serbian, are names for certain maple trees (including what the British call 'sycamore'). The word was borrowed from a Germanic form such as German *Ahorn*, related to Latin *acer*.

**Sorb:** Serbian, Croatian and Slovene *brek*, Czech *břek*, and Slovak *brekyňa* are names for the sorb or serviceberry, the rowan that is native to central Europe. This word seems to be borrowed from a Germanic form like German *Birke* 'birch', adopted as a name for the sorb because the two trees have similar bark.

Based on Zbigniew Golab, *The origin of the Slavs: a linguist's view* (Columbus, Ohio: Slavica, 1992) pp. 273–80

# SLOVAK

**5,500,000 SPEAKERS**

## *Slovakia*

**S**eparated from Poland by an almost continuous mountain range, a continuum of dialects occupies the territory that, for most of this century, was united as Czechoslovakia. They are closely similar to one another and mutually intelligible: however, two standard languages have emerged from them, Czech and Slovak. They are Western SLAVONIC LANGUAGES.

*Slovák*, plural *Slováci*, the name that Slovak speakers use for themselves, is a variant of the term *Slavonic* that is applied to the whole language group. It is first recorded in a document of 1485. They call their language *Slověnský* (a name easily confused with that of Slovene).

Slovak is spoken in the mountainous region that lies due north of Hungary (see map at POLISH). The territory has no tradition of political independence, having previously existed as a state only briefly under German domination in 1939–45.

In the 9th century the Slovak region belonged to the Kingdom of Great Moravia. But for many hundreds of years after that it was part of Hungary, and Hungarian was its official language. However, documents in Slovak, or in Czech with recognisable Slovak features, are known from the 15th century onwards. The origin of modern standard Slovak may thus be traced to unofficial use, in local administration and by a local intelligentsia, of a language which, though close to Czech, was developing to serve their own needs on the basis of their spoken dialects. Czech and Slovak standards continued to compete: in a religious context Czech may be seen as the Protestant language, Slovak as the Catholic lan-

guage in early modern Slovakia. It was only in the 19th century that a widely accepted literary standard Slovak emerged. A landmark is the appearance of L'udovít Štúr's *Nárečia slovenskua alebo potreba písania v tomto nárečí* 'Slovak speech, or the necessity of writing in this dialect' in Bratislava in 1846.

When the region was at last detached from Hungary to become part of the newly formed Czechoslovakia, in 1918–38, Slovak language and literature were able to flourish, yet their separate status was not fully accepted by the Czech majority. In post-war Czechoslovakia Slovak had a higher status. It is now, almost for the first time, the national language of an independent country, one whose people have in the past been held together by little more than their language. An inherited defensiveness leads to strong discrimination in favour of standard Slovak, against local dialects and against minority languages such as ROMANI.

Although Slovak and Czech are so similar (see the table of numerals at CZECH), Slovak is one of the easiest of languages for other Slavonic speakers to learn, while Czech is one of the more difficult. This is partly because of the *česká přehláska* 'Czech vowel mutation', a comprehensive series of sound changes beginning around the 14th century which affected vowels that followed palatalised consonants. Slovak, on the other hand, geographically central to the Slavonic world, has links with Southern and Eastern Slavonic languages as well as with the adjacent Czech and Polish.

# SLOVENE

2,000,000 SPEAKERS

## *Slovenia*

**S**lovene, one of the South SLAVONIC LANGUAGES, is the national language of the Republic of Slovenia, until recently part of Yugoslavia.

*Slovenski* is now the local name for the language – in origin the same name as that of the *Slavonic* language family. Owing to its political fragmentation in earlier centuries, dialects of Slovene were sometimes identified as separate languages with their own names: *Kranjski* for the speech of Krajnska or Carniola, *Windisch* (a German word with pejorative overtones, see SORBIAN) for that of Styria and Austrian Carinthia.

Slavonic speakers must have settled in this region soon after the fall of the Roman Empire, perhaps at the same time as in Slovakia. The two languages have some similarities although they have been geographically separated ever since the arrival of the Hungarians in the 9th century.

For a thousand years most Slovene speakers were Austrian subjects. Hungarian, Italian, but most of all German, were the languages of the cities and the elite in this part of the Austrian Empire.

Slovene was often seen as nothing more than a rustic dialect. Yet publishing in Slovene began quite early, with a *Catechism* and an ABC printed by Primož Trubar in 1551. The great Bible translation of 1584 included a glossary explaining the Krajnska and other local words that were used, so that Southern Slavonic speakers in general would also be able to read it. Slovene speakers are largely Roman Catholic, but it was Protestant publishing that contributed most to the early development of written Slovene. From the beginning the Latin alphabet was used.

In the following centuries there was a jostling for

position – a ‘battle of the forms’ – between the local Slovene dialects and a generalised literary language. The latter, which had begun to emerge even in Trubar’s books, was based on the Krajnska dialect of Ljubljana more than on any other, but it has features drawn from several dialects. There was also a need to fix the spelling of Slovene: writers in local dialects had based their spelling practice on Italian, German or Hungarian patterns. The influence of German word order and syntax was gradually cut back.

Some writers and theorists went further, planning for a unified future South Slavonic or even ‘pan-Slavic’ language. But the Vienna Accord of 1850, uniting the SERBIAN AND CROATIAN literary languages, left the remaining South Slavonic languages – including Slovene – to find their own future. Meanwhile, after the short-lived French invasion of 1811, Slovene had been accepted for the first time as a language of education, local administration and the press. It naturally became the official language of Slovenia, first as one of the constituents of federal Yugoslavia, and at last as an independent republic.

There are communities of Slovene speakers in Hungary, Italy and Austria near the Slovene borders (for map see SERBIAN). The Austrian Slovenes live in three valleys, Gailtal, Rosental and Jauntal. By recent counts there are 20,000 of them, but they form a majority only in one or two villages. In Italy, in and around Trieste, there are about 30,000 Slovene speakers. They are a majority in the small towns of San Dorligo della Valle and Sgonico. Trieste itself, once the seaport of the multilingual Austro-Hungarian Empire, has a large Italian majority.

*Kar ne pride iz srca, se ne prime srca*: What does not come from the heart does not move the heart.

**Numerals in South Slavonic languages**

In this table Macedonian and Bulgarian are given in Latin transliteration. Slovene appears in its original Latin script, while the Serbo-Croat form is given in both Latin (for Croatian) and Cyrillic (for Serbian).

	<b>Slovene</b>	<b>Croatian</b>	<b>Serbian</b>	<b>Macedonian</b>	<b>Bulgarian</b>
1	en	jedan	један	eden	edin
2	dva	dva	два	dva	dva
3	trije	tri	три	tri	tri
4	štirje	četiri	четири	chetiri	chetiri
5	pet	pet	пет	pet	pet
6	šest	šest	шест	shest	shest
7	sedem	sedam	седам	sedum	sedem
8	osem	osam	осам	osum	osem
9	devet	devet	девет	devet	devet
10	deset	deset	десет	deset	deset



# SOGDIAN

## EXTINCT LANGUAGE OF CENTRAL ASIA

**S***ogdiane*, as it is called in Greek texts, or *Suguda* in Persian inscriptions, was a region of central Asia between the rivers Amu Darya and Syr Darya. It belonged to the Persian Empire, and afterwards to Alexander and his successors. Little is known of its history in the millennium that elapsed between then and the Islamic conquest in the 8th century: but coins, inscriptions and a few historical references show that it was at times independent and influential.

Among the Middle IRANIAN LANGUAGES, the importance of Sogdian – our name for the language of this little-known region – comes from its geographical position. Sogdiana lay across the Silk Road, and was deeply involved in the trade. Sogdian became a lingua franca: texts in it are now found at oases far to the east, in Chinese Xinjiang, where Sogdian trading communities must have lived and where Sogdian language and literature must have been cultivated.

*Chorasmian* was the Iranian language spoken west of Sogdiana, on the lower course of the Amu Darya. Chorasmia or Khwarezm was likewise for a long time an independent state, eventually succumbing to the Islamic expansion. There is no Chorasmian literature, but the language is known from inscriptions, documents, and short quotations by Arabic authors. These sources range from the 3rd century BC to the 13th century AD.

Sogdian had a rich literature, long forgotten but now rediscovered in early manuscripts from the Silk Road cities – Buddhist texts, Manichaean texts and Christian texts, all in distinctive scripts. These poems and scriptures, in a once well-known language, helped to spread the beliefs and philosophy of three great religions well beyond Sogdiana itself to the heartlands of Asia.

Most of the surviving manuscripts come from finds at Tunhuang and Turfan in Xinjiang.

Sogdian eventually died away, replaced by the Tajik variety of PERSIAN and by UZBEK. But one community of about two thousand people, settled in northern Tajikistan, still speaks *Yaghnobi*, the direct descendant of medieval Sogdian (see map at IRANIAN LANGUAGES).

The *Bactrian* language, spoken to the south of Sogdiana, is found on coins and a few inscriptions of the early centuries AD in Greek script – a reminder that this region had been the centre of a once-powerful Greek kingdom. The letter *ϐ* was added to the script to represent the sound *sh*, unknown in Greek.

### Numerals in Sogdian and Chorasmian

Sogdian		Chorasmian
'yw	1	'yw
'ðw	2	'ðw'
'ðry	3	šy
ctf'r	4	cf'r
pnc	5	pnc
xwšw	6	'š
'βt	7	'βd
'št	8	'št
nw'	9	š'ð
ðs'	10	ðys

Vowels are not fully recorded in the original scripts and cannot be reconstructed with certainty. The transliterations are those conventionally adopted by Iranian scholars.

I. Gershevitch, *A grammar of Manichean Sogdian*. Oxford, 1954; Helmut Humbach, 'Choresmian' in *Compendium linguarum iranicarum* ed. Rüdiger Schmitt (Wiesbaden: Reichert, 1989) pp. 193–203

# SOMALI

**5,500,000 SPEAKERS**

*Somalia, Ethiopia, Kenya, Djibouti*

**S**omali is one of the CUSHITIC LANGUAGES of the Afroasiatic family. It is the national language of Somalia, acting as a lingua franca for speakers of minority languages throughout the country.

According to one theory, the 'Horn of Africa' is the region from which the earliest Afroasiatic languages began to spread, many thousands of years ago. From the time of the first historical records, two thousand years ago or more, all available evidence suggests that Somali has been spoken precisely where it is spoken today, in what is now Somalia and by a significant minority in south-eastern Ethiopia. Its speakers are traditionally nomadic pastoralists; their mobility has been only partly arrested by the frontiers that the French, Italians, British and Ethiopians have drawn across the Somali plateau.

People of this region were once known as *Berber*, hence the name of the town of Berbera. They call themselves *Soomaali*. The word is first found, in an Ethiopic form, in a 15th-century praise poem for an Ethiopian king's victories over his eastern neighbours. It seems to derive from the name of a legendary ancestor, *Soma* or *Samale* – but many other origins have been suggested, including the two Somali words *so* 'go' plus *mal* 'milk' as an indication of a pastoral way of life.

There are no significant early records of the language. Henry Salt's *Voyage to Abyssinia* (1814) contained a wordlist of Somali: the French linguist Antoine d'Abbadie published further work on the language in 1839.

Although unwritten until recently, Somali is the vehicle of a rich oral literature. Poetry is alliterative: verse forms seem to be home-grown,

uninfluenced by Arabic or other foreign forms. Both poetry and prose focus on Somali history and legend. Tales of local Sufi saints are also popular. Traditional bards, some of them now specialising in political poems and propaganda, are prominent in the media.

Close contact and migration between the Horn of Africa and the Arabian peninsula dates back at least two thousand years: over all that time Arabic has influenced Somali. As the vehicle of the Islamic religion, in which the Koran must be recited, Arabic is a highly important language of culture in modern Somali towns but it tends to be known by men only. During the early years of British and Italian occupation Arabic acted as the medium of communication between Somalis and invaders.

On independence in 1960 Somali seemed destined to be the natural language of government, yet no national writing system was agreed on until 1972. A foreign-educated elite, literate in English or Italian, retained influence during this early period: Arabic literacy was also prized.

Until 1972 Somali had been written in several ways: in the Latin alphabet, with various spelling rules influenced by English or Italian; in the Arabic alphabet; and in the *Far Soomaali* or *Cusmaaniya* script which was invented specifically for Somali. Orthodox Muslims often saw the Latin alphabet as a vehicle of Western influence, but Arabic script, poor in its marking of vowels, was unsuitable for the vowel-rich sound pattern of Somali.

The Latin orthography finally adopted in 1972 is easy to write and print, having no accents or

diacritical marks. At the same time Somali was made the sole official language of Somalia, and most civil servants met the three-month deadline for literacy in their mother tongue.

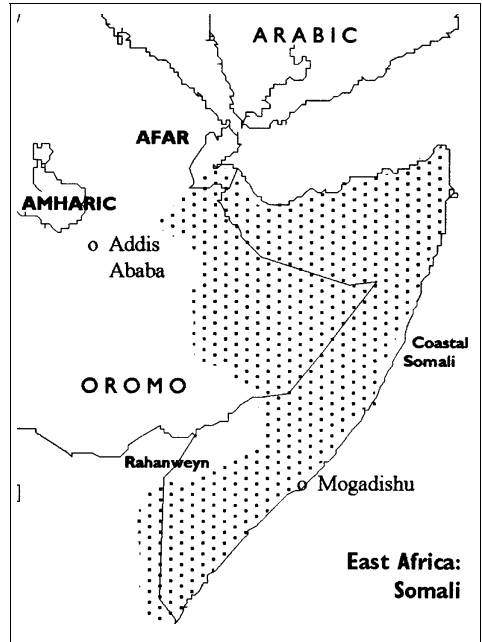
Somali has naturally many Arabic loanwords. Other borrowings include *miis* 'table' from Portuguese *mesa*.

Somali has an unusual vowel pattern of five 'normal' and five 'fronted' vowels. There is also a long-short distinction, giving twenty vowel phonemes in all. The deep-throat pharyngeal consonants, typical of the language, are written *q x*. There are four tones: tone in Somali has grammatical uses, helping to indicate number, gender and case: *Sōōmááli* 'a Somali', *Sōōmāāli* 'Somalis'; *Áràb* 'an Arab', *Ārāb* 'Arabs'.

The first ten numerals are: *ków*, *lábba*, *sáddex*, *áfár*, *shán*, *lís*, *toddobá*, *siddéed*, *sagaal*, *toban*.

### Somali dialects

'Common Somali' is the dialect of most Somalians (notably the *Ishaak*, *Dir*, *Darod* and *Hawiya* people) and of most Kenyan speakers of Somali. There are two other important dialects: 'Central' or *Rahanweyn* Somali, whose speakers are to be found in Upper Juba province, and *Coastal Somali* of Benadir province and of Mandera district of Kenya. Many speakers of these also know



Common Somali, which is the language of broadcasting in Somalia and in Kenya.

About 900,000 Somali speakers live in Ethiopia, 300,000 in Kenya and 200,000 in Djibouti.

OROMO is the closest linguistic relative of Somali. Also related to Somali are Saho and AFAR (see map there).

# SONGHAY

**2,000,000 SPEAKERS**

*Niger, Mali, Nigeria, Burkina Faso, Benin*

**S**onghay, one of the NILO-SAHARAN LANGUAGES though with no close relatives, is a language of trade along the middle Niger, from Mopti in Mali to Gaya on the border of Niger and Nigeria.

The three peoples who speak the language call themselves *Sōngai*, *Zarma* and *Dendi*, or *koyroboro* 'one of the country'; their language is *Sōngai kine* 'language of the Songhay' or *koyra kine* 'language of the country'. In French its name has sometimes been spelt *Sonrhai*.

The Songhay people had been converted to Islam in the 11th century. Some Songhay words are recorded by Arabic authors who had travelled in the Western Sudan in the 12th to 14th centuries.

The importance of Songhay as a language, however, dates from the brief flowering of the Songhay empire under Sonni Ali and Askia Mohammed I (Mohammed Ture) in the 15th and 16th centuries. From its old capital of Gao, in western Mali, the empire now extended its influence from Senegal all the way to northern Nigeria – until, soon afterwards, a Moroccan invasion destroyed it.

Ever since that collapse, Songhay has remained an important means of communication along the middle Niger. It is still learnt as a second language by some of the peoples living along the river. It is an official regional language in Niger, Mali and Benin; the Djerma dialect has separate official status in Niger.

Songhay has two tones and a distinction of vowel length. A noun is marked as definite or as plural by a change of suffix: *gorgo* 'hen', *gorga* 'the hen'; *haw* 'cow', *hawiyān* 'cows'. Although Arabic and Songhay have coexisted for many centuries,

and although many speakers have learnt Arabic as the language of Islam, Songhay does not have a large number of Arabic loanwords: it has, however, borrowed freely from Hausa.

## Songhay and its neighbours

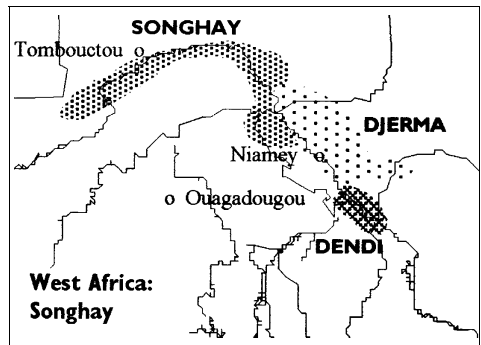
Speakers of *Songhay* as a first language are concentrated in south-western Mali and the modern state of Niger (where the dialect is *Djerma* or *Zarma*) and in northern Benin (where the dialect is *Dendi* or *Dandawa*). The third dialect zone, further up the Niger from Tombouctou to Mopti, is that of Songhay proper. The Zabarima, who raided and eventually settled in northern Ghana in the 19th century, speak a Djerma dialect.

### ... end of the story

ây-cā-bō kā dānjī!

my mouse's head has fallen in the fire!

The traditional sentence with which the Songhay storyteller ends his tale.



### Numerals in Songhay and Djerma

Songhay		Djerma
afo	1	-a fo
ahinka	2	-ihinka
ahinza	3	-ihinza
ataty	4	-itaaci
agu	5	-igu
iddu	6	-iddu
iye	7	-iyye
ahaku	8	-ahakku
yagga	9	-egga, yagga
away	10	-iway

In the newly standard written form of Djerma, numerals are written as suffixes to nouns, linked by the definite suffix *-a*, *-e*, *-i* (the Djerma numerals are from information by Hamidou Ko in M. Malherbe, *Les langues de l'humanité* (Paris, 1995) p. 1662). But the basic numeral forms are practically identical in Songhai proper and Djerma.

# SONINKE

**1,000,000 SPEAKERS**

*Mali, Senegal, Mauritania*

**O**ne of the Mande Tan group of NIGER-CONGO LANGUAGES, Soninke is a language of inland West Africa, at the point where the frontiers of Mali, Mauritania and Senegal meet (for map see MANDEKAN). Soninke is called *Sarakollé* in Wolof, the main language of Senegal: the word means 'light-coloured person' in Soninke. The language is called *Maraka* in Mandekan.

The SUSU and Soninke were among the early enemies of the Empire of Mali in the 13th century. Some Soninke words are recorded by Arabic authors who had travelled in the Western Sudan in the 12th to 14th centuries. European records of the language date back to the 17th

century: under the name *Saracole* it figures in a multilingual glossary made for the French Royal Senegal Company in 1690.

Since their territory was acquired by France (and divided among three administrative units, now independent states) many Soninke have emigrated to France: so many that their language serves as a lingua franca in West African communities in French cities.

Soninke has borrowed from Arabic, the language of Islam, and from French. The first ten numerals in Soninke are: *baane, fillo, sikko, nax-ato, karago, tumũ, ñeru, segu, kabu, tamũ*.

# SORBIAN

100,000 SPEAKERS

*Germany*

**S**orbian marks the far north-west of the SLAVONIC LANGUAGES – the only living linguistic reminder that the north German plain was once partly Slavonic-speaking.

Sorbian speakers call themselves *Serbja*, a name identical with that of the Serbians of former Yugoslavia which perhaps is that of an early Slavonic tribe. The language was traditionally called *Wendisch* in German, a name that goes back two thousand years to the *Venethi* mentioned by the Roman historian, Tacitus, as eastern neighbours of the Germani. *Wendisch* has a pejorative slant and is now less used. In some older books Sorbian is called *Lusatian* after the province where it is spoken.

There are in reality two Sorbian languages, each of which has its own literary form. In the German countryside to the south-east of Berlin the speakers of Lower Sorbian are settled around the city of Cottbus or *Chošebuz*. Further south, in and around the town of Bautzen or *Budyšin*, Upper Sorbian is spoken (for map see POLISH). Although they share a name, the two languages are quite distinct: see, for some examples, the table of numerals at CZECH.

After periods of Polish and Czech rule in the 11th and 12th centuries, Lusatia became part of the German state of Saxony, which traditionally allowed an official role to Sorbian (the Crown Prince was expected to learn the language). The Cottbus region, however, passed to the Kingdom

of Prussia in 1815, and Lower Sorbian speakers found themselves under pressure to turn to German. Later, under Nazism, the use of Sorbian was officially forbidden. Communist East Germany recognised Sorbian as an official language, and it is still used in local schools.

The first major text in Sorbian is a manuscript translation of the New Testament into Lower Sorbian, made in 1548 by Miklawuš Jakubica. The first grammar of Upper Sorbian was published in Prague in 1679. There has been a Sorbian newspaper press for over two hundred years.

In the 18th century Sorbian students were found particularly at Leipzig and Breslau and at the Catholic Sorbian Seminary in Prague: the student societies kept in touch with one another and with the increasing momentum of Slavonic nationalism. Strong local organisations, which have helped to keep the language and literature alive in the 19th and 20th centuries, include the cultural academy *Mačica Serbska*.

Further reminders of the former Slavonic presence in present-day Germany are well-known place names. *Leipzig*, host to international trade fairs ever since medieval times and the second city of former East Germany, was once *Lipsk* from Slavonic *lipa* 'lime tree'. Its neighbour *Dresden*, which suffered so severely in the Second World War, is in origin *Drěždžane*, 'people of the marshy woods'.

# SOTHO

**7,400,000 SPEAKERS**

*South Africa, Lesotho*

**S**otho is the name of two related BANTU LANGUAGES, both of them now official languages of South Africa. Speakers of the dialects called Northern Sotho form the fifth largest language community there, just ahead of speakers of English as a mother tongue. Southern Sotho is the national language of Lesotho and also has a large number of speakers in South Africa. There is a Sotho-Tswana television channel in South Africa.

In Sotho itself the language is called *seSotho* and the people *baSotho*: hence the older English names – *Sesuto* for the language, and *Basutoland* for the now independent country that is called *Lesotho*, ‘country of the Sotho’.

Sotho and TSWANA form one of the two major groups of Bantu languages in South Africa, the other being ‘Nguni’ (see maps here and at ZULU). The two groups, though related, do not make up a dialect continuum, suggesting that their differences crystallised at a time when they were geographically separate – and perhaps one or both groups of speakers were not yet in precisely their present position.

Overall, in recent centuries, Nguni languages have been in the ascendant and Sotho-Tswana in retreat. The separate history – the resurgence – of Southern Sotho comes from the establishment and steady expansion of a powerful kingdom by Moshoeshoe in the early 19th century. In the following decades, parts of this kingdom were seized by Afrikaans and English speakers, becoming the ‘conquered territories’ of the Orange Free State. But the remainder, as Basutoland, was governed as a British protectorate from 1884, and Moshoeshoe’s descendants retained power.

Swiss and French missionaries of the Paris Evangelical Mission Society assisted Sotho diplomacy at this period; they also helped to spread literacy and education. Basutoland regained independence, as Lesotho, in 1966: Sotho and English are its official languages.

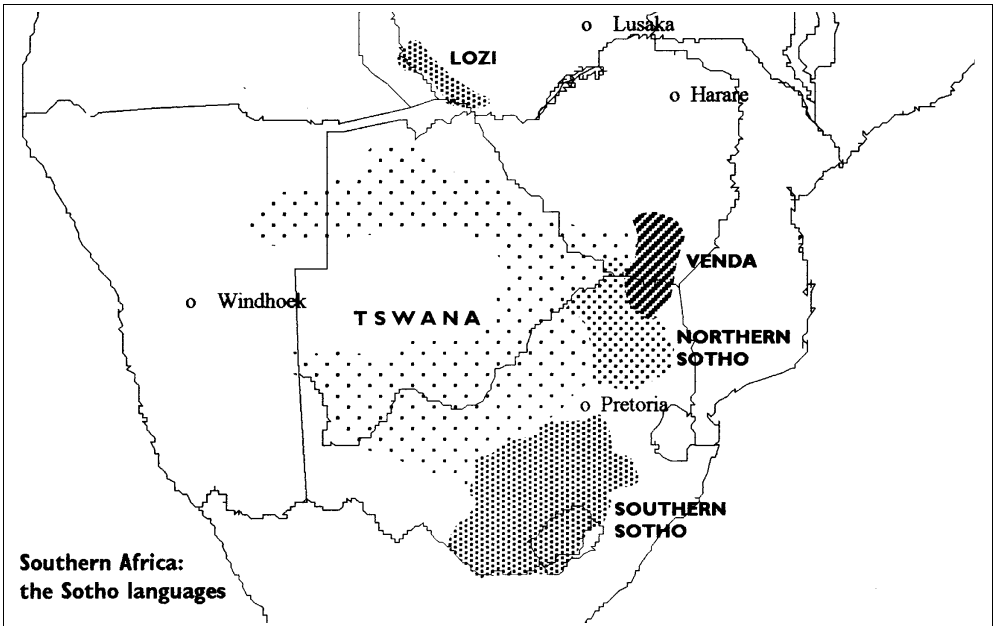
Historically, the Southern Sotho are Moshoeshoe’s people. Sotho is their common language – whatever their origins – and they are traditionally cattle farmers. Lesotho is marked by massive mountain plateaus, but it is in the lowlands, to the west of these, that most Southern Sotho speakers live.

The political dominance of Zulu, during the centuries preceding Moshoeshoe, led to extensive Zulu influence on most dialects of the Sotho and Tswana group. The large numbers of Zulu speakers who became subjects of the baSotho kingdom in the 19th century and eventually adopted Sotho as their language, increased this effect on Southern Sotho in particular. Hundreds of Zulu loanwords are to be found in Southern Sotho: *hosasa* ‘tomorrow’, *haholo* ‘a lot’, *-kgolo-phala* ‘become fat’. Southern Sotho now has click phonemes such as are typical of Zulu – sounds which Zulu in turn had borrowed from Khoisan languages. However, for the three Zulu clicks *c* (dental) *q* (palatal) *x* (lateral), Southern Sotho has only one, *q*.

Southern Sotho, unlike the rest of the group, has developed an ‘avoidance language’, *hlonpho*, used by women in the presence of some male relatives. This practice is widespread in the Nguni languages (see box at XHOSA) and is clearly borrowed by Southern Sotho from Zulu.

Written Southern Sotho uses *o* and *e* not only as vowel signs but also for the semivowels *w* and





y, a vestige of the origins of written Sotho in the work of the French-speaking missionaries in the 19th century. Like other Bantu languages, Sotho has a system of noun classes marked by prefixes: see table at LOZI.

### Lessons in verse

The 'puberty schools' of the baSotho were traditionally the occasion for the recital of didactic songs, learnt in advance from elder brothers and sisters. In this traditional Sotho literature the rules of daily life were transmitted from each generation to the next.

'The opening of the [boys'] school is marked by a formal party at the local chief's meeting place, *lekhotla*. An ox will be presented to the chief and another slaughtered for the party . . . The following morning, the candidates for instruction are initiated into the school by the performance of the collective ceremony of circumcision.

'After these opening ritual activities, the real work of the school begins. The period of instruction is nowadays some two or three months. The boys are taught obedience to their fathers and mothers. They are introduced to the

rules of seniority and respect. Loyalty and obedience to the chieftainship take second place to the importance of correct lineal and family relationship behaviour. The curriculum also includes instruction in sexual morality and the rules of sexual behaviour. All this instruction is conveyed through the medium of songs which the candidates sing all through the night . . .

'Information about the girls' schools is scanty . . . Their instruction concerns itself with matters relating to womanhood, domestic and agricultural activities, sex and behaviour towards men.'

V. G. J. Sheddick, *The Southern Sotho*  
(London: International African Institute, 1953)  
pp. 31–2

### Northern Sotho

Northern Sotho, formerly 'Transvaal Sotho', now in full *seSotho na Leboa*, is a modern concept: a label for several related dialects, one of which, *Pedi*, is the basis of a standard written language. Pedi speakers, historically forming one of several kingdoms in the Transvaal, migrated southwards

as labourers in large numbers and drew the attention of 19th-century missionaries to their language, whose written form was therefore based on the migrants' speech. Speakers of other dialects initially saw no special prestige in it.

For a table of numerals in Northern and Southern Sotho see TSWANA.

### The Sotho languages

*Northern Sotho* has 3,500,000 speakers in Transvaal. *Pedi* is the dialect that gave rise to the first written form of Northern Sotho: its origins are in the Lulu mountains west of Ohrigstad and Lydenburg.

*Southern Sotho* has 2,700,000 speakers in South Africa and 1,200,000 in Lesotho. Local dialects are said to have been more noticeable in the past than they are now.

TSWANA or Western Sotho, although mutually intelligible with Pedi, is usually treated as a separate language. It has 3,600,000 speakers in South Africa and about 850,000 in Botswana.

LOZI, one of the national languages of Zambia, spoken by 450,000 in the upper Zambezi valley, is in origin the Sotho speech of the Kololo, who conquered this region in the early 19th century.

VENDA, spoken in Zimbabwe and the northern Transvaal, is not closely related to any of its neighbours.

#### Sotho praise poetry

Preserving the memory of leaders and battles recent or long past, praise poetry is an important part of Sotho oral literature, as already observed by the French missionary Casalis in 1833. This is the opening of a modern praise poem commemorating Moshoeshe II:

Lona le ratang ho roka baholo,  
le roka hampe le siea mohale,  
le siea Thesele oa Mokhachane . . .

You who are fond of praising the ancestors,  
your praises are poor when you leave out the warrior,  
When you leave out Thesele son of Mokhachane . . .

*Lithoko: Sotho praise-poems* ed. M. Damane, P. B. Sanders (Oxford: Clarendon Press, 1974) pp. 73, 270

# SPANISH

**225,000,000 SPEAKERS**

*Spain, Costa Rica, Cuba, Dominican Republic, Mexico, Guatemala, Honduras, El Salvador, Nicaragua, Panama, Colombia, Venezuela, Ecuador, Peru, Bolivia, Chile, Argentina, Paraguay, Uruguay, United States, Puerto Rico*

**S**panish, one of the ROMANCE LANGUAGES and thus a direct descendant of Latin, is now the major language of the Iberian peninsula, the south-western extremity of Europe. As a result of Spanish discoveries and conquests five hundred years ago, it is also the ruling language of most of America, south of the United States – Mexico border and west of Brazil. In terms of the number of speakers who learn it as a mother tongue, Spanish is the third most popular language in the world after Chinese and English.

The peninsula south of the Pyrenees was called *Iberia* by the Greeks but *Hispania* by the Romans. *H* was not pronounced in later Latin, and a short *i* regularly became *e*. Thus, in Spanish, the country is called *España* and its language *Español*.

*Latin America*, from Mexico to Argentina, is so called because its two ruling languages, Spanish and Portuguese, are descendants of Latin.

As Latin spread across the western Roman Empire it gradually superseded the local languages of prehistoric western Europe. In Spain, inscriptions (as yet not fully deciphered) show that several languages existed, one of them – BASQUE – still surviving in the Pyrenean foothills. Iberian, apparently spoken over much of central Spain, was probably unrelated to Basque. Spanish *huerga* ‘spree’ and *huelga* ‘industrial strike’ both originate from a pre-Roman, perhaps Iberian word \**folga*: *huelga* reached Spanish by way of the regional dialect of Andalusia. Spanish *izquierdo* ‘left’ is also a pre-Roman word, paralleled by Basque *ezkerr*, Gascon *kêrr*.

The Vandals and Visigoths, Germanic invaders who successively occupied Spain after the end of the Roman Empire, contributed rather little to Spanish vocabulary. Far more important was the Islamic conquest of 711–18, after which much of Spain was ruled by an Arabic-speaking elite for five hundred years – indeed the Islamic Kingdom of Granada survived until 1492.

As a result of this long Arabic domination many Arabic loanwords are to be found in Spanish, particularly for aspects of high culture – government, science, architecture, dress and jewellery, music, food. Examples include *alcalde* ‘mayor’, *algodón* ‘cotton’, *almirante* ‘admiral’: this last comes from the Arabic *amīr*, which is also the origin of English *emir* and (by way of French) *admiral*.

It is curious that these Arabic words are nearly always borrowed complete with the definite article, *al-*. Surely, at any period, Spanish speakers could have identified and separated off the Arabic definite article, which is so similar to the Spanish *el*? A possible explanation is that the words were first learnt, in the early days of Islamic rule in Spain, from what was probably a majority of BERBER troops in the Islamic armies. Berber itself has no similar definite article (its nearest approach is the feminine suffix *-t*), so these north African speakers of army Arabic would have learnt the words, complete with article, as if they were indivisible. In this form they will have become part of MOZARABIC, eventually to reach Spanish in the course of the Reconquest.

The Christian reconquest of Spain, *Reconquista*, began from the northern mountains, which the Muslims had never been able to subdue. It was conducted by a row of five small states: Portugal in the west, Catalonia in the east, and between these Leon, Castile and Aragon. A sixth state, Navarre, astride the Pyrenees, took little part. In their historic regions of northern Spain the distinctive dialects of Leon and Aragon can still be heard. But history brought Castile to the fore: thus, as the Mozarabic- and Arabic-speaking people of the south adopted the speech and the way of life of their new Christian rulers, Castilian was the dialect to which they turned.

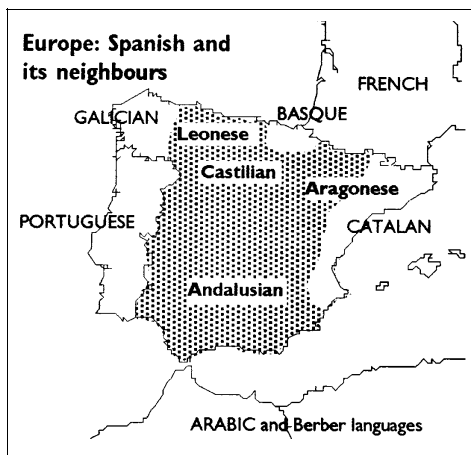
So medieval Castilian is the main source of modern Spanish, and *Castilian* (Spanish *Castellano*) is an alternative name for the language. The name of the old kingdom derives from Latin *castella*, 'the fortresses'.

The first written records of early Spanish are in glosses, translations written between the lines, in Latin religious manuscripts from 10th-century Navarre and Castile. Law codes (*fueros*, from Latin *forum*) began to be translated into Spanish in the 12th century: the early Spanish epic, *Poema del Cid*, was perhaps first written down about the same date, though the surviving manuscript seems to have been copied in 1307. In real life its hero, the 'Cid', captured Valencia from its Islamic rulers in 1092.

The flowering of Spanish prose in the 13th century was largely the work of King Alfonso 'the Wise' of Castile (reigned 1252–84) and his court. Spanish grammars and dictionaries began to appear in the 15th and 16th centuries.

Because of its importance across the world, particularly in exploration and trade, Spanish is the origin or the transmitter of numerous loanwords that are now part of modern English. Examples are *potato*, which comes via Spanish *patata* from the word *batata* 'sweet potato' in a native language of Haiti; *maize*, which comes by way of Spanish *maiz* from Carib, the main native language of the early Caribbean; *canyon* or *cañon*, a special meaning given in the Spanish dialect of New Mexico to a word originally meaning 'tube, pipe' (*caña* 'cane').

For a table of numerals in Spanish see CATALAN.



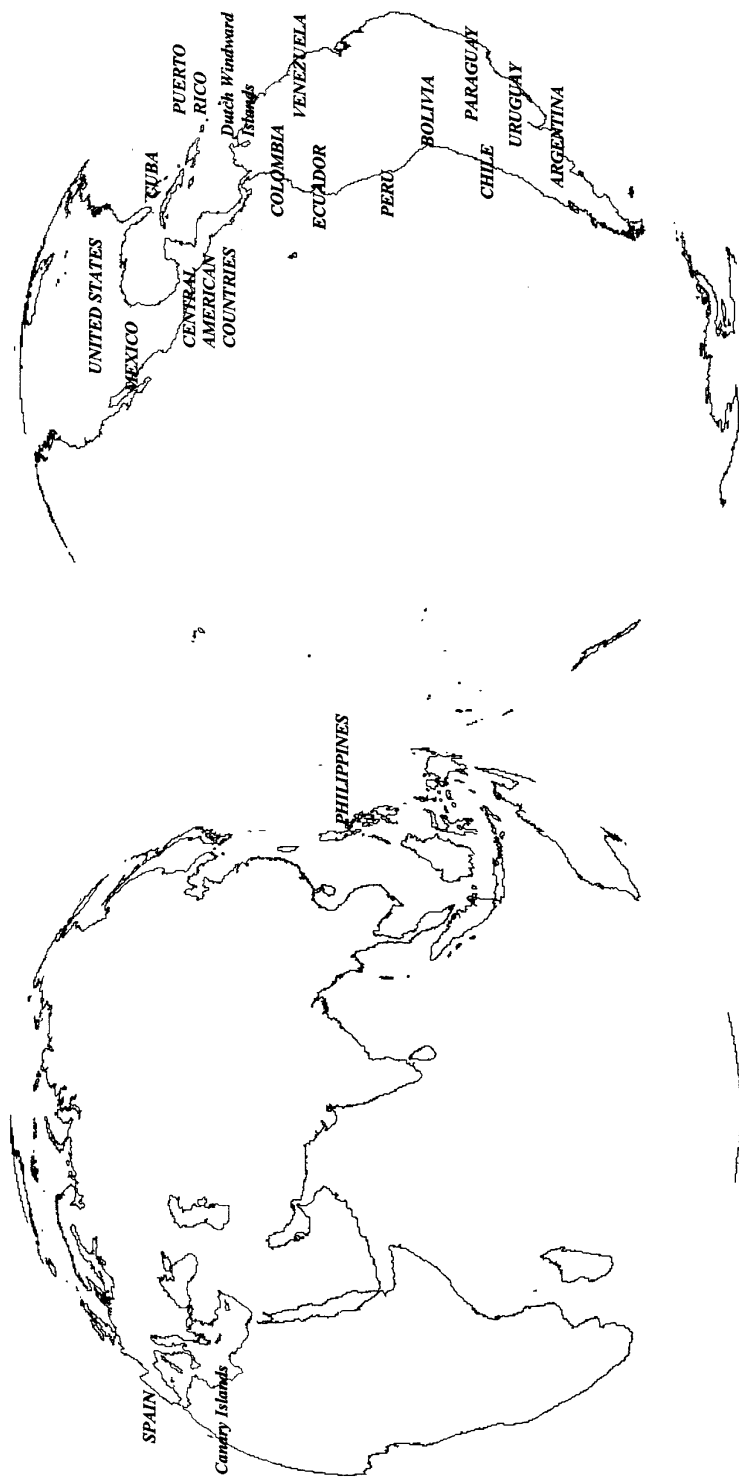
### The borrowing of cultural concepts

Spanish *trabajo* 'work' was borrowed into many American Indian languages – and Nahuatl *tequetl* 'work' was borrowed into Lenca of Guatemala, where the Spaniards used Nahuatl as a lingua franca. Yet American Indians laboured hard before the Europeans came. Why, then, did they need the term? William Bright explains by citing Marshall McLuhan: ' "Work" . . . does not exist in a preliterate world. The primitive hunter or fisherman did no work, any more than does the poet, painter, or thinker of today. Where the whole man is involved there is no work.'

William Bright in *Native languages of the Americas* ed. Thomas A. Sebeok (New York: Plenum, 1976)

### Spanish in Spain, Latin America and elsewhere

The ascendancy of Spanish in the Iberian peninsula comes from the Reconquest, completed in 1492 – and the marriage of Ferdinand of Aragon with Isabella of Castile, in 1479, which made Castilian the ruling language of both Castile and Catalonia. From this date onwards, Catalan fought to survive as a language of literature



Spanish as a world language

and culture. Galician, in the far north-west, has a far less flourishing status than Catalan; Leonese and Aragonese have not in modern times been considered anything more than dialects of Spanish.

The dialects of southern Spain are dialects of Castilian, not of Leonese or Aragonese. Andalusian is sometimes claimed as the immediate parent of American Spanish varieties, but this is an oversimplification: migration to America came from all regions of Spain, and the Latin American dialects reflect this.

It was also in the 1490s that the Jews of Spain (the *Sephardim*) were expelled. Most of them settled in the Ottoman Empire (the majority of their descendants are now in Israel). Since that time, their language has developed independently of its parent Spanish. It is now known as JUDEZMO.

In the same fateful year of 1492 the Genoese adventurer, Christopher Columbus, under Spanish patronage, discovered America. Spanish conquerors, settlers and missionaries were soon spreading across the Americas from Florida ('country of flowers') and California southwards to Chile and Argentina ('silver country'). Alongside NAHUATL and QUECHUA Spanish immediately became an essential lingua franca throughout this huge territory, with the exception of Portuguese-ruled Brazil. Until around 1800 Spanish America was part of a Spanish empire, ruled from Madrid. By the time that independence came, the status of Spanish was unquestioned, and it is now the official language in all Spain's old American territories, with the exception of those that now form part of the United States.

American Spanish has existed in many varieties. At one end of the spectrum are the regional 'standard languages', which differ from the Castilian of Madrid more as regional accents than as true dialects. It is by 'accent' rather than by special choices of words that Latin Americans can most easily recognise one another's country of origin. Each has its peculiarities: Mexican Spanish, because of its many borrowings from Nahuatl and other local languages, has a consonant cluster *tl-*: *tlaco* 'coin'.

At the other end, there are numerous mixed

languages, pidgins and creoles which have grown up and died away in the course of the five centuries of Spanish-speaking rule and settlement: Nahuatl Spanish and Quechua-Spanish mixtures (*Chaupi Lengua* 'half language' and *Media Lengua* 'middle language'); Slave and Caribbean Spanish; the Spanish of Puerto Rico and Cuba with its creole elements. Puerto Rican Spanish is now heavily influenced by American English in its grammar and vocabulary. The local language of the Dutch Antilles, PAPIAMENTU, is generally regarded as a Spanish creole.

Ironically, new mixed forms of European languages are found in Spanish America. *Dialecto Fronterizo* was a mixed Spanish-Portuguese dialect spoken on the Brazil-Uruguay border.

Argentina was for many years a popular destination for Italian emigrants. Around Buenos Aires *Cocoliche*, a pidgin Italian-Spanish, used to be spoken by great numbers of settlers. It was never likely to develop into a stable language or even a creole: after all, true Spanish is relatively easy for Italians to learn. Yet *Cocoliche* has had a long history, for with each new wave of Italian migrants the pool of Italian speech was refreshed. There were many stages of *Cocoliche* between good Italian and good Spanish: speakers soon learnt to add a Spanish plural, *-s*, to their words; it took much longer to adopt an unfamiliar sound like the fricative *j* so typical of Spanish.

North of 'Latin America', Spanish is well known to have served as a lingua franca for contacts between Europeans and Indians in what are now the western United States. It was used, perhaps as early as the late 17th century, for communication among the Pueblo Indians themselves, since their own languages were quite different from one another. Spanish loanwords are to be found in many North American Indian languages. Spanish place names (*Nevada*, 'snow-covered'; *Colorado*, 'coloured, red'; *Los Angeles*, 'The Angels'; *San Francisco*, 'Saint Francis') are the most obvious reminder of Spanish exploration and influence here.

Nowadays there are several varieties of Spanish in the United States. *Pachuco* or

*Pochismo*, which originated as a Spanish-English contact language, is nowadays used as an argot by 'Chicano' young people in Arizona and southern California. *Chicano* – in several variant forms – is spoken in New York and Florida by people from Cuba and Puerto Rico, and in California, Texas, New Mexico and Arizona by people of Mexican descent. Chicano naturally has English loanwords: *flonquear* from American English *flunk* 'fail an examination'; *Quit* from American English *Kid*, often as part of a nickname.

'Tirilones', men of Mexican origin living in El Paso, Texas, use a secret dialect called *Caló* (which is properly the name of the para-ROMANI language of Spanish Gypsies). Originally based on Spanish, the *Caló* of El Paso contains many words that would be quite unfamiliar to speakers of Mexican Spanish, and many words used in new senses. *Caló* seems to adopt new words frequently so that it remains difficult for outsiders. Women do not speak it. Other secret and criminal languages based on Spanish are also called *Caló*.

## Philippine Creole Spanish

Spain ruled the Philippine Islands for over 300 years. Yet Spanish was never very widely used there: the Philippines used TAGALOG and other languages in long distance trade and there was little need for another. By the time the United States took power in the Philippines, only about one-tenth of the population knew Spanish.

However, some communities did develop a Spanish creole, known as *Chabacano*, and a quarter of a million people still speak it. On the south side of Manila Bay Chabacano is spoken in Cavite City and in Ternate: this is the point of origin of the creole, for Ternate (named after one of the Moluccas) was founded by migrants from the Moluccas, under Spanish domination, in the 17th century. The original migrants evidently spoke Portuguese pidgin, and its influence can still be traced in Chabacano.

When the Spanish took far-off Zamboanga on

the western tip of Mindanao in 1718, it must have been settled by people from Manila Bay and also by HILIGAYNON-speaking troops from the islands of Panay and Negros. The creole called *Zamboangueño*, which is the language of the Christian communities of Zamboanga region and of Basilan island, is still identifiably a variant of Chabacano, but it shows strong influence from Hiligaynon and Tagalog, both spoken far to the north. *Zamboangueño*, with over 150,000 speakers, is one of the trade languages of the archipelago between Mindanao and Sabah.

### The naming of accents

Spanish speakers have a range of terms to identify special features of one another's pronunciation and vocabulary.

**Seseo and ceceo:** Seseo is the pronunciation of *c*, before *e* or *i*, like *s* and not like English *th* (which is the pronunciation in Castilian and the standard Spanish of Madrid). The *s* pronunciation is widespread in Spain and Latin America. Ceceo is what an English speaker would call 'a lisp': the pronunciation of *s* with the *th* sound. In a sense it is the converse of seseo. It is typical of the Spanish of Seville and Cadiz.

**Voseo:** The use of *vos* instead of *tú* as the singular equivalent of 'you': most Latin American speakers do this, but not those in northern and central Mexico, in northern and central Peru or in northern Uruguay.

**Yeísmo versus lleísmo:** The use of a *y* sound for the written *ll* of Spanish is now much more widespread than the traditional 'palatal *l*' (the nearest English sound to this is *ly*). Yeísmo is gaining ground rapidly in both Spain and Latin America, but it has not reached speakers from the Andean highlands (all the way from Colombia to northern Argentina).

**Žeísmo:** The use of a *zh* sound instead of standard *y* and sometimes instead of the 'palatal *l*'. This is typical of Central America, south-eastern Mexico, Ecuador, Paraguay, Uruguay and part of Argentina.

**Value judgments?**

In Zamboangueño, the Spanish creole of Christian communities in Mindanao, a scale of moral judgments was once built into the vocabulary: favourable words seem to come from Spanish, unfavourable from Philippine languages.

grande	large	dyutay	small
alto	tall	pandak	short
lihero	fast	mahinay	slow
agudu	sharp	mapurul	dull
liso	smooth	makasap	rough
dulse	sweet	mapa'it	bitter
manso	tame	ma'ilap	wild
bunito	pretty	umalin	ugly
derecho	straight	tiku'	bent
byeho	old person	bata'	child
plores	blossom	putut	bud
soltero	unmarried boy	dalaga	unmarried girl

After Charles O. Frake in *Pidginization and creolization of languages* ed.  
Dell Hymes (Cambridge: Cambridge University Press, 1971) p. 232



# SUKUMA

4,000,000 SPEAKERS

*Tanzania*

**S**ukuma is the most important of a closely related group of BANTU LANGUAGES spoken in central and south-western Tanzania. It is in a linguistically varied neighbourhood, with Khoisan, Cushitic, Nilotic and Nilo-Saharan languages in the vicinity.

In the 19th century, when German missionaries and administrators were exploring Tanganyika, Nyamwezi was the first of this group of languages to be noticed. The Nyamwezi themselves were long distance traders, and other traffic passed through their territory. Sukuma, with a far larger number of speakers but in a less explored region, was regarded as a dialect of Nyamwezi, and its name *kiSukuma*, 'language of the north', describes it from the Nyamwezi point of view.

Sukuma now has by far the largest number of speakers of any regional language of Tanzania. But more and more Sukuma speakers are becoming bilingual in Swahili, the national language of Tanzania.

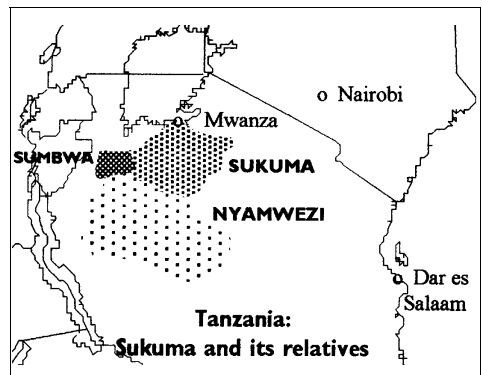
Sukuma has a remarkable tonal pattern. In most words one syllable has, grammatically, a high tone: but this high tone is heard not in its own place but one, two or (most often) three syllables further on in the spoken sentence. Thus the three words *akabóna*, *babítì*, *bataále*, have a quite different tone pattern when spoken in a sentence: *akabona bábiti batáale*, 'he saw tall passers-by'. In Nyamwezi, by contrast, 'the intonation is different', as local speakers say: the high tone only shifts one syllable on.

Example from Derek Nurse, *Description of sample Bantu languages of Tanzania* (African languages vol. 5 no. 1, London: International African Institute, 1979) p. 48

## People who shit the moon

'The term Nyamwezi is of Swahili origin, and is fairly recent. It arose in the last century during the trade caravans. My grandfather told me that in those days a caravan would leave Tabora at new moon to arrive in Bagamoyo or at the Dar es Salaam coast at the following new moon. Since this was a regular occurrence, the Zalamo started teasing the caravanists, calling them 'the people who excrete the moon', *wanyamwezi* (from the verb *ku-nya*) because their arrival at the coast nearly always coincided with the new moon. Apparently, since there had already developed a joking relationship, *utani*, between the Zalamo and the people from Tabora, the term was not resisted. Since my grandfather did in fact take part in the trade caravans, I have every reason to consider his explanation a viable one.'

C. Maganga in Nurse (1979) p. 57



**Sukuma and its relatives:  
Bantu languages of western Tanzania**

Sukuma, Nyamwezi and Sumbwa are best considered dialects of one another. They are locally regarded as separate languages, though they are so close that speakers of any one have little difficulty with the others.

*Sukuma* itself, with 4,000,000 speakers, is usually divided into four dialects: *kiMunasukuma*,

language of the northerners; *kiMunangweli*, of the westerners; *kiMunadakama*, of the southerners, closest to Nyamwezi; *kiMunakiya* or *kiMunantuzu*, of the easterners.

*Sumbwa* has 190,000 speakers. It is unlike the other languages of the group in having five rather than seven vowels. It shows signs of influence from Zinza, a Bantu language of the Rutara group, to the north.

# SUMERIAN

## Extinct language of Iraq

**S**umerian is the oldest human language on record, its earliest texts dating from 3100 BC. It is a language with no known relatives, living or extinct.

Once the ruling language of an ancient culture in southern Iraq, Sumerian perhaps ceased to be spoken as early as 2000 BC. But it continued to be written for another two thousand years. The Assyrian king Ashurbanipal, in the 7th century, was proud of his ability to read classical Sumerian. After that it was forgotten, and the very name of Sumer was forgotten.

When AKKADIAN cuneiform texts were gradually deciphered, in the 19th century, scholars realised that an even older language was to be found on some of these tablets. Fortunately, this mysterious language had been difficult for the Akkadian scribes to learn: so they had compiled parallel texts and glossaries, and these have been extremely helpful in the work of decipherment. It was eventually realised that the language belonged to an ancient city state, Sumer, in southern Iraq. It was therefore given the name Sumerian.

The Sumerian language, and the Sumerian cuneiform script, can be seen to have developed gradually over their eleven hundred years of active life. At first used mainly for business and administrative texts, the script came to be used after 2500 BC for royal inscriptions and for religious poetry. Two hundred years later the Akkadians conquered Sumer. From that point onwards the old language was less used in government, while its script was adapted for Akkadian – and eventually for other languages of the Near East. But most Sumerian literature – including the *Epic of Gilgamesh* and other myths and legends – survives to be read today in texts that were written many centuries after this, and long after Sumerian had ceased to be an everyday spoken language.

As the classical language of a highly developed state, Sumerian had a considerable influence on Akkadian. Sumerian words were borrowed into Akkadian, and, even more frequently, Sumerian written word forms ('Sumerograms') were made to stand for the equivalent Akkadian words in cuneiform writing. Sumerograms are found in Hittite and other cuneiform texts too.

It is not surprising that no relatives of Sumerian have been discovered. Nearly all the languages of the world of 3000 BC are now completely unknown. Not long afterwards, at any rate, it is clear that the Near East was linguistically highly fragmented, for Hittite, Urartian, Akkadian and Elamite represent four completely separate language families, and Sumerian simply makes a fifth.

### How we read Sumerian script

'Akkadian is a Semitic language, but Sumerian is a language isolate of a very different type, and with a very different phonemic inventory. The values we give to cuneiform signs in Sumerian texts are based on Akkadian values and on ancient glosses. Since most of these glosses date from periods when Sumerian was no longer spoken, i.e. from a milieu speaking Akkadian or other Semitic languages, it is said that we view Sumerian phonology through Akkadian glasses. However, since the signs used to write Akkadian had been adapted from an originally Sumerian system of cuneiform writing, we might also say that our Akkadian glasses were made by Sumerian opticians.'

Jerrold S. Cooper in *The world's writing systems* ed. Peter T. Daniels, William Bright (New York: Oxford University Press, 1996)

p. 37

Sumerian was an ‘agglutinating’ language, one in which a basic word form is given a series of affixes to indicate its function in the sentence. Two main dialects can be recognised in the language written down in the tablets: *eme-gir* is the administrative language, while *eme-sal* is

used for hymns and chants. It has been suggested that *eme-sal* originates not as a regional dialect but as the distinctive speech of Sumerian women.  
The first ten numerals in Sumerian were *desh*, *min*, *pes*, *lim*, *i*, *i-ash*, *i-min*, *i-ush*, *i-lim*, *hu*.

**Once upon a time**

Storytellers in many languages begin with a set phrase or sentence, like English ‘Once upon a time’, that draws listeners’ attention. This stylistic feature can be traced five thousand years back, to the opening of Sumerian tales such as *Ašnan and her seven sons*:

U <sub>4</sub> re u <sub>4</sub> re na-nam	It happened in those days, in those days,
gi <sub>6</sub> re gi <sub>6</sub> re na-nam	It happened in those nights, in those nights,
mu re mu re na-nam . . .	It happened in those years, in those years . . .

In standard Sumerian transliteration, subscript numbers (*u<sub>4</sub>* and *gi<sub>6</sub>* above) are used to mark duplication in the script: the scribe used the fourth alternative form of *u*, the sixth alternative for *gi*.

After Jeremy Black, ‘Some structural features of Sumerian narrative poetry’ in *Mesopotamian epic literature: oral or aural?* ed. Marianna E. Vogelzang, Herman L. J. Vanstiphout (Lewiston: Edwin Mellen Press, 1992) pp. 71–101

# SUNDANESE

**27,000,000 SPEAKERS**

*Indonesia*

**S**undanese belongs to the group of major AUSTRONESIAN LANGUAGES of the Greater Sunda Islands. It is the everyday speech of the western third of the island of Java, and, in terms of numbers, is Indonesia's third language, spoken by nearly one-sixth of the national population. But Indonesian (MALAY), not Sundanese, is the most used language of Jakarta, the capital of Indonesia, which thus forms a large enclave in Sundanese territory.

Sundanese has for many centuries had a history very like that of JAVANESE (see map there), which is spoken in central and eastern Java. Yet

there has been much argument among linguists as to whether Sundanese belongs with Javanese, or rather with Malay, in the family groupings within Austronesian.

Like Javanese, Sundanese shows the effects of Sanskrit, Malay and Arabic influence – and also has many loanwords from Javanese itself. Sundanese has the separate 'formal' and 'informal' speech registers typical of neighbouring languages. As with Balinese and Madurese, the formal vocabulary consists very largely of Javanese loanwords.

## Numerals in Javanese and Sundanese

	Javanese Ngoko	Javanese Krama	Sundanese
1	siji	setunggal	hiji
2	loro	kalih	dua
3	telu	tiga	tilu
4	papat	sekawan	opat
5	lima	gangsals	lima
6	enem	enem	gänăp
7	pitu	pitu	tujuh
8	walu	walu	dalapan
9	sanga	sanga	salapan
10	sepuluh	sedasa	sapulu

## Sundanese in writing

Sundanese has a long written history. It has been written in three quite different scripts. The *aksara* or alphabet of Indian type is nowadays

identical with that of Javanese, though the spelling rules are different. Arabic script is also used. But commonest now is the Latin alphabet, which uses an orthography similar to that of Indonesian.

# SUSU

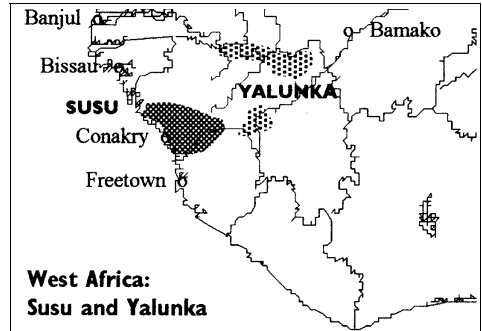
**700,000 SPEAKERS**

*Guinea, Sierra Leone*

**O**ne of the NIGER-CONGO LANGUAGES of the Mande Fu group, Susu is a coastal lingua franca of Guinea and the main language of the capital, Conakry. It is also spoken on the north-western borders of Sierra Leone.

The Susu, an inland people who have fairly recently migrated towards the Atlantic, were among the early enemies of the Empire of Mali in the 13th century. In the 18th century Susu had reached the coast, being spoken as far south as the Sierra Leone estuary. Susu speakers came under the British sphere of influence on the establishment of Sierra Leone Colony in 1788, but most Susu territory was acquired by the French in 1889 and is now in Guinea.

Susu speakers are largely Muslim, and the language shows Arabic as well as French influence. The first ten numerals are: *keren, firin, sakhan, nani, suli, seni, solo masakhan, solo manani, fu*.



## Susu and Yalunka

*Susu* (in French *Soussou*) is spoken in Guinea and northern Sierra Leone.

*Yalunka*, closely related to Susu, has 30,000 speakers further inland on the borders between Guinea and Sierra Leone.

# SWAHILI

## PERHAPS 4,000,000 SPEAKERS OF SWAHILI AS A MOTHER TONGUE

### *Tanzania, Kenya, Congo (Kinshasa)*

One of the BANTU LANGUAGES, and strongly influenced by Arabic, Swahili is probably the most widely spoken indigenous African language. Swahili is the official language of Tanzania, and it is the most widely spoken language in Kenya and Uganda. Well over 40,000,000 speakers use Swahili either as a first or a second language.

The name *kiSwahili* means 'coastal language' from Arabic *sawāḥil* 'coasts'. Until very recently, this is precisely what Swahili was – a language of the coastal peoples, which, from its geographical position, tended to absorb foreign influences and to be used more and more widely in trade. The language group to which Swahili belongs is sometimes called *Sabaki*, a name suggested by Christopher Ehret, from the Sabaki river of eastern Kenya.

The Swahili-speaking Muslim trading communities considered themselves Arabs and they were often so called by Europeans in colonial times. The Portuguese called them *Moros* and the English *Moors* or *Black-moors*.

*Zanzibar* was once the name of the whole coast, as well as of the island. It is an Arabic form of Persian *zangī-bār*, 'country of the negroes'. *Zingis* and *Zanj* are recorded as the name for the islands, the coast and their empire from the 2nd century onwards.

From a postulated prehistoric Eastern Bantu centre of dispersal, in the Great Lakes region, early Sabaki dialects were probably beginning to spread into their present locations in the first centuries AD. It was some centuries after that when northern Mozambique and the Comores

were reached, as archaeological evidence confirms.

As that settlement was continuing, the speakers of one Sabaki dialect, 'proto-Swahili', were already becoming ubiquitous in the coastal trade: they, or at any rate their language, began to radiate widely from their original excellent location at Zanzibar and Pemba. Thus the oldest island Swahili dialects are probably just as old as Mwani and Comorian – but they have differentiated less from the parent language, because their communities tended to remain in close touch.

East African seaborne trade has been, for two thousand years, dominated by people of Arabic culture, the actual languages involved in this trade being Persian, Gujarati, Konkani, Swahili itself and, most honoured, Arabic. According to Greek sources, around AD 100 Arabs were already established at the port of Rhapta (Dar es Salaam), where they ruled 'by some ancient right' and were already familiar with the local language, whatever it then was. Islam was introduced very early. The oldest surviving Arabic inscription of this coast dates a mosque at Zanzibar to 1107. By this time the Zanj Empire, centred on Zanzibar, dominated the trade of the western Indian Ocean – and it is said that Swahili influence can be traced in the Arabic dialects of southern Arabia and coastal Iraq.

Until around 1700 Arabic was the only *written* language of this coastal, island and ocean culture – a culture in which literacy was already relatively widespread. The first known manuscript of Swahili poetry, an epic, dates from 1728. It is naturally in Arabic script. The Latin script was introduced and Swahili printing began, under

European influence, in the 19th century: the first newspaper, *Habari ya Mwezi*, a missionary publication, appeared at Magila in 1895. Arabic script is still very widely used.

At first concentrated on coastal islands, Swahili-speaking traders began to found more coastal settlements in the 13th century (Mombasa is one). When the Portuguese arrived, in 1498, Arabic and Swahili were important languages all along the coast. Alongside the spread of everyday Swahili, Koranic schools and the religious use of Arabic spread also. Many Swahili speakers of this heartland are descendants of slaves, including, probably, a large proportion of YAO and NYANJA origin. Immigration of Arabic speakers may have reached a peak in the 18th century: their descendants also now speak Swahili.

Tabora, founded about 1820, was a staging post on the longer inland routes. During the 19th century, and maybe earlier, Swahili-speaking slave traders and other adventurers crossed Lake Tanganyika and ranged beyond as far as the Lualaba valley. The name of Tippu Tip, who made trading expeditions into Congo (Kinshasa) from 1870 to 1884, is rightly linked to the early expansion of Swahili in Congo, though the language was already known there. His Muslim Swahili-speaking warriors, who eventually formed colonies in what became the Congo Free State, were known to the Belgians as *Arabisés*, to themselves and others as *wangwana* 'free men' – hence the modern name *kiNgwana* for Zaire Swahili, 'language of free men'. Here too, Islamic schools were founded and Arabic has its place as the special language of religion.

In Tanganyika, under German occupation from 1884, and in the whole of East Africa as it fell under British and Belgian rule, Swahili was taken up as the most convenient available means of communication with inland peoples, at least some of whom already knew the language. Far fewer learnt German or English. In British times Swahili came to replace the various mother tongues of Tanganyika as the language that the administration prescribed for school use.

Thus when independence came Swahili was the natural choice as the national language of Tanzania.

## National varieties and creolised forms of Swahili

Nowadays Swahili is learnt as a first language on the islands of Zanzibar, Lamu, Pate and Pemba and by a fairly small number of people on the Kenya coast. More important, it is the first language for a rapidly growing number of city-dwellers in Tanzania. More important still, it is the language that all Tanzanian schoolchildren need to know. It remains true that most of the speakers of Swahili learn another African language first, but more and more now use Swahili for preference. All over East Africa, the learning in childhood of at least three languages – a local language, Swahili and English – is the usual rule.

Within Tanzania the older, coastal dialects of Swahili, particularly those of the islands, persist. New regional varieties, influenced by various inland languages and reflecting cultural differences, can be seen emerging, but they will remain closer to the prestigious standard language of education and the media.

The colonial British of Kenya took a completely different view of Swahili: they found it a useful tool of domination, but despised it. Linguistic surveys of the early post-colonial period retained this view, evaluating Swahili as a useful language for Kenya, but not one that carried prestige. Later history has shown the emptiness of this inherited viewpoint. Kenya, too, has adopted Swahili as its national language.

A variety of Swahili that spread as a lingua franca in Kenya and Uganda was once characterised as 'Up-Country Swahili', *kiSwahili cha bara*. 'Up-Country Swahili' shaded into a pidgin language, Kenya Pidgin Swahili, once very widely used by speakers who had no opportunity to learn a more standard variety – and into a special pidgin, *kiSettla*, that was used between the European settlers of Kenya and their African employees. *KiVita*, 'war language', was the jargon of the British East African army.

*Cutchi-Swahili* or *kiHindi* has been used as the name of the speech variety of South Asians in Kenya, its name recalling their predominant origin from the Cutch peninsula. Another pidgin Swahili, *Barracoon*, a mixed language with



elements of Arabic, Portuguese and Malagasy, was used in the Mozambique ports in the 19th century. It scarcely survives today. But a form of Swahili nearer to the current standard remains a lingua franca of trade in north-western Madagascar and in northern Mozambique, where the *kiMwani* dialect is influenced by local MAKONDE.

Swahili was briefly a school language of Uganda. Uganda's British rulers once planned to make it the local official language, but abandoned an idea which, to the speakers of Ganda, was unpleasantly reminiscent of the settler ethos of Kenya. Swahili is still the language of the Ugandan army and police force, and an essential lingua franca in Uganda generally.

Swahili is important in trade in Rwanda and Burundi, a position it achieved as a result of the 19th-century expansion of trade from the East African coast. Rather surprisingly, this position has not yet been undermined either by French, which was the language of power here for nearly a century, or by the local languages (see RUNDI) common to almost the whole population.

Zaire Swahili, known locally as *Ngwana* or *kiNgwana* or *Kiungwana*, is the most extensive lingua franca of eastern Congo, important in the media and in everyday contact between people of different languages. It is a distinct regional variety of Swahili, heavily influenced by Congolese languages and differing considerably from the standard Swahili of Tanzania.

In Lubumbashi (Elisabethville), the capital of the Shaba region of Congo, because of the early recruitment of mine workers from north-eastern Congo, Rwanda, Burundi and East Africa, a distinct form of Swahili, perhaps a true creole, has become the city's lingua franca. Nowadays more workers come from Luba-speaking districts, but they adopt Lubumbashi Swahili as their second language. It is used in the popular press, in plays, and especially among schoolchildren. The standard Swahili of Tanzania is difficult for Lubumbashi speakers even to understand.

The first ten numerals in Standard Swahili are: *moja*, *mbili*, *tatu*, *nne*, *tano*, *sita*, *saba*, *nane*, *tisa*, *kumi*. It is a sign of Swahili's history as a language of trade that three of these ('6', '7', '9') are

borrowed from Arabic, thus abandoning the quinary counting system, more usual in Bantu languages, in which 'six' is formed as 'five plus one' or 'over to the right hand'. The three borrowed numerals are invariable: the smaller numerals, as usual in Bantu languages, must agree in class with the noun that they qualify. Thus *mayai matatu*, 'three eggs'; *vitabu vitano*, 'five books'.

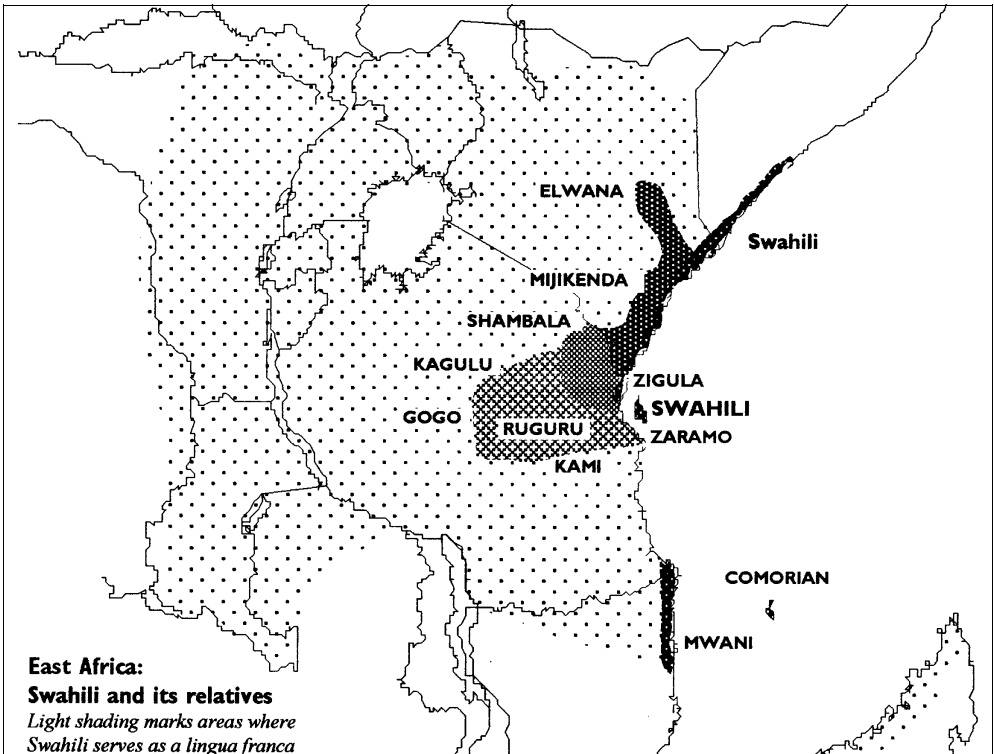
The Swahili word for 'book' in this example is borrowed from Arabic, like many cultural and religious terms: Arabic *kitāb* fitted well with the Swahili 'inanimate' noun class, giving singular *kitabu*, plural *vitabu*. Swahili also has numerous Portuguese loanwords, for example *gereza* 'prison' from Portuguese *igreja* 'church'. English loanwords include *buluu* 'blue', *kilabu* (plural *vilabu*) 'club, nightclub'. Swahili has naturally borrowed heavily from MALAGASY, the language of the southern island of Madagascar with which there has been constant interchange and trade: *divai* 'wine' is from Malagasy, where it derived originally from the French *du vin* 'some wine'.

The rapid spread of Swahili is partly due to slavery and the need for a lingua franca among communities of captured slaves, *watumwa*, home-bred slaves, *wakulia*, freed former slaves, *wahuru*, and their descendants, *wahadimu*. One of the Swahili dialects of Zanzibar takes its name from the *wahadimu*. *Wahuru*, with a different prefix, becomes *uhuru*, 'freedom' or 'land of the freed', a potent political catchword at the end of the colonial period.

### Bantu languages of the East African coast

Some *North East Coast Languages* differ from the rest of the Bantu family in their sound pattern: Zalamo and Swahili, for example, have no tones, but instead a stress accent.

Asu or *Pare*, known to its own speakers as Chasu, has 300,000 speakers in Pare district. A northern and a southern dialect can be distinguished.



### 'God' and 'Gods'

'The Swahili prefer to use the Arabic word for God, *Ilahi* or *Allahu*. There is a word of Bantu origin – *Mungu* – which has been used exclusively by Christian missionaries, who naturally did not feel inclined to use an Islamic term. Mohammedans would point out, however, that *Mungu* has a plural *miungu* "gods, i.e. idols". A plural of Allah is absolutely inconceivable and this may explain the preference of the Swahili Islamic writers for this word.'

Jan Knappert, 'Contribution from the study of loanwords to the cultural history of Africa' in D. Dalby and others, *Language and history in Africa* (London: Cass, 1970) pp. 78–88

The Ruvu group includes Gogo, Kagulu, Kami, Ruguru and Zalamo. *Gogo*, with almost 1,000,000 speakers, is the language of the Dodoma district, with three recognised dialects: *ciNyambwa* to the west, *ciNyaugogo* near Dodoma,

*ciTumba* to the east. *Kagulu* has 225,000 speakers. *Kami* has 300,000 speakers south of Dar es Salaam. *Ruguru* or *Lugulu* has 500,000 speakers in the Uluguru mountains. *Zalamo* or *Zaramo* has 450,000 speakers in the country south, west and north of Dar es Salaam.

The Seuta group includes Ngulu, Shambala and Zigula. *Ngulu* has 130,000 speakers. *Shambala* or *Shambaa* has 500,000 speakers in the Usambara mountains. Shambala is notable for its complex verb forms. A verb may have around eighty combinations of infixes marking tense and aspect, in addition to negative markers and personal prefixes: *mwadika*, 'you cooked'; *nemudike* 'you will cook'; *titazadika*, 'we will cook (contrary to assumption)'; *watavyadika*, 'these will cook (by contrast with others)'; *hemuzedika*, 'you did not cook'. *Zigula* has over 300,000 speakers north of Dar es Salaam.

The Sabaki group is a chain of languages running roughly north–south, from the Tana river valley of Kenya to the Comores Islands: Elwana, Mijikenda, Swahili, Mwani and Comor-

ian. The larger languages are Mijikenda, Comorian and Swahili itself. *Mijikenda*, comprising *Nyika* (a subgroup including *Giriama* and *Digo*) and *Pokomo*, has nearly 1,000,000 speakers in eastern Kenya and Tanzania. There has long been a tendency for speakers of these languages to shift to Swahili, so statistics are difficult to fix. *Comorian* or Komoro has 450,000 on the Comores archipelago. *Ngazija* or Shingazidja is the dialect of Grande Comore, while *Nzwani* is the dialect of Anjouan. The points of origin of *Swahili* itself are Pemba and Zanzibar islands.

### The spread of Swahili

The older Swahili dialects are coastal ones. The language was once spoken at ports and harbours all the way from Mogadishu in Somalia to Sofala (Beira) and Chibwene in middle Mozambique.

Surviving coastal dialects are marked with dark shading on the map.

In the 19th century the language spread rapidly inland, at first with Swahili trade and conquest, afterwards with German and British administration and the migration of workers. Shading indicates the regions where it is regularly used today.

Swahili is the official language of Tanzania, the usual medium of secondary education, government and broadcasting. *Unguja* or kiUnguja, one of the three dialects of Zanzibar island, owing its origin to the highly mobile population of 19th-century Zanzibar, is the basis of Standard Swahili. The other two dialects spoken on Zanzibar, *Hadimu* ('language of the descendants of slaves') and *Tumbatu* ('of Tumbatu Island'), have a more archaic character.

### Spelling it out

Swahili, language of a rapidly growing population and for many speakers a second or third language, has far fewer complex verb forms than its relatives. While Shambala uses affixes to show verbal 'mood', Swahili tends to use separate adverbs or auxiliary verbs:

#### Shambala

**Tedika** inu manga

Ziya mbogha taaz**azidika**

Uja mushi **netidikiye** manga

**Please** cook this cassava

We **really did** cook the vegetables

That day we **actually** cooked cassava

#### Swahili

**Hebu** pika huu mhogo

Zile mboga **tulija** zipika

Siku ile **tulikuwa** tumepika mhogo

Examples from Ruth M. Besha, 'Mood in Bantu languages: an exemplification from Shambala' in *Studies in Tanzanian languages* (Tokyo: ILCAA, 1989) pp. 205–22

The further simplification typical of pidgin languages can be seen in Kenya Pidgin Swahili in the question words. A pattern already found in Standard Swahili (first two examples) has been extended to the whole range:

#### Kenya Pidgin Swahili

saa gani?

namna gani?

siku kani?

kitu gani?

sababu gani?

'hour which?' = when?

'kind which?' = how?

'day which?' = when?

'thing which?' = what?

'reason which?' = why?

#### Standard Swahili

saa ngapi?

namna gani?

lini?

nini?

kwa nini?

Examples from Bernd Heine in *Readings in creole studies* ed. Ian F. Hancock (Gent: Story-Scientia, 1979) p. 95

# SWAZI

1,600,000 SPEAKERS

*South Africa, Swaziland, Mozambique*

One of the Nguni group of BANTU LANGUAGES, Swazi or Swati is the national language of Swaziland, where there are 600,000 speakers. It also has about 900,000 speakers in neighbouring districts of South Africa, where it is one of the eleven official languages (see map at ZULU). Swaziland is a country of rocky highlands, *inkangala*, and of bush-covered lowland plains, *lihlandza*, which are cut off by the Lubombo mountains from the Mozambique coast.

When Afrikaans- and English-speaking invaders first approached the country it was ruled by Mswati II (1839–68). Their Zulu contacts called its people *emaSwazi*, ‘people of Mswati’. The modern names for the language and the country derive from this.

*Swati* and *siSwati* are the forms now preferred by linguistic purists, since *Swazi* with a *z* was of Zulu origin; but, as names for the language, all these are foreign. The speakers have traditionally called themselves *ebantfu bakaNgwane*, ‘people of Ngwane’, the founder of the monarchy, and their language *siNgwane*.

It is probable that Nguni language speakers have occupied the region of modern Swaziland for about a thousand years. Cattle are their traditional wealth: ‘epigrams, symbolism, riddles and praise songs are built around them; the king is honoured as the Bull of the people’ (H. Kuper).

The ruling family of Swaziland traces its origin to a migration from the north in the late 15th century, and first occupied a part of Swaziland under Ngwane III, who died c. 1780. His burial site in a cave in a tree-covered hill at Lobamba is a place of pilgrimage. His grandson Sobhuza I

(ruled 1815–39) settled near the modern capital, and established a powerful monarchy which managed to survive the damaging land seizures of the late 19th and early 20th centuries.

About a third of the kingdom escaped incorporation in South Africa and remained a British protectorate until independence in 1968. Here, few speak English, though many know some Zulu. The Swazi-speaking people of the rural Transvaal have tended to retain their language, and their emotional ties with the monarchy, while those more widely spread across South Africa, especially city-dwellers, may be gradually absorbed by majority Sotho, Zulu and Xhosa speakers. In the Swazi-speaking districts of Mozambique Tsonga is the dominant language.

Swazi, like the other Nguni languages, has the click consonants that are otherwise typical of KHOISAN LANGUAGES. Their presence indicates close interaction with speakers of one or other of those languages in the past.

## Praise poetry and legendary history

Swazi praise poetry is composed in a special Zuluisd literary dialect, while prose myths and legends are typically told in a purer form of Swazi.

‘The main literary productions are the praise poems, *tibongo*, developed in each reign round the character and activities of the rulers, and recited, together with those of their predecessors, on national occasions. Praise poems of other personalities have a local circulation. Animals, particularly cattle . . . have also provided suitable subjects. There may be one or more versions, and the individual composer is often

not known . . . Official praisers are chosen by the rulers for their memory and dramatic abilities, and are listened to with admiration.

'Swazi prose is in the form of myths, legends, fables and tales, handed down from generation to generation. Swazi myths of origin are few, simple, and generally accepted. Legends dealing with clan and tribal history are often complicated and usually recalled by the old men for specific legal or ritual purposes; fables and tales are reserved for entertainment in the evening, and are told mainly by old women. In the tales, the main characters are often of royal birth; in the fables, the most popular animal is the hare. While tales and fables are told essentially for entertainment, they may also point a moral. For this purpose, however, the Swazi . . . have a wealth of proverb and idiom. Riddles and verbal memory games are popular with children.'

H. Kuper, *The Swazi* (London: International African Institute, 1952) pp. 14–15

### Revising Bantu prehistory

'We have come a long way from believing that the Bantu-speaking people crossed the Limpopo river around 1652. This is one of the great myths of twentieth-century South African historiography. Archaeological research has shown that Bantu-speaking people herding livestock and practising cultivation were already established in the Transvaal and Natal before AD 300,' writes R. Bailey in *Language and social history: studies in South African sociolinguistics* ed. Rajend Mesthrie (Cape Town: David Philip, 1995) p. 39.

The revised view may be correct (see also map 1 at BANTU LANGUAGES) but Bailey's brief statement obscures the discontinuity between two types of evidence. Archaeological research cannot really show what language these people spoke. It may be significant that *inkomo*, the word for 'cattle' in Nguni languages, seems to be a KHOISAN loanword.

# SWEDISH

**9,000,000 SPEAKERS**

*Sweden, Finland*

**S**wedish is one of the descendants of OLD NORSE. Specifically it derives from the East Norse dialect, like Danish. It is the official language of the Kingdom of Sweden, whose origins go back to the 7th century.

The Germanic tribe *Swiones* is mentioned by the Roman historian Tacitus in the 1st century AD. *Sverige*, the modern native name of Sweden, comes directly from the medieval *Sveariki*, the 'kingdom of Svea', whose centre was the province still called *Svea*.

The range of modern standard Swedish is very different from its historical extent. This history is still partially reflected in surviving dialect boundaries.

To the west, much of the region that we think of as southern Sweden was part of Denmark until 1658, when the provinces of Skåne, Blekinge and Halland were conquered. Denmark also ruled the Baltic islands of Bornholm (which has remained Danish) and Gotland, acquired by Sweden in 1645.

To the east, Finland was Swedish territory from the 12th century until 1809, and Swedish was for much of this time the only language of government and education. In the 12th, 13th and 14th centuries Swedes had settled in large numbers in the Åland Islands, which lie to the east in the Gulf of Bothnia, and also on the western and southern coasts of Finland and in coastal parts of Estonia.

Two of these areas of Swedish speech still survive, both of them now within the borders of Finland, where there are nowadays 300,000 speakers of Swedish. The Åland islands are almost completely autonomous: Swedish is their

only language, and Finnish is not even taught in most schools. Mainland Finland is, by its constitution, bilingual, with equal rights for Swedish and Finnish. In addition there is now a large number of Swedish-speaking immigrants in the United States.

Modern Swedish literature may be said to begin with 'Gustav Vasa's Bible', translated and published under King Gustav Vasa's patronage in 1540–1.

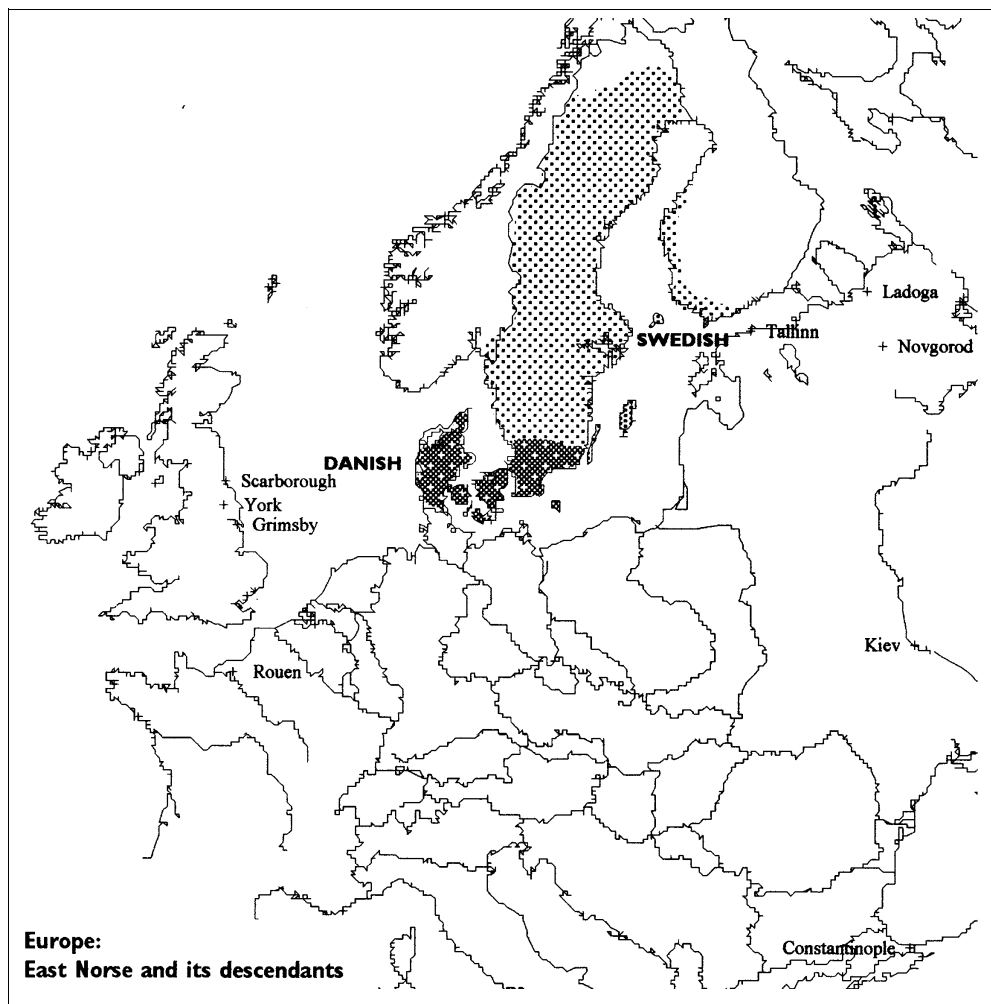
Swedish vocabulary includes Greek and Latin loanwords, particularly religious terms (*biskop* 'bishop'; *brev* 'letter'); Low German, introduced at the period when the Hansa towns controlled Baltic trade (*frukost* 'breakfast', *bädd* 'bed'). More recently, High German influence appears in the way that technical terms are formed.

## Descendants of East Norse

The modern descendants of East Norse are DANISH and SWEDISH. There are no firm frontiers between the two: they form a single dialect continuum.

The three principal dialects of Danish are Jutish (of the Jutland peninsula), Island Danish and the *Sydsvenska mål*, 'South Swedish' or Dano-Swedish dialect group, which includes the island of Bornholm. Some linguists count the latter group as Swedish.

Asterisks on the map mark former centres of East Norse speech where the language is now extinct. They include Scarborough (*Skarðaborg*), Grimsby (*Grimsbær*), York, Rouen, Ladoga (*Aldeigjaborg*), Novgorod (*Holmgårdr*) and Kiev (*Kænnugarðr*). Norse and Anglo-Saxon warriors formed the Varangian Guard at Constantinople (*Miklagårdr*).



### Numerals in Scandinavian languages

	Swedish	Danish	Norwegian	Faroese	Icelandic
1	en, ett	en, et	en, ei, et	ein	einn, ein, eitt
2	två	to	to	tveir	tveir, tvær, tvö
3	tre	tre	tre	tríggir	þrír, þrjár, þrjú
4	fyra	fire	fire	fýra	fjórir, fjórar, fjögur
5	fem	fem	fem	fimm	fimm
6	sex	seks	seks	seks	sex
7	sju	syv	syv, sju	sjey	sjö
8	åtta	otte	otte	átta	átta
9	nio	ni	ni	níggju	níu
10	tio	ti	ti	tíggju	tíu

# TAGALOG

10,500,000 SPEAKERS

*Philippines*

**A**mong the AUSTRONESIAN LANGUAGES of the Philippines. Tagalog happened to be the one that was spoken in the Manila region of southern Luzon. It was also the language of the early manifestos of the resistance against Spain and the United States. As early as 1897 came the proposal that Tagalog should be the national language of the independent Philippines. When independence was in sight, Tagalog was indeed declared the basis of the new 'national language', and it has been taught as such in schools from 1940 onwards, though in some provinces opposition to it remains. By now nearly three-quarters of the population can use Tagalog – a remarkable advance for what was, a century ago, a regional language with no official recognition at all.

*Tagalog* is said to mean 'river people', from *taga-* 'place of origin' and *ilog* 'river'. For the national language the name Tagalog has been officially replaced by *Pilipino*. This, incidentally, is a local form of the name of the islands, a name commemorating Philip II, who was King of Spain when the islands came under Spanish rule in 1565. In theory, *Pilipino* is to be replaced in due course by *Filipino* (see box).

The Philippine archipelago was ruled by Spain until 1898. Spanish is not now a major language of the Philippines, but the inheritance of Spanish culture, and of Catholic Christianity spread by Spanish missionaries, is to be traced everywhere. The typical *barrio fiesta* (Spanish loanwords: 'village festival') of the Philippines is celebrated annually in Holland Park, London, by Filipinos in Britain. It has been argued that the complex system of polite address forms in

Tagalog is borrowed from or modelled on that of Spanish.

After an abortive declaration of independence American rule followed in 1898–1946 (interrupted by Japanese occupation), and American military bases remain. English is thus a major language of the Philippines still. Education was officially bilingual from 1957, combining Tagalog with a local language. Since 1974 trilingual education is the general practice, in a local language for the first years, in Tagalog and English later. This has influenced Tagalog pervasively. English loanwords occur frequently in everyday speech and in journalism. 'Taglish' is the name given to a mixed jargon fashionable among young people.

Tagalog was already a language of written culture at the Spanish invasion. A 1593 bilingual publication in Spanish and Tagalog, *Doctrina christiana*, prints the Tagalog in an Indic-type script, not unlike that still used for MAKASAR. However, no historical records survive from that period or before. It is only clear that Tagalog, and the other Philippine languages to which it is related, had gradually spread over the islands in the preceding three or four thousand years, having originally been introduced in a migration from Taiwan, the presumed origin of all the Austronesian languages.

The script soon fell out of use. In the Spanish period, publishing in Tagalog and the other regional languages was largely religious in inspiration – yet it can be seen, from the beginning, to incorporate Tagalog poetic forms. Spanish in origin, the *corrido*, metrical romance, *pasyon*, Christian passion play, and *komedya*, theatrical 'comedy' of Christian–Muslim warfare, all three became lively and fully naturalised forms



of Tagalog literature. Manila has now a flourishing press both in Tagalog and in English.

In recent decades Tagalog has spread from its original location both as a second language and as the language of migrants, who have settled widely in Luzon and in coastal parts of Mindoro and Palawan. It is much heard in cities all over the Philippines and on the proliferating broadcast media. Large Tagalog-speaking communities exist in the United States, and migrant workers are to be found in many countries of the world.

'National language' and 'regional language of southern Luzon' are simply two ways of looking at the same language, Tagalog. There has been a movement to replace Spanish and English loanwords with native coinages: *salumpuwig* for the Spanish loanword *siya* 'seat'; *banyuhay* for the English loanword *metamorphosis*. But it is clear that Tagalog, as a lingua franca, will naturally draw on the vocabulary of Spanish, English and the other regional languages of the Philippines. Some sounds (such as the *f* of *Filipino*) occur only in loanwords.

The linguist Leonard Bloomfield produced a grammar of Tagalog as one of his earliest pieces of research, in 1917. He broke away from the Latin and Spanish tradition of Tagalog grammar, analysing the language afresh and bringing out its contrasts with European languages and their grammar.

For a table of numerals see BIKOL.

Examples from R. David Zorc, 'Tagalog' in *Comparative Austronesian dictionary: an introduction to Austronesian studies*, ed. Darrell T. Tryon (Berlin: Mouton De Gruyter, 1995–) pt 1 pp. 335–41

### Filipino: language of the future

In its 1973 constitution the Philippine government looked forward to a new language, *Filipino*, declaring: 'The National Assembly shall take steps towards the development and formal adoption of a common national language to be known as Filipino . . . Until otherwise provided by law, English and Pilipino shall be the official languages' (1973 Constitution of the Philippines, article xv, section 3, paras 2–3).

To forestall regional opposition to Tagalog-Pilipino, the plan was to invent a truly inclusive national language, with a grammar and vocabulary based somehow on all the regional languages. Some work has been done since 1973, but language planning of this kind usually fails. Redesigned Filipino is not likely to progress far: instead, 'Filipino' may gradually replace 'Pilipino' as a name for Tagalog viewed as the national language.

### Tagalog and its regional rivals

The eight 'major languages' of the Philippines are Tagalog, Ilocano, Pangasinan, Kapampangan, Bikol, and the three Bisayan languages, Cebuano, Hiligaynon and Waray-Waray: for the last three see map at CEBUANO. Together these eight are the mother tongues of nearly 90 per cent of the population of the Philippines.

Tagalog dominates in Greater Manila, Bataan, Batangas, Bulacan, Cavite, Laguna, Marikina, Nueva Ecija, Mindoro Occidental, Mindoro Oriental, Quezon and Rizal. It is becoming the lingua franca of many cities and towns throughout the Philippines. Standard Filipino is based on the Tagalog dialect of Manila. Tagalog and Filipino are now used, as lingua franca, throughout the Philippines, even where they are no one's mother tongue.

BIKOL is the major language of southern Luzon: Albay, Camarines Sur, Catanduanes, Sorsogon, and parts of Camarines Norte and Masbate.

ILOCANO, the language of north-western

#### The Philippines: Tagalog and its rivals



Luzon, is dominant in the provinces of La Union, Ilocos Sur, Ilocos Norte and Abra. It is also widely spoken further to the south-west in Tarlac, Pangasinan and Zambales, and in parts of Mindoro island. It is also spoken in far-off Cotabato on the island of Mindanao.

PANGASINAN is the language of the central part

of Pangasinan province in north central Luzon.

KAPAMPANGAN is spoken in Pampanga province, north-west of Manila, as well as in four cities of Tarlac (Bamban, Concepcion, Tarlac itself and Capas) and two cities of Bataan (Dinalupihan and Hermosa) near the western borders of Pampanga.

# TAHITIAN

100,000 SPEAKERS

*French Polynesia*

**T**ahitian, one of the Oceanic branch of AUSTRONESIAN LANGUAGES, is spoken by about two-thirds of the population of French Polynesia. This island group of the central Pacific is known in English as the Society Islands: better known is the name of the largest island, Tahiti.

The first printed book in Tahitian was an ABC, *Te Aebi no Tahiti*, by John Davies, published in 1810. Much literature, mostly religious and educational, has appeared since then.

Since the first European contact Tahitian has been changing rather rapidly and has been considerably influenced by French. Tahitian has some official status locally, since knowledge of the language is required for Polynesians (though not for French speakers) in government service. As the language of local administration, of central health, social and educational services, and of local broadcasting, Tahitian is a lingua

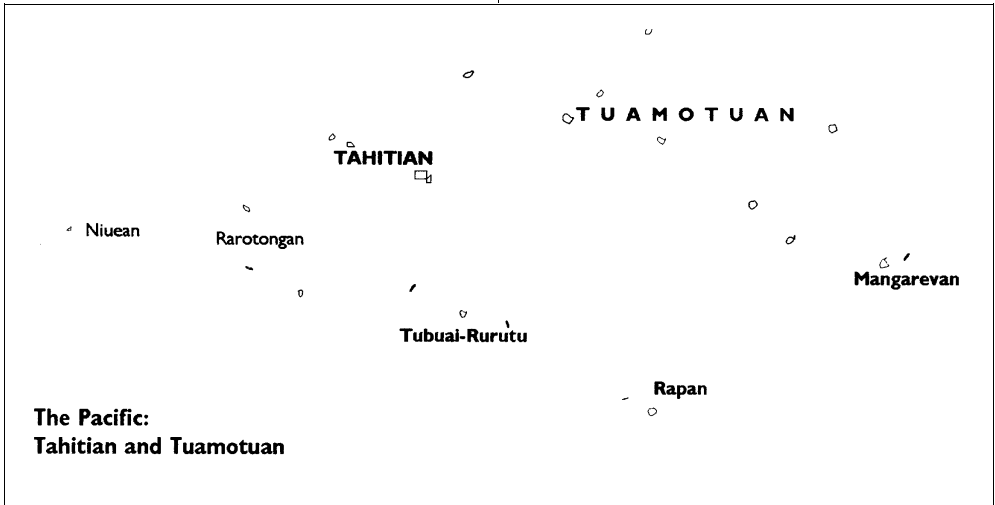
franca in the territory and the local languages of the smaller islands are giving way to it. Mangarevan, Rapan, Tuamotuan and Tubuai-Rurutu (all of them Polynesian languages) are among those that show signs of succumbing to Tahitian in its lingua franca form, *Neo-Tahitian* as it is sometimes called.

Even among Polynesian languages Tahitian is unusual in its sound pattern, having only nine consonant phonemes, *p t ' f v h m n r*. It has five vowels and, since a syllable may contain one or any two vowels, a total of twenty-five diphthongs.

The first ten numerals in Tahitian are: *tahi*, *piti*, *toru*, *maha*, *pae*, *ōno*, *hitu*, *va'u*, *iva*, *hō'e ahuru*.

## Tahitian on the map

Like the West Indies, the Society Islands are divided by mariners into a Leeward and a Wind-



ward group. There are some dialect differences between the two.

*Tuamotuan*, with 15,000 speakers, is the language of the Low or Dangerous archipelago.

Along with the small island languages *Mangarevan*, *Rapan* and *Tubuai-Rurutu*, it is now in decline: Tahitian is taking their place.

# TAI LANGUAGES

**T**he major constituent of the KADAI LANGUAGES, one of the groupings that probably belongs to the postulated AUSTRO-TAI family, is Tai – a group of closely related languages usually classified into three roughly geographical divisions.

The South-western division includes the two best known, THAI or Siamese, spoken in central Thailand, and Lao of Laos and north-eastern Thailand. A continuum of dialects in areas to the north of these begins with AHOM, an extinct language of the Indian state of Assam, and SHAN; it continues eastwards with LANNA THAI and ends with the minority languages of Laos and Vietnam usually known as Black Tai, Red Tai and White Tai (see map at LAO).

The Central group includes THO, NUNG and Southern Zhuang. Northern Zhuang and BUYI are classified as the Northern group. These are all minority languages of southern China and the northern borders of Vietnam (see map at ZHUANG).

Because of some obvious similarities with Chinese, and because, like Chinese, they are monosyllabic and tonal languages, it has often been assumed that Tai languages and SINO-TIBETAN LANGUAGES are ultimately related. Most scholars now agree that these similarities are due to

borrowing. Instead, evidence has accumulated for a link between Tai and other languages of south-east Asia in a postulated family of AUSTRO-TAI LANGUAGES.

*Tai* or *Thai*, ‘free’, is the people’s own term for themselves: linguists have chosen the variant *Tai* as the name for the language family, while *Thai* normally means the national language of Thailand.

Many foreigners use another term for Tai peoples. In Assam they were called *Asam*, locally pronounced *Ohom*; in Burmese and Khmer, they are called *Shan*, *Syam*; in Chinese they have been called *Sien*. *Assam* and *Siam*, as country names, both derive from this term.

*Dai*, in Chinese, is the name for the South-western members of the family: *Zhuang-Dai* is the Chinese term for the grouping called *Tai* by linguists writing in English.

Tai chronicles trace the history of individual Tai-speaking states back no further than the 12th century, and this really does seem to be the approximate date of the appearance of Tai ruling groups in south-east Asia. The details of Tai history before this time are not known, but

## Numerals in Tai languages

	Shan	Thai	Lao	Northern Zhuang
1	nǒng, ét	nǒng, ɛt	nǒng, ɛt	ɪt, dǝw
2	sǒng	sǒng	sǒng	sǒng, ngey
3	sám	sám	sám	sám
4	sǐ	sǐ	si	séy
5	hà’	hà’	hə	hā, ngù
6	hók	hok	hok	rèk, gok
7	cét	cet	cet	cāt
8	pēt	pēt	pēt	pét
9	kàw’	kàw’	kàw	kōw
10	síp	sip	sip	cip

The second form for ‘one’ is used in compounds, e.g. in Shan *síp ét* ‘eleven’. Forms in the three south-western Tai languages (Shan, Thai and Lao) differ only in the tones. For Thai numerals in original script see table at THAI.

the ultimate origin of the language group must be in south-eastern China, where many Tai speakers are still to be found: it is from here, several thousand years ago, that early Austronesian language speakers must also have dispersed.

Many Thai historians hold to the theory that the Yunnanese kingdom of Nanchao, once powerful and independent, destroyed by a Mon-

gol invasion in 1253, was Tai-speaking. The fall of Nanchao was thought to explain the apparent rapid spread of Tai speakers towards the south at about this period. It now seems more likely that the languages of Nanchao were BAI and YI, or something like them. It is not unlikely that some Tai states with long histories, including Hsip Hsong Pan Na (see LANNA THAI), were in early times tributary to Nanchao.

# TAMASHEQ

**PERHAPS 1,000,000 SPEAKERS**

*Niger, Mali, Burkina Faso, Algeria, Libya*

**T**amasheq, one of the BERBER LANGUAGES (see map at TAMAZIGHT), is spoken by the *Tuareg* (singular *Targi*) of the Sahara.

The name *Tamasheq* identifies it as the language of the *Amaziɣ* ‘Berbers’.

In spite of the unbroken tradition of writing in the *tifinagh* alphabet, no long texts are known from earlier periods of Tamasheq. Interest in the language stemmed from European exploration of the Sahara.

The major dialects of today centre on Timbuktu, Tahoua, Aïr (*Tayart*) and Hoggar. But, as the language of a nomadic people, Tamasheq has surprisingly small dialect differences across the vast area over which it is spoken. There are about 450,000 speakers in Niger, about 300,000 in Mali and 150,000 in Burkina Faso. In all three countries, Tamasheq is recognised as a regional official language and is used in schools and adult education. Numbers in southern Algeria and Libya are much smaller.

Tamasheq has, naturally, fewer Arabic loanwords than the Berber languages of the north, where the long term influence of Islamic culture, radiating from the cities, has been almost overwhelming. By contrast, Tamasheq draws loanwords from African languages such as Hausa.

Like the other Berber languages, Tamasheq has a gender distinction in which feminine forms are typified by a *-t* suffix, as seen in the table of numerals.

Tamasheq (Tayart) numerals		
Masculine		Feminine
īyan	1	iyāt
əsshin	2	sənatāt
kərād	3	kəradāt
əkkoz	4	əkkozāt
səmmos	5	səmmosāt
səḡdis	6	səḡdisāt
əssa	7	əssayāt
əttam	8	əttamāt
təza	9	təzayā
məraw	10	mərawāt
M. Malherbe, <i>Les langages de l'humanité</i> (Paris: Laffont, 1995) p. 1545		

## Tifinagh

The *Tifinagh* (‘Phoenician’) script still used by some Tuareg communities is descended from the alphabet used for the ‘Old Libyan’ Berber languages in pre-Roman times. It is not taught in schools: it is used for private notes, love letters, and in decoration. It is astonishing that it has survived, with uses that seem to observers so unimportant, for two thousand years – but it may be a crucial factor that women do not learn to read Arabic, the language and script used by most Tuareg for public purposes.

Attempts to use Tifinagh for other purposes, whether in Bible translations or in Berber nationalist movements, have failed.

# TAMAZIGHT

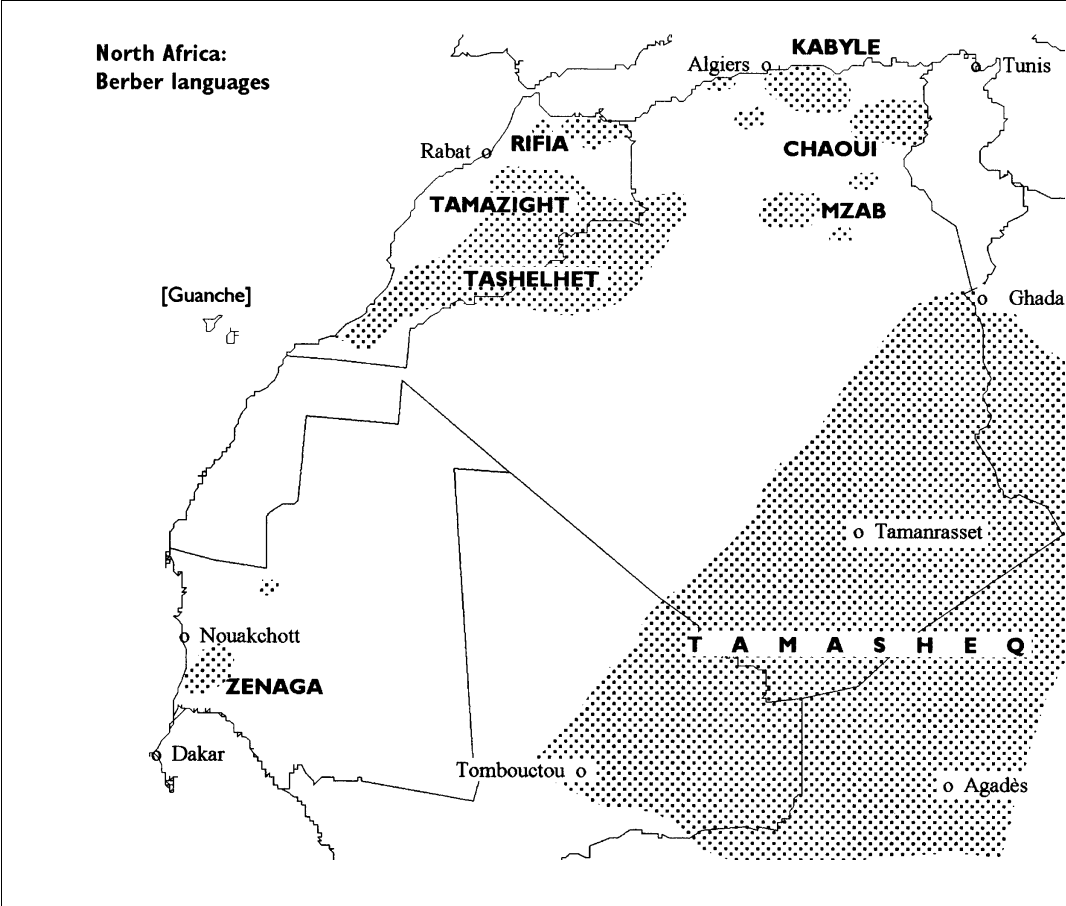
3,000,000 SPEAKERS

*Morocco, Algeria*

**T**amazight is one of the BERBER LANGUAGES. Its speakers gave their name to the whole language group, for it is spoken by the people called *Berber* and is sometimes simply referred to as ‘Berber’. But this was originally a derogatory Greek word, ‘barbarian’, and the speakers’ own name for their language, *Tamazight* ‘lan-

guage of the *Amaziy* or Berbers’, is more often used. The name occurs in Ibn Khaldūn’s 13th-century *History of the Berbers*: ‘their ancestor was *Mazigh*’.

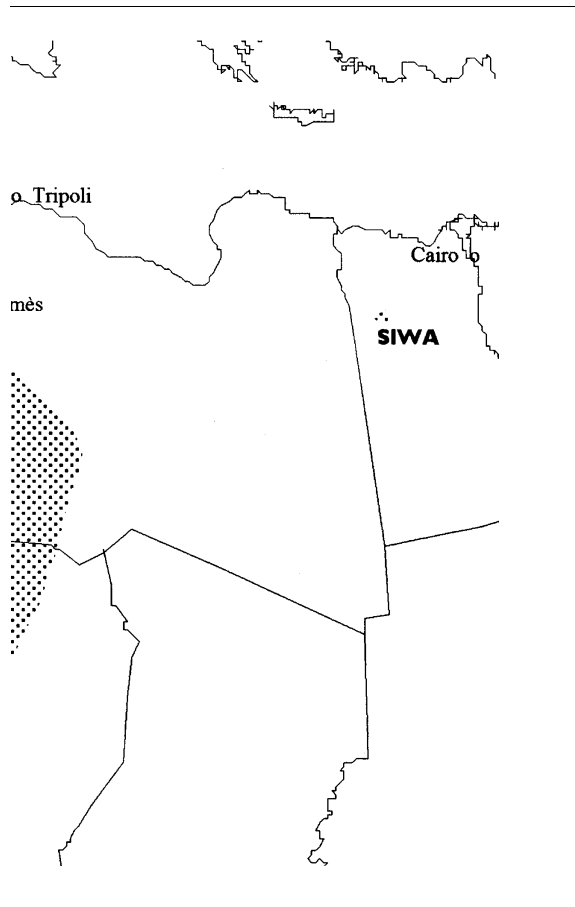
The Berber languages of Morocco are sometimes grouped under the name Shilha: in these terms Tamazight is ‘Central Shilha’. It has about





1,800,000 speakers in the Middle Atlas region of northern Morocco, and perhaps another 1,200,000 in western Algeria. It has no official status in either state: in Algeria, a constitutional amendment in 1996 declared Arabic the sole language of business, education and administration. Men of the Tamazight country are usually bilingual in Arabic, but Tamazight remains the language of the household.

Tamazight literature is largely oral. Some manuscript collections of poetry in Arabic script have been collected and published. *The forgotten hill*, the first feature film in Tamazight, shot in Algeria, appeared in 1997.



# TAMIL

**52,000,000 SPEAKERS**

*India, Sri Lanka, Malaysia, Vietnam, Singapore*

**B**elonging to the family of DRAVIDIAN LANGUAGES, Tamil is the first language of the Indian state of Tamilnadu and is spoken by a minority of over 2,000,000 speakers in north-eastern Sri Lanka.

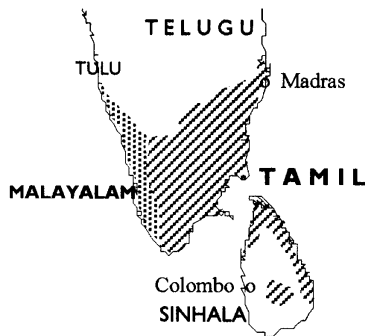
*Tamiḻ* or *Tamiḻ* is a Tamil form of the old Sanskrit name for the southern Indian kingdom, *Drāviḍa*, and is thus the term that foreigners have applied to the region and the language for nearly two thousand years.

For all this time, and even for longer, Tamil has been spoken in the area of modern Tamilnadu. Cave inscriptions, in a mixture of early Tamil and Prakrit, date back as far as the 3rd century BC. The earliest Tamil literature goes back perhaps to the 1st century AD: partly Buddhist-inspired, it consists of lyric poems, epics – and a grammar, *Tolkappiyam*.

Medieval Tamil is dated from AD 700 to 1500: Hindu religious influence now supplanted Buddhism, but Tamil, unlike its twin language MALAYALAM, remained not especially receptive to Sanskrit loanwords. It is from this period that the written and spoken forms of the language have grown apart, for literary Tamil remains close to the prescriptions of the grammarian PavaNanti-munivar's *Nannul*, compiled in the 13th century and itself drawing on the more ancient texts.

The 19th century saw attempts to bring modern spoken Tamil into written form, partly in the context of Christian missionary activity. The style of, for instance, school readers and of the dialogue passages in fiction still differs very strongly from that of formal prose.

Tamil has a much smaller range of consonant phonemes than other Indian languages (see box). Verb forms include suffixes for person,



**Southern Asia: Tamil and Malayalam**

number, and in the case of the third person also for one of three genders (male, female, non-human). Verbs also have a negative conjugation. There are many differences in noun and verb forms between literary and spoken Tamil.

### Beyond India

Tamil speakers have been among the readiest of the inhabitants of India to migrate for work and to settle abroad. The first records of Tamil trading colonies overseas are in Tamil inscriptions from southern Thailand, in the 9th century, and from northern Sumatra, dating from 1088 onwards. Tamil is now spoken in Malaysia, in Vietnam, in Singapore (where it is one of the four national languages), on the island of Zanzibar, and in many other countries.

In nearby Sri Lanka there are two layers of Tamil speech. The 'indigenous' dialects, long established in the country, have over 2,000,000 speakers in the north of the island. The 900,000 speakers of 'Indian Tamil' are the descendants of tea plantation workers brought to the central hills in the mid 19th century. The dialects may begin to converge as speakers of indigenous Tamil become familiar with Indian Tamil through imported films, radio and television.

Sri Lanka has suffered serious linguistic unrest. The Sinhala-Only Bill of 1956, requiring the

use of Sinhala in official contexts, eroded the position not only of English (its stated intention) but also of Tamil. Rebellion flared. Over 200,000 Tamil speakers fled abroad in the course of the 1980s. In 1987 Tamil and English nominally regained their old position as official languages, but the dispute, once aroused, has not been easy to calm.

### Tamil and Malayalam

Tamil is spoken in the Indian state of Tamilnadu, in Sri Lanka (3,100,000), and by large communities in Malaysia (1,750,000), Vietnam (perhaps 1,000,000), Singapore (200,000) and other countries across the world.

Malayalam, historically an offshoot of Tamil, is the language of the Indian state of Kerala. Malayalam communities abroad tend to be counted with Tamil communities and to assimilate to Tamil as their home language.

#### The twelve initial vowels of Tamil

அ ஆ இ ஈ உ ஊ எ ஏ ஐ ஒ ஓ  
a ā i ū ē ē ai o ō au

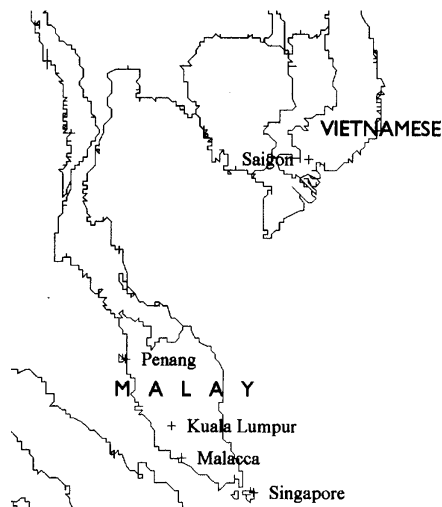
#### The eighteen consonants of Tamil

கஙசஞடணதநப  
மயரலவழளறன  
k ṅ c ṇ ṭ ṇ t n p m y r l v z ! ṛ ṇ

### The Tamil script

As with other scripts descended from Brahmi, consonants are combined with following vowels to make a single character group. Unlike others, Tamil has no compound consonant symbols.

The Tamil script, admirably efficient in plan, is strictly limited to what is necessary to write Tamil phonemically. It works well in writing the literary language and for traditional vocabulary. Speakers from written texts, actors for example,



learn to make a fairly complex set of conversions, subconsciously, as they speak.

However, Tamil script does not go well with the way that Tamil speakers now use language. Sanskrit words can be inserted in Tamil texts with the help of extra ‘grantha’ characters; but

English and other loanwords have no such traditional help and are difficult to write. So the modern spoken language, now with many international loanwords and an extended range of sounds to accommodate them, cannot be fully written in Tamil script.

# TANGUT

## Extinct language of western China

**T**angut, known in Chinese as *Xixia* (older transliteration *Hsihsia*), was one of the SINO-TIBETAN LANGUAGES and apparently a member of the Burmese-Lolo group.

The kingdom of the Tangut was a major power in north-western China in early medieval times. In AD 1037 a script of about 6,600 characters was devised for their national language, credited to 'the Teacher Iri'. It flourished for only two centuries. Genghis Khan was killed in 1226 at the siege of the Tangut capital. His son Kublai wreaked fierce revenge on the Tangut. Those who remained continued to use their

Buddhist texts until about the 16th century, after which time the language clearly ceased to be known.

The Tangut script has been deciphered with the help of a Chinese–Tangut dictionary, the *Sea of characters*, compiled by the Chinese diplomatic service in the 12th century. Like Chinese, it was a pictographic script extended by the phonetic use of rhymes. Character combinations were formed by borrowing the consonant value from one character and the vowel from another. The sounds of Tangut have been only partly recovered. Like Chinese, it was clearly a tonal language.

# TASHELHET

PERHAPS 3,000,000 SPEAKERS

*Morocco, Algeria, Mauritania*

**T**ashelhet or Tachelhit is one of the BERBER LANGUAGES (see map at TAMAZIGHT).

The Berber languages of Morocco are sometimes grouped under the name Shilha. Tashelhet is the one from which this name derives, for the two words are identical in origin: speakers call themselves *Shlhi*, sometimes rendered *Chleuh* in French or *Shluh* in English. They call their language *T-ashllhi-t*, with the *t-* prefix and *-t* suffix typical of a feminine noun.

Tashelhet has no official status in Morocco or Algeria. Male speakers are usually bilingual in Arabic, but women are not.

The first ten numerals in Tashelhet, in their masculine forms, are: *yan, sin, kraḍ, kkuḥ, smmus, sḍis, sa, tam, ttḥa, mraw*. Feminine numerals are formed with a final *-t*: *yat, snat, kraṭt, kkuḥt* and so on up. Arabic tends to be used for counting above ‘10’.

### The twelve months in Tashelhet

The influence of imperial Latin, at the south-western extremity of the Roman Empire, seems to be indicated by the month names in Tashelhet. It has been argued that they are borrowed, not directly from Latin, but, much later, from Spanish, which had inherited these names from Latin.

Latin	Tashelhet	Spanish
januarius	innayr	enero
februarius	xubrayr	febrero
martius	maṛṣ	marzo
aprilis	ibrir	abril
maius	mayyuh	mayo
junius	yunyu	junio
julius	yulyuz	julio
augustus	ʁušt	agosto
september	šūṭambir	setiembre
october	kṭubṛ	octubre
november	nuwambir	noviembre
december	dujanbir	diciembre

After J. Bynon, ‘Linguistics and Berber history’ in *Language and history in Africa* ed. D. Dalby (London: Cass, 1970) pp. 64–77

# TATAR

**6,000,000 SPEAKERS**

## *Russia*

**T**atar is one of the **TURKIC LANGUAGES**, spoken in European Russia to the west of the Urals. For a table of numerals see **BASHKIR**.

Ruled and led by Mongolians, the warriors subject to the Golden Horde were largely Tatar-speaking and Muslim. They overran Rus' in the 13th century, destroying the old capital, Kiev, in 1242. Kiev and Ukraine were soon lost to them, but they dominated Moscow for two hundred and fifty years – and ever after, Muslim peoples encountered by the Russians were usually called by them 'Tatar'. At this period there was inter-marriage and intermixture between Russian and Tatar nobility. The Russian language still shows the influence of Tatar, and such figures as Boris Godunov and the novelist Fyodor Dostoyevskii could claim Tatar descent.

In the 15th century Moscow threw off the Tatar overlordship, and soon afterwards, as its power grew, conquered the Golden Horde's khanates of Kazan and Astrakhan. The khanate of Crimea was annexed by Catherine the Great in 1783.

Turkic and Mongolian speakers had settled in these three neighbourhoods. For the Kalmyks of Astrakhan see **MONGOLIAN LANGUAGES**; for the 'Tatars' of the Crimea see box. The true Tatars were those of Kazan, and it is south-east of Kazan that most Tatar speakers are to be found today. After submitting to Russia themselves they were active once more in the subsequent spread of Russian power to the Caucasus, Azerbaijan and Central Asia. In all these regions Tatars served as interpreters, administrators and – most significantly – teachers and missionaries in a 19th-century Muslim renewal.

The Tatar diaspora is numerous, both inside

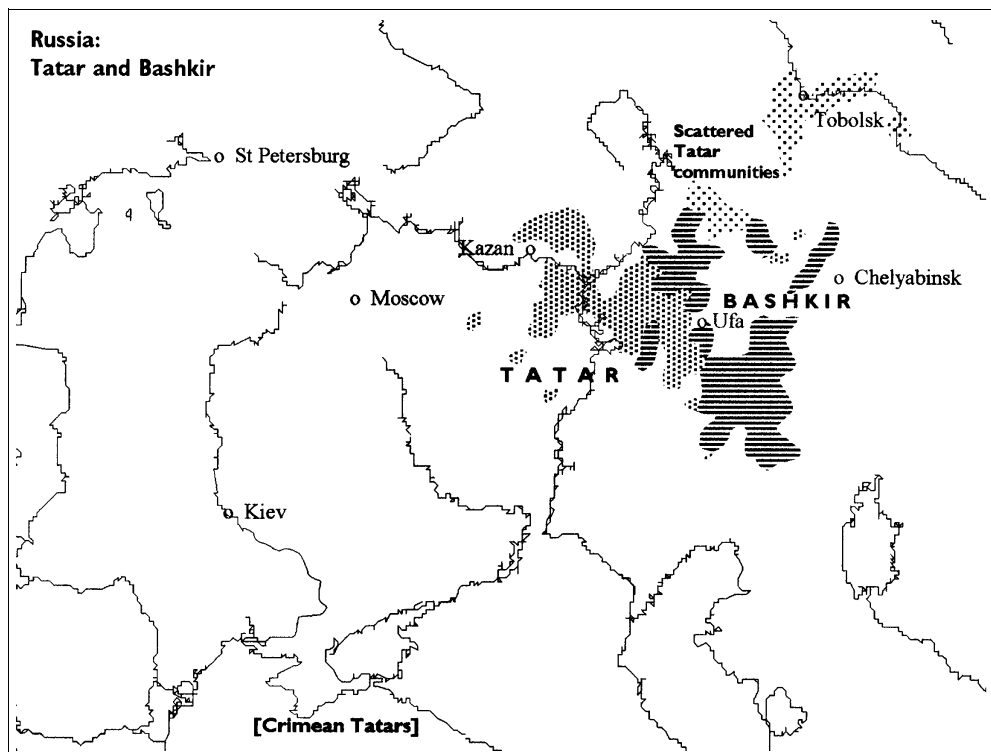
Russia (see map) and beyond its borders. There are Tatar communities in the Bulgarian Dobruja, for example. A World Congress of Tatars assembled in Kazan in 1992.

Tatarstan, now an autonomous republic within Russia, has an independent-minded government and a powerful Muslim party. One politician, Fauzia Bairamova, said in 1992: 'I would like to be able to say that Russia has contributed more to Tatar civilisation than prostitution and alcoholism. Sadly, I cannot.'

## **The dispersal of the Crimean Tatars**

That ill-fated group of peoples, the Crimean Muslims and Jews, once had six distinct languages. The Muslims of the south coast of the Crimea spoke Crimean Ottoman, a branch of the south-western group of Turkic languages. Those in the centre around Bahçesaray spoke Central Crimean Turkic, and those from the steppe spoke Crimean Tatar, a Kypchak language: both of these were western Turkic. The nomadic Nogai, of the northern Crimea and beyond, spoke a language of their own (see **KAZAKH**); their history is traced to Amir Noghay in the 13th century.

One Jewish people of the Crimea, the Krymchaks, had their own language, close to Uzbek, but gradually assimilating to Crimean Tatar. These were orthodox, Talmudic Jews. Another, the Karaim, were 'heretical', owing their religious beliefs to a sect of Judaism that split off in Basra in the 8th century. Imperial Russia recognised the Karaites as a separate religion, centred at Evpatoriya in the Crimea, led by a *hakam* (Hebrew: 'sage'). The language of the



Karaim is a quite distinct offshoot of Turkic, and the Karaim traced their historical origin to the mysterious Khazar Empire, which had indeed adopted Judaism as its state religion and ruled the south Russian steppe in the 7th to 10th centuries. Both of these Jewish peoples had scriptures in Hebrew, but they used their own Turkic languages – written in Hebrew characters – in everyday life.

There were also Gothic, Greek, Slavonic, Arab and Iranian communities along the Crimean coast. This complex, multilingual society was doomed when Ottoman protection for the once-autonomous Khanate of the Crimea gave way to Russian rule around 1800. Population movements then were intensified by the Crimean War in 1853–6, the German occupation during the Second World War, and the Soviet deportations that followed.

Already in the 1780s many Crimean Tatars and Jews had migrated to the European provinces of the Ottoman Empire – and, as the empire continued to shrink, moved on again

to Turkey. A majority of the remaining population fled, following these same routes, in the disastrous 1850s. Crimean Ottoman is so close to Turkish that its speakers soon gave up their language. Karaim had already spread to Lithuania and Poland.

Of the minority that remained in the Crimea, the Krymchaks, as Jews, were killed by the Germans in 1944. The Karaim, accepted by them as Turks whose religion alone was Jewish, were not killed. There are perhaps 20,000 Karaim today, most of them now in Israel – but few still speak their ancestral language. The Tatars, numbering about 190,000, were deported en masse to Central Asia in 1944; well over a quarter of these died in the first five years of exile. They were not officially allowed to return until 1990. Since then as many as 50,000 have found their way back to the eastern Crimea, but over 200,000 remain in Central Asia.

Based on Peter Alford Andrews, *Ethnic groups in the Republic of Turkey* (Wiesbaden: Steiner, 1989) and other sources



# TELUGU

45,000,000 SPEAKERS

*India*

**O**ne of the DRAVIDIAN LANGUAGES, Telugu is the language of the Indian state of Andhra Pradesh (for map see KANNADA).

*Telugu* is a later, local form of the name *Telinga* once given to this region and its people in Sanskrit texts. *Andhra* is an alternative, used officially in the state name (*pradeśa*, 'province') and locally for the people and their language. In some older English writings it is called *Gentoo*, from Portuguese *Gentio*: from the point of view of early European traders, speakers of this language were the 'gentiles' of the Empire of Vijayanagar, in contrast with the *Moros*, the 'Moors' or Muslims, who were more in evidence on the western coast of India.

Within India Telugu perhaps has a larger number of speakers than any other Dravidian language. It is known from inscriptions beginning in the 6th century AD and from literature beginning around 1100. The Middle Telugu period is reckoned from the 11th to 15th centuries. It was in this period that the literary language became fixed in form. Among the early classics are versions of the Sanskrit epics and *purāṇas*, notably the 11th-century *Andhra Mahābhārata* of Nannaya. Early modern Telugu court poetry was interwoven with music and dance; the mixed verse and prose form of the *prabandha* is a special feature of this literature.

By the beginning of the 20th century literary and spoken Telugu were almost two different languages. After long dispute between classicists and modernists, a modernised literary language, much closer to everyday speech, is now almost universal in current writing and in the media.

Telugu has three genders – but only two distinct gender forms. In the singular of nouns, feminine and non-human genders are combined, so that the distinction is masculine/non-masculine; in the plural, masculine and feminine are combined, and the distinction is human/non-human. The positive/negative opposition is incorporated in some verb forms: *ammutāḍu* 'he sells it', *ammaḍu* 'he does not sell it'.

Telugu has undergone lengthy, pervasive influence from Sanskrit, the traditional learned language of the subcontinent, and also from the Prakrits, the spoken Indo-Aryan languages of medieval northern India. As a result, the vocabulary of traditional literary Telugu is heavily Indo-Aryan, and so is that of the modern spoken language. At least one Telugu scholar argues that Sanskrit, not proto-Dravidian, is the true parent of Telugu.

## Polite address in Telugu

Politeness is expressed by the choice of pronouns: *nunuvu* 'you (singular)' is informal, while *mīru* 'you (plural)' also functions as honorific singular. The third person pronouns have a four-way distinction for males and a three-way distinction for females: listed from low to high these are *wāḍu*, *atanu*, *āyana*, *wāru* 'he', and *adi*, *āme*, *wāru* 'she'. The choice depends on the age and education of speaker and hearer, the social context and the speaker's intentions.

After Bh. Krishnamurti, 'Telugu' in *International encyclopedia of linguistics* (New York: Oxford University Press, 1992) vol. 4 pp. 137–41

Numerals in Telugu and Kannada		
Telugu		Kannada
oṇḍu, oka	1	oṇḍu
reṇḍu	2	eradu
mūḍu	3	mūru
nālugu	4	nālku
eidu	5	aidu
āru	6	āru
ēḍu	7	ēḷu
enimidi	8	eṇṭu
tommidi	9	oṇbhattu
padi	10	hattu

Telugu script

Telugu script is very similar to that of Kannada. As with other descendants of the Brahmi script, a consonant is combined with a following vowel in each symbol. Compound consonants are usually easy to pick out as they generally involve subsidiary symbols written below the line.

The Telugu consonants	
క ఖ గ ఘ చ ఛ జ ఝ ట ఠ డ ఢ త థ ద ధ న ప ఫ బ భ మ య ర ల వ శ ష స హ	
k kh g gh c ch j jh ṭ ṭh ḍ ḍh ṇ t th d dh n p ph b bh m y r l v ś ṣ h ḷ	

# TEMNE

1,000,000 SPEAKERS

*Sierra Leone, Guinea*

**T**emne or Themne or Timne, a member of the Atlantic group of NIGER-CONGO LANGUAGES, is the predominant language of north-western Sierra Leone.

The first European notice of Temne came in a wordlist made by an English mariner in 1582; the language was also seen as important by the Jesuit mission to Sierra Leone in 1610. Temne speakers (*aTemne*, as they call themselves) remained significant on a regional scale in the 17th century, when they clearly ruled the coastal hinterland of southern Guinea and northern Sierra Leone, enriched by trading contacts with Europeans. At this period the related Bullom or Sherbro was the dominant language on the coast itself, but it soon gave way to Temne in the north and later to Mende in the south. It was the Temne who ruled at the site of Freetown, which they ceded to the British in 1788. This marks the establishment of Sierra Leone Colony, under which some Temne lived: most remained independent, though at times under Mende domination, until a British Protectorate was proclaimed in the hinterland in 1896.

Christian missionaries began to work among Temne speakers in the 1790s. Many Temne are Muslim, and the language shows Arabic influence. Temne serves as a second language to as many as a quarter of a million speakers of local languages including Bullom, Susu, Lokko and Limba.

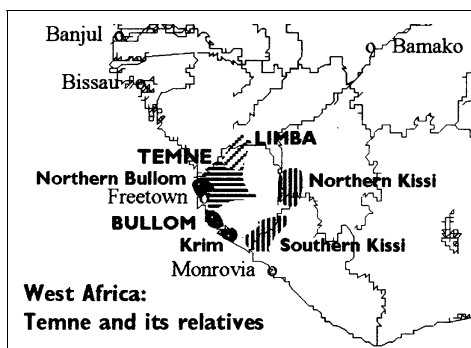
## Temne and its relatives

*Temne* has a northern (*Sanda*) and a southern (*Yonni*) dialect. *Baga* and *Landoma* are a series of dialects of southern Guinea that are best counted as belonging to Temne.

*Limba* or *Yimbe* has about 250,000 speakers in inland northern Sierra Leone.

*Bullom* (as the speakers call themselves) or *Sherbro*, now a relatively minor language with about 175,000 speakers, was once the lingua franca of the coast. There are several early wordlists in Bullom compiled by European mariners and traders. *Northern Bullom* has a small number of remaining speakers in coastal parts of Temne country. *Krim* or *Kim* is a southern dialect of Bullom.

*Northern Kissi* or *Gizi* has nearly 300,000 speakers in Guinea. The *Southern* dialect is spoken by about 150,000 people in north-western Liberia and Sierra Leone.



# TERNATE

**70,000 SPEAKERS**

*Indonesia*

**T**ernate belongs to the North Halmaheran group of West PAPUAN LANGUAGES. The other groups of this family are spoken on the Bird's Head peninsula of West Irian.

Ternate and its relatives (see map) are the only languages of any of the Papuan families that are spoken outside the immediate vicinity of New Guinea. It is not known whether they came to Halmahera from New Guinea, or whether early speakers of the Bird's Head languages migrated there from Halmahera. At any rate, there are still close contacts between the two regions.

The importance of Ternate, and its twin language Tidore, is that they were spoken on the two small islands off western Halmahera which are known to history and legend as the 'Spice Islands'. Until the 16th century, these were the only places in the world where cloves grew. The prized spice was known to India, China and the Roman Empire by the last centuries BC – though its buyers long believed that it came from Java. Clearly the first stage of its long journey was handled by local traders.

The rulers of Ternate and Tidore grew rich and powerful on the profits from their luxury export. They competed for the overlordship of all of Halmahera and the central Moluccas. Both island states became Islamic sultanates in the 15th century. Ternate and Tidore are recorded in writing – in Arabic script – from this period onwards.

A layer of Ternate loanwords is naturally found in Tobelo and Galela, the related languages of the coastal peoples of north-east Halmahera and Morotai. Tobelo and Galela seamen manned the war canoes with which Ternate

raided the north coast of New Guinea. Ternate had an even stronger influence on neighbouring Sahu: Sahu-speaking communities supplied rice and domestic servants to the nobility of Ternate. Ternate loanwords are found in many of the Austronesian languages of the Moluccas.

From its earliest recorded history Ternate was rich in Malay loanwords, an indication of already longstanding trade contacts. Portuguese and Dutch, languages of successive European intruders, have also influenced Ternate.

Besides its use as a mother tongue, Ternate is still spoken today by many of the inhabitants of northern Halmahera. Their own languages (see map) were not written down until the Christian missionaries of the Utrechtse Zendingsvereeniging began studying them, and making Biblical translations, in 1866.

## **Ternate as language of rhetoric and ritual**

Ternate is no longer a major language of trade: this role is taken by the local variety of MALAY. But its prestigious history as the language of the Sultans of Ternate has ensured its place in ceremonial.

'Many Tobelo are anxious to learn words or phrases of Ternatese. Some of this may be due to whatever oratorical tradition makes some Tobelo want to incorporate a few basic words of several languages into their conversation . . . Ternatese predominates, specifically, in the language of traditional chants, of magical formulae, and of the give-and-take of marriage consultations between families of the bride and groom.

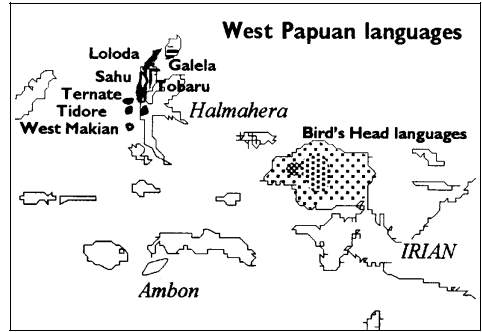
'One proper setting for the traditional chants

is the proud festivity of the last step of the marriage process, when the new daughter-in-law arrives at the house of her groom's parents and (traditionally, for three separate nights) she is 'displayed' perfectly motionless in her finest regalia of heirlooms (borrowed for the occasion from as many relatives as possible), constantly waited upon by her new sisters-in-law, while the older generation among her new male in-laws, chewing the slightly narcotic betel nut and drinking the palm-wine that ideally should flow freely on such occasions, sing their chants in Ternatese to the hearty beating of gongs and deer-skin drums.'

Paul Michael Taylor, *The folk biology of the Tobelo people* (Washington: Smithsonian Institution Press, 1990) p. 14

### The West Papuan languages

*Ternate* is spoken on Ternate itself and in several towns on the facing coast of Halmahera. The *Tidore* dialect, likewise, is spoken on Tidore and in three coastal enclaves. Ternate is known throughout northern Halmahera.



*Sahu* (10,000 speakers), *Loloda* or *Loda* (12,000), *Tobaru* or *Tabaru* (15,000), *Galela* (25,000, also known more widely as a language of love poetry) and *West Makian* (12,000) are all members of the North Halmaheran group. So is *Tobelo* or *Tugutil* (at least 25,000 speakers), the language of the north-east Halmaheran coast and of the eastern highlands, which is itself known in many parts of the island. It is spoken as far off as Ambon to the west and Raja Ampat to the east.

The Bird's Head group of the West Papuan family includes *Moraid* or *Hattam* (12,000 speakers), *Mai Brat* (20,000 speakers) and some smaller languages.

# TESO

1,300,000 SPEAKERS

Uganda, Kenya

One of the Eastern Nilotic group of NILO-SAHARAN LANGUAGES (see map at TURKANA), Teso is traditionally said to have been brought to eastern Uganda in a migration from further east, led by the culture hero Teso.

The language was once called *Dum* by outsiders. Speakers call themselves *Iteso* (singular *Etesot*). The form *Ateso* is sometimes found. In Kenya, where a distinct dialect is spoken (see map at TURKANA), the official term is *Itesyo*. Among Ganda speakers, the Teso to their east are known as *bakide*, a word said to mean ‘naked people’ – but it may derive from Teso *kide* ‘east’.

At the end of the 19th century Ganda-speaking agents brought most of Teso territory under the British protectorate of Uganda, and reorganised its political system. From then onwards

Teso has adopted an increasing number of Ganda loanwords. Other Bantu languages, such as Luyia, have also influenced Teso.

The first ten numerals in Teso, in their feminine forms, are: *adiope*, *aarei*, *auni*, *aojom*, *akany*, *akany kape*, *akany kaarei*, *akany kauni*, *akany kaajom*, *atomon*.

**The Teso alphabet**

a b c d e g i j k l m n ŋ ny o p r s t u w y

Under British rule a Teso alphabet was developed, under missionary influence, by the Teso Orthography Committee. The Christian sects could not agree over spelling, however: Catholic and Protestant orthographies differ.

# THAI

**PERHAPS 25,000,000 SPEAKERS**

## *Thailand*

**T**he best known of the TAI LANGUAGES, Thai is the national language of Thailand, spoken by the great majority of the population in the central and southern parts of the country (see map at LAO).

Historically the Thai kingdom is the largest and most powerful of the many large and small Tai principalities that emerged in south-east Asia in the 12th and 13th centuries. The Chao Phraya valley, the centre of Thailand, was once ruled by the Mon kingdom of Dvaravati, and a small MON-speaking minority remains not far to the west of Bangkok.

From successive capitals – Sukhotai, Ayutthaya, Bangkok – Thai monarchs have achieved suzerainty over distant provinces and subject states, including Nan and Chiangmai (both now part of Thailand) and most of Laos. Bangkok was the only Tai capital that remained entirely independent through the colonial period, though it lost the Laotian states on the left bank of the Mekong, which were surrendered to France.

Thai society is firmly Buddhist. Linguistically, Sanskrit and Khmer were the early vehicles of Indic cultural and religious influence, now largely supplanted by Pali. The linguistic influence of Chinese is also very important. The Thai writing system (see box) is a careful adaptation of that of Khmer to a language with a very

different sound pattern.

Thai was from the beginning of its known history the official language of a monarchical state. It has a special vocabulary of respect, used in court ritual and in the addressing of royalty. Thai literature and art, which show Indian and Khmer inspiration, have developed a distinctive flavour. The older literature includes chronicles and histories as well as poetry and religious texts. There is a flourishing modern publishing industry. One distinctive feature is the production of funeral or memorial volumes. These are often the vehicle for publishing a work of literary scholarship or local history, dedicated to the memory of a friend or patron.

### **Numerals in Thai**

Thai numerals	Thai script	Transcription
1	๑	หนึ่ง nǒng
2	๒	สอง sǒng
3	๓	สาม sám
4	๔	สี่ sǐ
5	๕	ห้า hà'
6	๖	หก hok
7	๗	เจ็ด cèt
8	๘	แปด pǎet
9	๙	เก้า kǎw'
10	๑๐	สิบ sǐp

### **The Thai alphabet**

ก ข ค ด ม ง จ ฉ ช ซ ฌ ญ ฎ ฏ ฐ ท ธ น บ ป ผ ฝ พ ฟ ภ ม ย ร ล ว ส ห

k kh ng c ch s y d t th n d t th n b p ph f m y r l w s h

กะ กิ กี ฤ เกะ แกะ โกะ เกาะ เกอะ ก่า ไก ไก เกา กา กิ กิ ฤ เก แก โก กอ กัว เกีย เกื่อ เกอ

In the top line, of Thai consonants, many symbols appear to duplicate one another: they are used for borrowed Pali and Sanskrit words, and to help in representing tone distinctions. There are 25 vowel signs. To exemplify them, all 25 are shown in the bottom line in conjunction with the consonant *ṇ k*.

**The inscription of Ram Khamhaeng**

‘Previously these Thai letters did not exist. In 1205, year of the goat [AD 1283], Prince Ram Khamhaeng set all his will and all his heart on inventing the Thai letters, and the letters exist because the Prince invented them.’

The history of the Thai language and of Thai script begins with the inscription of Prince Ram Khamhaeng of Sukhotai, who in 1292 boasted of having established peace, law, trade, monasteries – and a new alphabet – at his capital city. Some Thai scholars believe the inscription (extract above) is ‘too good to be true’, and is a 19th-century forgery.

Whoever was its inventor, the Thai writing system was designed specifically for Thai (though redundant consonants were retained in order to render Pali and Sanskrit loanwords accurately). Until the present century, not Thai but Khmer script was used in Thailand in the writing and printing of Pali.



# THO

**1,000,000 SPEAKERS**

*Vietnam*

**T**ho is one of the TAI LANGUAGES, forming the southern end of the Northern and Central Tai dialect continuum (see map at ZHUANG). It is a major language of northern Vietnam. Tho speakers are a population of the plains and river valleys to the north and north-east of the Red River delta.

The name *Tho*, ‘soil’ in literary Vietnamese, is exactly comparable with Chinese *T’u-jen*, ‘people of the soil’, an older name for Zhuang. In Vietnam *Tho* is now considered pejorative and *Tay* is preferred.

In the same way that Zhuang speakers are more Chinese than most other minority groups in China, so Tho speakers are more Vietnamese in their culture than the other Tai peoples of Vietnam. An elite class, *Tho-ti*, is said to descend

from the Vietnamese officials who were first sent out to govern the Tho country in the 15th and 16th centuries. The Tho-ti, unlike other Tho speakers, are traditionally Mahayana Buddhists – and they officiate at rituals honouring the spirit of the soil, for, among the Tho, Buddhism is mixed with Confucianism, with ancestor worship like that of the Chinese and Vietnamese, and with the propitiation of spirits.

Tho has traditionally been written in a modified form of Chinese script, apparently based on the method once used for writing Vietnamese. This script was put to use mainly for recording songs. Genealogy has been an important matter to Tho speakers, and families kept a genealogy book which had to be consulted before a marriage could be agreed on in order to ensure that the couple were not related.

# TIBETAN

**PERHAPS 6,000,000 SPEAKERS**

*China, Bhutan, India, Nepal*

**T**ibetan is the language of the ancient and unique culture of the highest country in the world, the plateau and mountain chains that lie north of the Himalaya. It is one of the SINO-TIBETAN LANGUAGES.

Tibetan has a long recorded history. It was first written down in the 7th century, when a king in southern Tibet, Sron-btsan sgam-po, dispatched his minister Thon-mi Sambhoṭa to India to bring back information on Buddhism. The minister is said to have invented the script and to have written a grammar of Tibetan – on Sanskrit lines – in thirty stanzas (if only all grammars were as short as that). Translation of Buddhist texts into Tibetan was immediately begun. A Sanskrit–Tibetan dictionary, *Mahā-vyutpatti*, was compiled in the 9th century. Printing from carved wood blocks, a technique introduced from China, began very early and is still practised in some Himalayan monasteries.

Tibetan literature now forms a vast corpus, partly translated from Sanskrit and Chinese, partly original and inspired by the special paths that monastic Buddhism has taken in Tibet. The non-Buddhist Bon religion has its own literature. An unusual genre is that of ‘rediscovered’ texts, *gter-ma*, claimed as the works of venerated teachers of the distant past which have lain hidden for centuries in mountain caves.

Tibet was for many centuries an autonomous theocratic state, occasionally entirely independent, more usually interdependent in various ways with Mongolia and with China. Tibet is now a province of China, and since 1959 Chinese policy has tended to favour the destruction of Tibetan civilisation. However, many Tibetan speakers fled into exile in the 1950s and

1960s. Tibetan Buddhism and its literature are thus at present maintained by a worldwide diaspora, drawing some strength from Tibetan communities of the southern Himalaya beyond the Chinese border. Classical Tibetan is a literary language for all of these, still in use in Bhutan, Sikkim, Nepal and Ladakh. The Namgyal Institute of Tibetology, in Sikkim, is said to have the largest collection of Tibetan books in the world outside St Petersburg and Beijing.

Tibetan texts may contain Sanskrit proper names, but the language has few direct loanwords. However, written Tibetan is influenced by the word-building and the sentence structure of Sanskrit, which are closely imitated in translations and original texts on Buddhism. Tibetan, in turn, exerts influence: to Mongolians it is the classical language of Buddhism, widely taught in Mongolia till a few decades ago.

Classical Tibetan, as first committed to writing, seems not to have been a tonal language. Ladakhi, even today, has no tones. But the modern Tibetan of Lhasa is very definitely tonal (see box). The script and the various dialects allow linguists to glimpse the kinds of changes through which tones can arise in a language.

Tibetan words and names are very variously spelt in foreign writings. This is because Tibetan spelling, little changed from its 7th-century origins, now has a highly complex – though still regular – relationship with the way the modern language sounds.

## How tones originate

Written Tibetan, Amdo Sherpa, gLo-skad and Central Tibetan all represent different stages of

### The centre of the world

རྒྱ་གར་ཤར་གྱི་བྱ།

ཀོང་ཡུལ་མཐོལ་གྱི་ནེ་ཙོ།

འཁྱུངས་ས་འཁྱུངས་ཡུལ་མི་གཅིག།

འཛོམས་ས་ཆོས་འཁྱུང་རྩ་ས།

Rgya-gar śar-gyi rma-bya

Kong-yul mthil-gyi ne-tso

'khrungs-sa 'khrungs-yul mi gcig

'dzoms-sa chos-'khor Lha-sa.

A peacock from eastern India,

A parrot from deepest Kong,

Born and bred far apart,

Live together in holy Lhasa.

*Rgya-gar*, said to mean 'white country', is the Tibetan name for India. *Kong-po* in eastern Tibet is a region of deep, thickly forested valleys. This poem by the Sixth Dalai Lama (1683–1706) exemplifies the typical two-syllable rhythm of Tibetan literary writing, both verse and poetic prose. The emphasis falls on the first of each syllable pair.

tonogenesis: 1. Written Tibetan represents the original pre-tonal stage, with its various prefixes and suffixes and its oppositions between voicing, voicelessness and aspiration. 2. Amdo Sherpa represents a second stage. The suffixes remain as in Written Tibetan, but the effect of the prefixes has been lost except for nasal initials; however, as a result tonal distinctions now exist for nasals. 3. gLo-skad represents a third stage. Tones can only be set up for the nasals and the sibilants, but contrasting phonetic pitch patterns are emerging for all initials. 4. Central Tibetan represents a final stage, where syllable structures are extremely simplified and tonal distinctions exist for all series.'

Yasuhiko Nagano in *Linguistics of the Sino-*

*Tibetan area: papers presented to Paul K.*

*Benedict* (Canberra: Australian National

University, 1985) p. 462

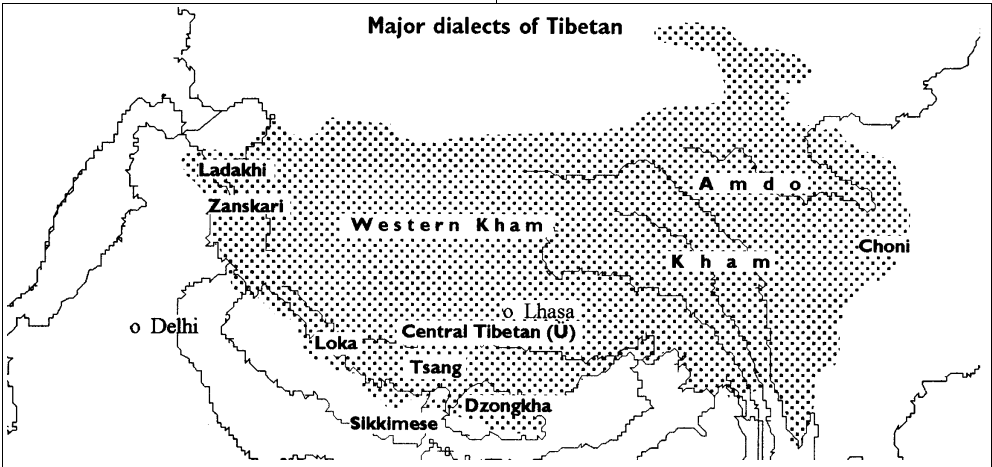
### Tibetan and its dialects

Tibetan is both a standard written language and a group of modern spoken dialects or languages, not all of which are mutually intelligible. Most of them share the culture and the classical literature of Tibetan Buddhism. As yet, too little is known of Tibetan dialects for adequate mapping. Besides 'Central Tibetan', the speech of Lhasa and its region, they include:

*Choni*, *Amdo*, *Kham* and other dialects spoken to the north and east of central Tibet, extending into the Chinese provinces of Qinghai and Yunnan.

*Bumthang*, *Tsangla*, *Ngalong*, the three major regional dialects of Bhutan (all mutually unintelligible), with over 1,000,000 speakers in total. *DZONGKHA* (see also map there), a lingua franca form of Ngalong, is now the national language of Bhutan.

### Major dialects of Tibetan



*Sikkimese*, the Tibetan dialect of Sikkim.  
*Kagate*, *gLo-skad* or *Loka* of Mustang, and other Tibetan dialects spoken by the ‘Sherpa’ and related peoples of Nepal.  
*Lahuli*, *Zanskari*, *Ladakhi*, *Purik* and *Balti*, Tibetan dialects of India and Kashmir, with perhaps 600,000 speakers in total. Many urban Ladakhi speakers are bilingual in Hindi, Urdu or

English. The standard form of Ladakhi is the dialect of Leh. Classical Tibetan literature can be read with a Ladakhi rather than a central Tibetan pronunciation.  
In Sikkim another Sino-Tibetan language exists alongside Tibetan: *Lepcha* or *Rong*, which has a local literary tradition and a distinctive alphabet. There are about 60,000 speakers.

The numerals in Tibetan				
Figures	Written form	Central Tibetan pronunciation	Ladakhi pronunciation	Dzongkha pronunciation
༡	གཅིག	chik	chik	dzong
༢	གཉིས	nyi	nyis	nyi
༣	གསུམ	sum	sum	sum
༤	བཞི	shi	zhi	zhi
༥	ལྔ	nga	shnga	nga
༦	དྲུག	dhuk	druk	ʈuk
༧	བདུན	dün	rdun	duin
༨	བརྒྱད	gyä	rgyat	gye
༩	དྲུག	gu	rgu	gu
༡༠	བཅུ	chu-thamba	schu	chu-thamba

The Tibetan alphabet	
ཀཁགང ཅཆཇཉ ཏཐདན བལའམ ཅཾཇཱ ཉེཊཱ འཟཊཱ རལའཤ ཏཱཱ	k kh g ng c ch j ny t th d n p ph b m ts tsh dz w sh z 'y r l ś s h a འཨཨཨ ཨཨཨ a i u e o

The Tibetan script

With its monosyllabic structure and complex initial consonants Old Tibetan was utterly different from Sanskrit and from medieval Indian languages. Thus, although recognisably based on an Indian model, the Tibetan script, devised in the 7th century, was built to a new specification, several separate characters making up a

typical single syllable, and each syllable marked off with a dot. There are 30 characters, shown in the top line, and five vowel signs, here shown with the character *a*.  
The formal version of the script, shown in the box, is ideally suited to woodblock printing, the traditional method of publication in Tibet. But it is very slow to write. Cursive forms exist for informal texts and for private use.

# TIGRINYA

**PERHAPS 3,500,000 SPEAKERS**

*Eritrea, Ethiopia*

**O**ne of the SEMITIC LANGUAGES, Tigrinya is the major language of Eritrea, and of the neighbouring province of Tigre in Ethiopia (for map see AMHARIC). The majority of Tigrinya speakers are Ethiopian Christians: there is a strong minority of Muslims.

Eritrea was conquered by Italy between 1882 and 1889. It was captured by the British in 1941 and transferred to Ethiopia in 1952, becoming independent, after a long and bitter civil war, in April 1993.

Both Tigrinya and Tigre are very similar to ETHIOPIC in their vocabulary. In its word structure, Tigre is closer to the classical language than is Tigrinya. The two modern languages are not mutually intelligible. Muslim Tigrinya speakers are often bilingual in Arabic.

Tigrinya, like Amharic, has seven vowels. As in other Semitic languages, plurals are normally

formed by internal vowel changes: *färäs* 'horse', *'afras* 'horses'.

To Eritreans, *Tigre* means 'the common people' or 'serfs': hence the name of the province Tigre. *Tigrinya* (sometimes written *Tigrīña*, or with the Italian spelling *Tigrigna*) is an Amharic form, 'the language of *Tigre* province'. The language is also called *Tigray* or *Tigrai* in earlier sources, but this is confusingly close to another language name, *Tigre*.

Tigre, with 100,000 speakers (or many more: estimates vary widely), is a quite distinct language of Eritrea. An early wordlist of Tigre was published by Henry Salt in his *Voyage to Abyssinia* in 1814. In Ethiopia people of Eritrea and of Tigre province are alike called 'Tigre': few are aware that Tigre and Tigrinya are different languages.

## Numerals in AMHARIC, ETHIOPIC and Tigrinya

	Amharic	Ethiopic: masculine and feminine		Tigrinya
1	and	'ahadū	'ahattī	hadə
2	hulätt	kələ'ētū	kələ'ētī	kələtte
3	sost	shalastū	shalās	sələste
4	arat	'arba'tū	'rbā'	'arba'te
5	amməst	khaməstū	khamś	hammushte
6	səddəst	sədəstū	səssū	shuddushte
7	säbat	sab'atū	sabəū	shob'atte
8	səmmənt	samantū	samānī	shommonte
9	zätäñ	təs'atū	təsəū	tesh'atte
10	asser	'ashartū	'ashrū	'asserte

# TIV

1,500,000 SPEAKERS

*Nigeria*

**O**ne of the NIGER-CONGO LANGUAGES, Tiv is the most important in the Southern Bantoid subgroup of Benue-Congo, and is thus one of the most closely related languages to proto-Bantu, the ancestor of the BANTU LANGUAGES of today. It has five tones, important in verb conjugation, and a system of nine noun classes.

*Tiv* is the speakers' own name for themselves: it is the name of the ancestor from whom they all claim descent. In the early 20th century most authors called them *Munshi*, the Hausa name for them. 'Hausa, noted for their folk etymologies, say it means "We have eaten them"', the answer Tiv are said to have made to Hausa traders seeking brothers who had disappeared in Tivland. Tiv today say that this name was first applied to them by Fulani, and that it was a herd of Fulani cattle which they had eaten. The term *Munshi* was used officially until the 1920s.'

By their neighbours, the speakers of Jukun, Tiv are known as *Mbitse* 'strangers' – possibly a sign that their present location is the result of migration. Certainly Tiv traditions told of a migration from hills in the south-east.

The earliest recorded visit to Tiv country by Europeans is in 1852, though earlier notes on the language had been made in Fernando Po in 1848 and in Freetown, Sierra Leone, in 1850, based on information from freed slaves.

The Tiv were an independent people, not grouped into a centralised state. Traditionally they are subsistence farmers, but many now migrate to work in the cities and towns of Nigeria. Tiv-speaking country was annexed by the British only between 1906 and 1912. It was

at this time, too, that Christian missionaries began to work among Tiv speakers, and produced translations and educational literature in the language, which had thus far been unwritten.

'The reputation of the Tiv has undergone several startling reversals. According to Baikie, Hutchinson and [the YORUBA linguist] Crowther they were a fearsome, savage lot of cannibals whose unsavoury reputation spread as far as the Gold Coast. This notion was current until some time after 1880 . . . During the later days of the Niger Company and the early days of the Protectorate, the Tiv became a solid, brave and respectable people defending their homes and preserving their independence against the constantly increasing danger of encroachment by the British. Some reports went so far as to say that the "Munshi" lived in dense forest land and formed an extensive, autocratic kingdom . . .

'In the opinion of the writers of this report, missionary influence has not been very great in Tivland; this, we feel, is attributable much more to the nature of the Tiv than to that of the missions.'

Law, inheritance and genealogy are complex matters in traditional Tiv Society; oratory is a crucial skill. 'An elder should be skilled in discussion (*ɔron kwagh*), that is, as a man of the moot-court (*or jil*) and as arbitrator. Such skill involves a knowledge of the individual and genealogical histories of all the people (and their immediate ancestors) of the lineage group (*non-go*) among whom he has influence.'

The first ten numerals in Tiv are: *mom*, *uhar*,

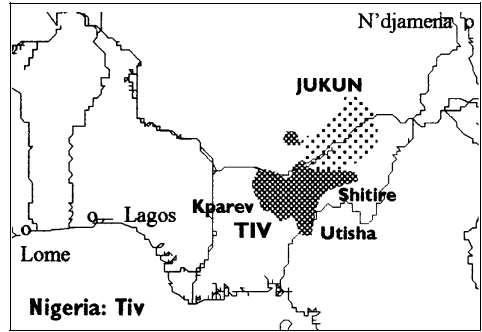
*utar, unyiyin, uta'an, uteratar, uta'an kar uhar, unyigenyi, uta'an kar unyiyin, puwe.*

Quotations throughout this entry are from Laura and Paul Bohannan, *The Tiv of central Nigeria* (London: International African Institute, 1953)

### The dialects of Tiv

Tiv is the language of Tiv Division in Benue-Plateau State, and has recently been spreading to Wukari, Lafia and Idoma Divisions. Some speakers live across the Cameroun border in Akwaya subdivision of South West Province.

'There are no dialects within Tivland itself. Tiv boast that they all speak the same language and that they can all understand one another. They tease one another, however, about accents (*liam*), which are roughly coordinated with area. There is the *Utisha* accent of south-eastern Tivland noted for its exaggerated palatalization of vowels. The *Shitire* accent (which includes most Tiv east of the Katsina Ala River) is much slower than other accents and has a tendency to merge all vowels into an indeterminate sound (like *a* in English *along*) . . . *Kparev* (central and south-central) accent is distinguished for rapidity and, according to some Tiv, contains additional sub-accent groups, like *Kunav*'s preference for *dj* sounds where other *Kparev* use *dz*. Vocabulary,



particularly the names of plants and of utensils, changes from one part of Tivland to another; this provides Tiv with conversation, however, rather than any major source of misunderstanding.

'Shitire accent has a slight dominance when Tiv is written, for it was here that Dutch Reformed Church missionaries first learned Tiv and reduced it to writing. However, *Kparev* accent is most widely known, in large part because *Kparev* trade more widely and hence travel more widely.'

The *Jukun* dialects are not closely related to Tiv but they have interacted with one another significantly. In total *Jukun* has about 125,000 speakers in the Gongola and Plateau states of Nigeria and in Cameroun; it is spoken by the *Wurku* river people of the middle Benue, and serves locally as a *lingua franca*.

# TOCHARIAN LANGUAGES

## EXTINCT LANGUAGES OF CHINA

In the finds of manuscripts at the old oases on the Silk Road, many religions and many languages were represented. There was great excitement when it was realised that two lost languages of about AD 600, first rediscovered in these texts, belonged to a previously unknown branch of the INDO-EUROPEAN LANGUAGES.

Incautiously, scholars named these languages *Tocharian* after a tribe encountered by Greeks in Afghanistan in the 2nd century BC. There is really no reason to link them to this name. They called their speech *Arshi-käntu*, apparently ‘Aryan’. The two languages are customarily distinguished as *Tocharian A* and *Tocharian B*.

Tocharian A, or *Turfanian*, was already a dead language when the texts we have were written down. It was used only in Buddhist scriptures and liturgies, and the people of Turfan and Karashahr needed glossaries or translations when they wished to understand it. Tocharian B or *Kuchean* was the ruling language of the Buddhist city-state of Kuqa, where it was soon to be replaced by UIGHUR. Tocharian B contributed loanwords to the Iranian language KHOTANESE, and influenced its script. Influences came in the other direction too: words such as *patsānk* ‘window’ and *sām* ‘enemy’ are of Iranian origin.

Although the Tocharian A and B manuscripts were found side by side, the two languages had actually been growing apart linguistically for hundreds of years. And it was soon found that Tocharian had quite astonishing relationships with the other Indo-European languages. Its strongest similarities were not with Indo-Aryan or Iranian languages, spoken a thousand miles away to the south-west, but with Celtic, Italic and Hittite, four thousand miles off. Like these, Tocharian belonged to the *centum* group (see box at INDO-EUROPEAN).

To explain this it is not necessary to imagine that the ancestors of the Tocharians once lived in western Europe. The true answer is that when the *satem* sound change occurred in Slavonic, Indo-Iranian, Albanian and Armenian, these were still Indo-European dialects spoken side by side. But by that time Hittite to the south, Celtic and others to the west and Tocharian somewhere to the east were all on the edge of the Indo-European dialect continuum, and none of them was affected by the change. So the Tocharian speakers were already on their way to their far-off destination among the cities of the Silk Road.

The first ten numerals in Tocharian A are: *sas*, *wu*, *tre*, *shtwar*, *pāñ*, *šāk*, *spāk*, *okāt*, *ñu*, *shāk*.



# TOK PISIN

## 2,000,000 SPEAKERS OF TOK PISIN AS A SECOND LANGUAGE

### *Papua New Guinea*

**T**ok Pisin is a pidgin language, or more correctly a creole, based on ENGLISH. It is the principal lingua franca of Papua New Guinea and one of the three official languages of the country, alongside English and HIRI MOTU.

*Tok Pisin* is the native form of the English words 'Pidgin Talk': in context this distinguishes it from the local languages, collectively referred to as *Tok Ples* 'language of the place'. It has also been called *Neo-Melanesian*, *Melanesian Pidgin*, *New Guinea Pidgin* or simply *Pidgin*.

Tok Pisin originated in the late 19th century as a form of the widespread English pidgin of the South Sea Islands, best known as *Beach-la-Mar* (see also Bislama). Once it had been introduced to eastern New Guinea – territories then under German and British influence – the pidgin spread rapidly. It served better as a lingua franca even than Motu, and was certainly more promising than English or German, because the clear structure of a pidgin aids rapid learning by speakers of differing linguistic backgrounds. This was all the more necessary, since, in the newly annexed colonies, people speaking numerous local languages had suddenly to learn to work together as servants, government employees, farm workers and seamen.

It may well be that the pidgin prospered better under German rule, in the north-east, than under the English in the south-east, since Germans, though they disliked using an English-based lingua franca, could not avoid taking the trouble to learn it properly. Australians, who ruled both the former British and the former German territories during most of the 20th cen-

tury, tended to disapprove of the pidgin equally strongly. The Japanese occupation during the Second World War forced a revision of this view: pidgin, already used by missionaries, now became the principal language of Australian as well as Japanese propaganda.

In independent Papua New Guinea Tok Pisin is the principal language of the media and of town life. It is the mother tongue of well over 100,000 young people, a number that is growing rapidly. Meanwhile, standard English is widely taught and is infiltrating Tok Pisin ever more strongly. Other, older influences can still be traced: it has been suggested that the absence of voiced final stops (so *rot* 'road'; *dok* 'dog') shows the phonetic influence of German. Like Austronesian languages, Tok Pisin has a rich series of personal pronouns: for the first person these are *mi* 'I', *yumitupela* 'we, including person addressed, total two', *yumitripela* 'ditto, total three', *yumi* 'ditto, indefinite total', *mitupela* 'we two, excluding person addressed', *mitripela* 'we three, ditto', *mipela* 'we all, ditto'.

Examples from John Holm, *Pidgins and creoles* (Cambridge: Cambridge University Press, 1989) pp. 529–34

#### **'Rude words' in Tok Pisin**

Standard Tok Pisin includes several important words whose English originals are 'improper'. In the 19th-century canefields, plantations and mines, language was used forcefully. The formerly 'rude words' are perfectly polite and proper in Tok Pisin. *As lo bilong gavman* ('arse law', basic law of the Government) means 'constitution'; *mi bagarap* means 'I'm tired'.

# TONGA

900,000 SPEAKERS

*Zambia, Zimbabwe*

**T**onga is one of the BANTU LANGUAGES and one of the eight official languages of Zambia. Speakers of Tonga, Ila and Lenje (see map at BEMBA) are together known as *Bantu Botatwe*, the ‘three peoples’. It is from this word *bantu* ‘people’ that the name of the whole language family is derived.

The history of these three peoples, so far as it can be traced, is one of subjection to raids and conquests, culminating in that of the British. The Kololo, who introduced the LOZI language to the Zambezi valley, annexed the Ila and the southern half of Tonga-speaking territory in the 1830s. The southern dialect of Tonga now shows noticeable Lozi influence. Until 1906 slavery was practised locally and Tonga and Ila speakers were often taken as slaves by the Lozi, some to be sold on to MBUNDU traders.

Under British suzerainty, beginning in 1890, Tonga-speaking territory, which lay astride the Zambezi with a majority to the north, was split between Northern Rhodesia (Zambia) and Southern Rhodesia (Zimbabwe). The most productive lands of the Three Peoples were seized by colonial settlers and the slave system was gradually replaced by one of agricultural and migrant labour, many Tonga speakers working in the mines of southern Africa.

The Twa of the Kafue and Lukanga swamps are perhaps of Bushman origin (see KHOISAN LANGUAGES), but they now speak Tonga, Ila and Lenje like their neighbours.

*Ka zuba o mwana kulu mwinzhila:* If you hide with a child, there’ll be a leg sticking out in the road – or, when you go into danger, choose your companions carefully. *Matako aswangene ta budi mutukuta:* When buttocks meet, there’s always sweat – or, if you spend your life with someone, you’re bound to quarrel.

After Edwin W. Smith, *A handbook of the Ila language* (London: Oxford University Press, 1907)

Numerals in Tonga and Ila		
Tonga		Ila
-mwi	1	-mwi
-bili	2	-bili
-tatu	3	-tatwe
-ni	4	-ne
-sanwe	5	-sanwe
-sanwe a -mwi	6	chisambomwi
-sanwe a -bili	7	chiloba
-sanwe a -tatu	8	lusele
-sanwe a -ni	9	ifuka
ikumi	10	ikumi

As with most other Bantu languages, a singular or plural prefix must be added to the numerals ‘1’ to ‘5’ agreeing with the class of the noun that is qualified. In Tonga, the numerals from ‘6’ to ‘9’ require two prefixes.

# TONGAN

130,000 SPEAKERS

*Tonga*

**T**ongan is a Polynesian language, belonging to the Oceanic branch of AUSTRONESIAN LANGUAGES. It is the official language of the kingdom of Tonga in the western Pacific.

Tonga was an influential place in medieval times, according to oral historical tradition. There was a Tongan Empire that extended to Samoa, Niue, Rotuma, Futuna and East Uvea. Recent linguistic research supports this, identifying Polynesian – most probably Tongan – loanwords in Vanuatu and in languages as distant as Kiribati, Ponapean and Mokilese.

Tonga, the ‘Friendly Islands’ of European explorers, became a British protectorate in the 19th century but remained self-governing. Wesleyan and French Catholic missionaries competed for converts, and developed different spellings in their grammars and religious texts in Tongan. The current spelling system was laid down by the Privy Council of Tonga in 1943.

Tongan does have loanwords from English, Fijian and Samoan. But all through the colonial period the language has remained in official use in education and government, and loanwords are balanced by new words formed from local roots: *tangata* ‘man’ gives such new terms as *tangata-tui-su*, ‘man who sews shoes, i.e. cobbler’, the *su* being a loanword from English.

The first ten numerals in Tongan are: *taha*, *ua*, *tolu*, *fā*, *nima*, *ono*, *fitu*, *valu*, *hiva*, *hongofalu*.

Based on Paul Geraghty, ‘Tongan’ in *Comparative Austronesian dictionary* ed. Darrell P. Tryon (Berlin: Mouton De Gruyter, 1995–) pt 1 pp. 937–42 and other sources

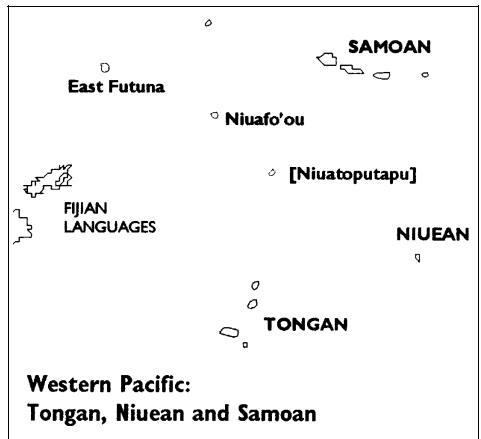
## Tongan, Niuean and Samoan

*Tongan* is spoken throughout Tonga, though the native language of one northern island, *Niuafo’ou*, is distinct and rather closer to Samoan. In 1616 the Dutch mariner Jacob le Maire made a wordlist of the language of another Tongan island, *Niuaotoputapu*. This also was close to Samoan: but Tongan is what is now spoken by the 1,500 modern inhabitants of Niuaotoputapu.

Le Maire also made a wordlist of *East Futuna*, a language of the Futuna or Hoorn Islands in the territory of Wallis and Futuna. It now has 3,000 or more speakers.

*Niuean* has about 10,000 speakers (but only about 2,000 on the island of Niue) and is related to Tongan. Niue is a self-governing territory dependent on New Zealand, where the great majority of speakers live.

SAMOAN is the language of Western Samoa and American Samoa. Linguistic differences between the two jurisdictions are very slight.



# TSONGA

**4,000,000 SPEAKERS OF TSONGA, RONGA AND TSWA**

*South Africa, Mozambique*

**T**songa, Ronga and Tswa are three BANTU LANGUAGES so close to one another that speakers of any one can understand the other two with little difficulty.

*Thonga*, the oldest form of the name, is of Zulu origin and means 'vassal'.

In the early 19th century the Shangana, warriors of ZULU and Xhosa origin, established the Gaza Empire in this region, with its capital at Mandlakazi on the lower Limpopo. Portuguese rule was established by encouraging civil war between rival sons of the conqueror, Manukosi, who died in 1859.

The Swiss missionaries of the Paris Evangelical Mission Society, who first developed a written form of this language in the 1870s, based their early work on the *Gwamba* variety, a lingua franca spoken by recent migrants from several parts of the Mozambique coast to the Spelonken district of Transvaal, south of the Levubu river. The missionaries learnt the language through Sotho, well known as a second language in northern Transvaal. Their first publication, widely known as 'the *buku*', was a Bible reader, printed in Switzerland in 1883. Later missionary writings incorporated features of the *Djonga* of the coastal plain between the Limpopo and Nkomati.

Partly as a result of rivalry among missions, partly because of local resistance to the despised Gwamba dialect and perhaps to its Sotho colouring, a separate religious literature was soon afterwards developed in the *Ronga* spoken south of the Nkomati in the Lourenço Marques (Maputo) region. The first primer in Ronga appeared in 1894.

The special features of *Tswa* are explained by the fact that it was the dialect of the Gaza

Empire. All forms of Tsonga came under Zulu influence, and now have Zulu loanwords, but even the phonetics of the language have been affected and in some dialects click consonants, typical of Zulu, can be heard.

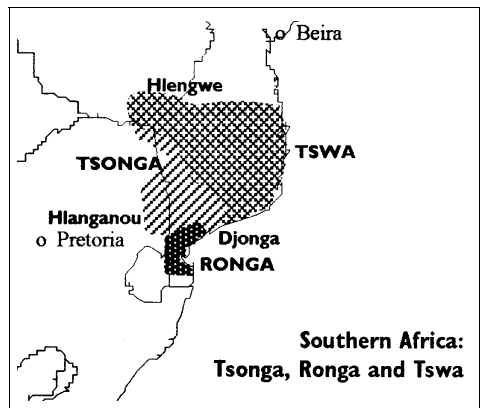
The first ten numerals are *-ngwe*, *-biri*, *-raru*, *mune*, *ntlhanu*, *ntlhanu na -ngwe*, *ntlhanu na -biri*, *ntlhanu na -raru*, *ntlhanu na mune*, *khume*.

## Tsonga, Ronga and Tswa

*Tsonga* or *Thonga* or *xiTsonga* counts 1,500,000 speakers in Mozambique, where it is known as *Shangaan*, and 1,300,000 in South Africa, where it is one of the eleven official languages. It is a grouping of dialects including *Djonga*, *Khosa* and *Hlanganou*.

*Ronga* or *Landim* is spoken north-west of Maputo, while the Maputo and Tembe dialects lie to the south. The three dialects may total 500,000 speakers.

*Tswa* has 700,000 speakers north of the Limpopo in Mozambique. *Hlengwe* is the inland dialect.



# Tswana

**4,500,000 SPEAKERS**

*South Africa, Botswana*

One of the BANTU LANGUAGES, Tswana is mutually intelligible with Northern SOTHO (see map there); but as the national language of Botswana, where there are about 850,000 speakers, it has developed a separate identity. Tswana functions as a lingua franca all over Botswana, and is important in South Africa.

The speakers of *seTswana* are *baTswana* in their own language. Some say that the word means 'those who went out' (*-tswa*, 'to go out') or 'the separators' (*-tswaana*, 'to separate') – and certainly the oral traditions of the Tswana peoples tell of repeated fragmentation of political groupings.

Southern dialects of Tswana have a palatal fricative: for example, Rolong dialect *-chwa* 'to go out' instead of standard *-tswa*. Since the southern *baTswana* were the first to be encountered by English-speaking explorers, the standard form of the ethnic term in British writings became *Bechuana*, with the southern palatal fricative: hence the former English names of colony and protectorate, *Bechuanaland*.

Historically, then, Tswana is the Sotho dialect of the people settled in the upper valley of the Limpopo. That settlement dates from at least four centuries ago, and probably more. They spread westwards in the 18th and 19th centuries until their territory included the south-eastern part of modern Botswana: at the same time Kololo (see LOZI) and Ndebele invaders were encroaching from the east, and Afrikaans and English-speaking colonists had begun to seize much of their

southern and eastern lands. When colonial boundaries were drawn Tswana-speaking territory was split between western Transvaal, north-eastern Cape Colony (briefly 'British Bechuanaland') and a third division whose chiefs successfully petitioned the British Government for direct rule. This became the British 'protectorate' of Bechuanaland and is now independent Botswana.

Ironically Tswana had low status in Transvaal, where the majority of its speakers live, yet became the 'national' language in Botswana, where speakers of Tswana as a mother tongue are the – politically dominant – minority. In Botswana many speakers are bilingual in local languages, both Bantu (e.g. Herero) and KHOISAN. The *Kgatla* dialect is the basis of the standard language in South Africa, where Tswana is now one of the eleven official languages.

Christian missions arrived in 1816, and gradually spread the knowledge of writing. The first book on Tswana, *A grammar of the Bechuana language*, was published by the Wesleyan missionary James Archbell in 1838. Tswana is written with frequent word divisions (*ke a e bona* for *kealebona* 'I see it') because of the missionaries' first attempts to grasp the grammar of the language, still enshrined in the orthography that they devised for it. Tswana was the African language learnt by the explorer David Livingstone.

There is now a well-established literature in Tswana, beginning with Bible translations but including a variety of formerly oral literature – legends, proverbs and riddles, praise poetry, songs and traditional history.

Numerals in Sotho and Tswana			
	Southern Sotho	Northern Sotho	Tswana
1	-’ngoē	-tee	-ngwe
2	-beli	-bedi	-bêdi
3	-rarō	-raro	-raro
4	-nè	-nê	-nê
5	-hlanō	-hlano	-tlhano
6	-tshêletsēng	selêla	-rataro
7	-supileng	shupa	shupa
8	-robileng mēnò ēlē ’meli	seswai	fêra mebêdi
9	-robileng mōnò ōlē mōng	senyane	fêra mongwe fêla
10	lēshōmè	lesomê	leshomê

# TULU

**1,200,000 SPEAKERS**

*India*

*Tulu* is one of the DRAVIDIAN LANGUAGES of India, and a geographical surprise. Sandwiched between the much larger numbers of Kannada speakers to the north and Malayalam speakers to the south, Tulu is spoken in a small district of coastal Karnataka – its traditional boundaries being the Chandragiri and Kalyānapuri rivers (see map at KANNADA). But it is very definitely a dialect neither of Kannada nor of Malayalam. It differs strongly from both, and is even in some ways reminiscent of the Central Dravidian languages of the inland minority peoples further north.

Missionaries at Mangalore published a few books in Tulu using Kannada script: on the few occasions when Tulu continues to be printed, this script is still used. Although there is very little written literature in Tulu, it is a language with a very powerful tradition of oral literature.

The dialect spoken by the Brahman caste is notably different from standard Tulu. Most speakers of the language are bilingual in Kannada, the state language of Karnataka.

The first ten numerals in Tulu are: *onji*, *raḍḍ*, *mūji*, *nāl*, *ein*, *āji*, *ēl*, *eṇma*, *ormba*, *patt*.

# TUMBUKA

**1,500,000 SPEAKERS**

*Malawi, Zambia*

**T**umbuka, one of the BANTU LANGUAGES, is spoken in Malawi and a few districts of eastern Zambia (see map at NYANJA).

According to oral tradition, a group of ivory traders from the Indian Ocean coast established a kingdom – the Chikuramayembe dynasty – in the Tumbuka country in the late 18th century, unifying what had until then been separate small chiefdoms. Much of the region was conquered in the 1830s and 1840s by Ngoni warriors from Zululand. Thereafter the Ngoni formed a ruling class. In much of Tumbuka country and around Chipata (Fort Jameson) people nowadays consider themselves to belong to the Ngoni people:

yet their languages are scarcely distinguishable from those of their Tumbuka- and Nsenga-speaking neighbours. They are said still to use the real Ngoni – a variant of 19th-century Zulu – in songs and royal praise poems. It certainly remained in use in lawsuits for nearly a century after the migration.

Tumbuka is not one of the official languages of Malawi or Zambia. Nyanja and English are used for all official purposes. The first ten numerals in Tumbuka are: *-moza, -wiri, -tatu, -nayi, -nkonde* or *-sanu, -sanu na, -moza, -sanu na -wiri, -sanu na -tatu, -sanu na -nayi, kumi*.

## Tumbuka praise poetry

Chakatumbu ka usipa,  
chituwi wakusonyora,  
cha maso nga ndi mere,  
cha mpuno nga ndi ngolo;  
sango za pa luji kwawika Nyirenda!

Slim-bellied as the whitebait,  
prickly-spined as the *chituwi* fish,  
red-eyed like the *mere*,  
long-nosed like the *ngolo*;  
the *sango* of the deep lake are offered to Nyirenda!

Praise formula for Kawunga, a vassal of Chikuramayembe, whose people became known as *nyirenda* ‘centipedes’.

T. Cullen Young, *Notes on the speech of the Tumbuka-Kamanga peoples in the Northern province of Nyasaland* (London: Religious Tract Society, 1932) pp. 171–3