

TUNGUSIC LANGUAGES

The last Imperial family that reigned in Beijing, the Qing or Manchu dynasty, seized power in 1644 and were driven out in 1912. Manchu was the ancestral language of the Qing court and was once a major language of the north-eastern province of Manchuria, bridge-head of the Japanese invasion of China in the 1930s.

It belongs to the little-known Tungusic group of languages, usually believed to form part of the ALTAIC family. All Tungusic languages are spoken by very small population groups in northern China and eastern Siberia.

Manchu is the only Tungusic language with a written history. In the 17th century the Manchu rulers of China, who had at first ruled through the medium of MONGOLIAN, adapted Mongolian script to their own language, drawing some ideas from the Korean syllabary. However, in the 18th and 19th centuries Chinese – language of an overwhelming majority – gradually replaced Manchu in all official and literary contexts.

The Tungusic languages

Even or *Lamut* has 7,000 speakers in Sakha, the Kamchatka peninsula and the eastern Siberian coast of Russia.

Evenki or *Tungus* is the major Tungusic language of Russia, with 12,000 speakers – widely scattered in the Siberian forests – and perhaps another 10,000 in China.

Manchu has the honourable status of an official nationality of China – but fewer than 1,000 surviving speakers. However, a variety known as *Colloquial Manchu* or *Xibo*, transplanted with a Manchu garrison to fortresses in Xinjiang centuries ago, still numbers about 27,000 speakers.

Nanai or *Goldi* has about 7,000 speakers on the banks of the lower Amur.

Orochen has about 2,000 speakers in northern Manchuria.

Several other Tungusic languages survive, with only a few hundred speakers apiece.

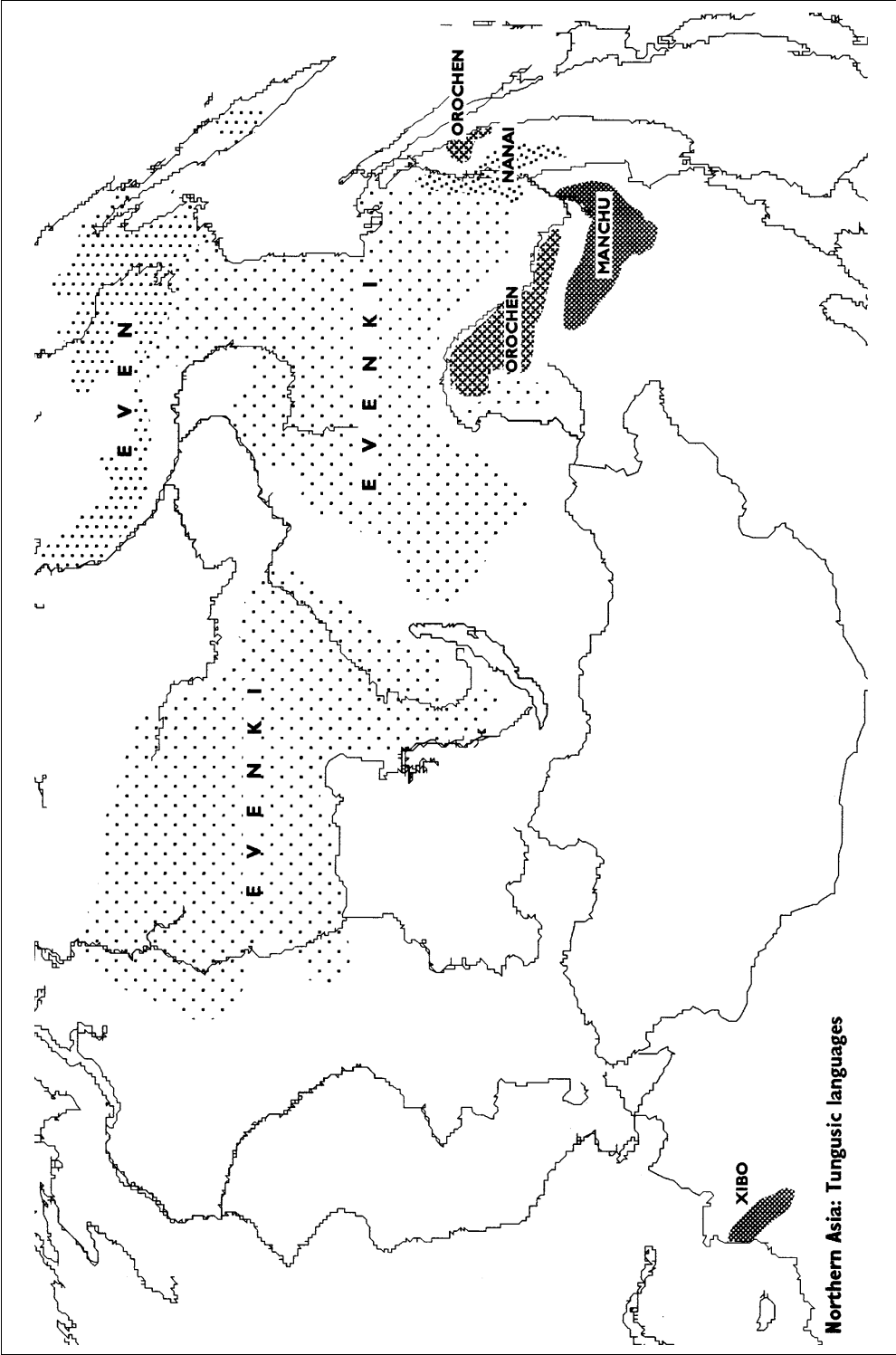
Numerals in Manchu, Evenki and Nanai

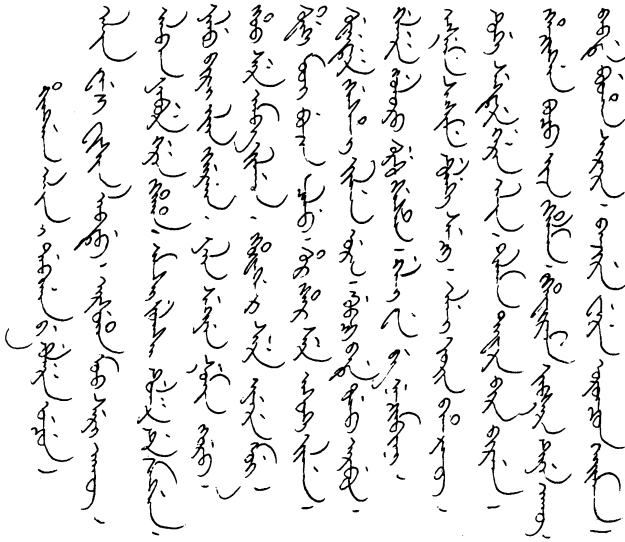
	Manchu	Evenki	Nanai
1	emu	umūn	emun
2	juwe	dyūr	dyuer
3	ilan	ilan	ilan
4	duin	digin	duin
5	sunja	tungga	toinga
6	ninggun	nyungun	nyungun
7	nadan	nadan	nadan
8	jakon	dyapkun	dyapkun
9	uyun	ēgin	khuyun
10	juwan	dyān	dyoan

From George L. Campbell, *Compendium of the world's languages* (London: Routledge, 1991)

The mountain forest

Manchu literature now has few readers. There was once a strong tradition of oral poetry, which can still be sensed in the written texts despite their dependence on Chinese genres. This is a rare example of Manchu lyric song, remembered from his childhood by Professor Yadamsürengiin Shariibuu. It reminds us that the Manchu had completely separated themselves from their linguistic relatives, such as the Orochen, who remained hunters and foragers in the Siberian forest.





Crossing a pass in the Hinggan Mountains

The mountain forest is dense and wide and full of thorns:
 We messengers suffer both coming and going.
 The Yalu has three streams, their sources deep and distant,
 Their sound as of water flowing, surging, roaring.
 Among the leaves of the forest a noise of horses,
 An echo of birdsong. Darkness all around.
 No one is safe from wild beasts and birds.
 We hurry endlessly – we never arrive.
 Every mountain is cloud-covered, every stream is hidden in fog,
 Gnats and mosquitoes attack us restlessly.
 And, all around, the *Orochen* live by their hunting.

After Juha Janhunen, 'A Manchu song from Mongolia' in
Journal de la Société Finno-Ougrienne vol. 77 (1981) pp. 207–18

TUPÍ

3,000 SPEAKERS

Brazil

Tupí, or Tupinambá, was originally the name for a group of dialects spoken along the eastern coast of Brazil when the Portuguese first began to explore the country. Tupí and GUARANI (see map there), with some minority languages, together form a group which may belong to the postulated family of AMERIND LANGUAGES.

The Portuguese settlement led to the evolution of a standardised, simplified form of Tupí, often called *Lingua Geral*, 'Common Language' – a development that has been compared to that of the Hellenistic form of GREEK, whose usual name *koinē* also means 'Common'. In the 17th and 18th centuries this standard Tupí was spoken throughout Portuguese Brazil, and religious texts were published in it: Antonio de Araujo's *Catechism* appeared in 1618.

Initially it was probably a second language everywhere: but it soon began to serve as a mother tongue for the children of mixed Portuguese-Tupí households, and for town-dwellers generally. Jesuit missionaries, in an attempt to isolate their Amerindian converts from the demoralising effect of contact with the Portuguese community, in some places preferred not to teach them Portuguese but *Lingua Geral*.

But the language of the elite was Portuguese. By the 19th century, especially with the establishment of an independent Brazilian court and government, *Lingua Geral* rapidly lost ground in coastal Brazil.

Meanwhile, however, it had been found useful by explorers, traders and settlers striking inland into the vast Amazon basin. Here, although the inhabitants did not originally speak Tupí dialects, the relatively simple *Lingua Geral* emerged as a more convenient medium of communication than Portuguese or any local language. Thus, rapidly changing, it was to have a new lease of life in the 19th and early 20th centuries as the *lingua franca* of the inland regions of Brazil.

Portuguese has now taken its place as first or second language almost throughout Brazil, though along the Rio Negro there are still about 3,000 speakers of modern Tupí, *Lingua Geral* or *Nheengatú*.

Many Tupí loanwords are to be found in modern Portuguese. Some have become international words, such as *mandihoca*, Portuguese *mandioca* 'manioc, cassava', which has spread as far as French and English – and all the way to Swahili *mahogo*.

TURKANA

650,000 SPEAKERS OF TURKANA AND KARAMOJONG

Kenya, Uganda, Sudan, Ethiopia

The Karamojong group of NILO-SAHARAN LANGUAGES is spoken in East Africa. Turkana, its best-known member, is the language of the hot, dry Rift Valley of Kenya: the massive western escarpment of this, which divides Kenya from Uganda, also divides Turkana from the related karamojong. Seen from the escarpment, Turkana country appears a vast sandy plain far below, where the flat scenery is relieved by isolated mountain blocks, and where dust devils rise in high columns for most of the day. On descending, it seems at first glance impossible for men or animals to live there.

Turkana speakers call themselves *ngiTurkana* (singular *eTurkanait*) and their territory *Turk-wen*. Throughout it, they are traditionally divided into two groups, *ngimonia* 'people of the forest' and *ngicuro* 'people of the rough plains'.

Turkana speakers are a pastoral people, keeping goats, sheep and cattle and living off their milk, blood and meat as well as off berries and nuts. They were visited by Swahili traders, in the 19th century and perhaps before. The country was sketchily under Ethiopian administration from the 1880s, but the British Africa Rifles conquered most of it between 1909 and 1926, and these conquests went to Kenya (the majority) and Sudan.

Apart from soldiers, few outsiders have spent long in Turkana country: among early expeditions that ventured in, there were several deaths from starvation. Only the bravest of linguists have been attracted to this part of the Rift Valley.

The first ten numerals in Turkana are: *a-pey*, *nga-arey*, *nga-uni*, *nga-omwon*, *nga-kànj*, *nga-*

kani-ka-pey, *nga-kani-ka-arey*, *nga-kani-ka-uni*, *nga-kani-ka-omwon*, *nga-tòmwn*.

How did Turkana come to be the language of the Rift Valley? 'There is a story that a poor Jie woman went down the escarpment looking for food and reached Moru Naiyece in the Tarash valley, where she found plenty of berries. Some Jie searching for a lost bull eventually found it with the woman and, seeing that there was food, some of them moved down and settled there, becoming the ancestors of the Turkana tribe.'

G. W. B. Huntingford, *The Northern Nilo-Hamites* (London: International African Institute, 1953) p. 12

The Eastern Nilotic (Nilo-Hamitic) languages

The *Bari* division of this group consists of Bari itself, *Kakwa* and *Mandari*. Its centre is in southern Sudan.

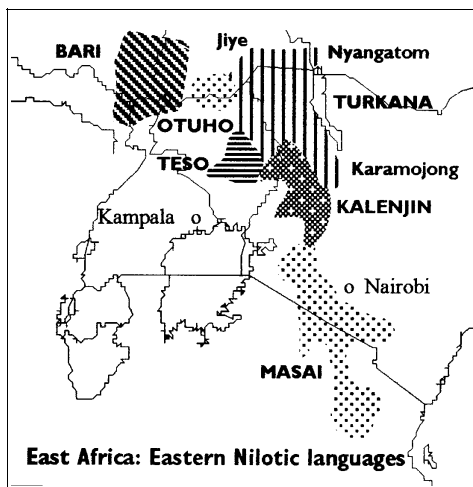
The *Karamojong* group includes Turkana of Kenya and – on the highlands surrounding the Rift Valley – Karamojong itself, which is spoken in slightly varying dialects by the Karamojong, Jie and Dodos of Uganda and the little-known Toposa and Jiye of Sudan. The form of Turkana spoken in Ethiopia and the south-eastern corner of Sudan is called *Nyangatom*.

TESO or *Iteso* is the language of a neighbouring, far more accessible region in Uganda. Its speakers share many cultural traits with their immediate neighbours, speakers of Bantu languages. The dialect of Ngora is the usually accepted standard. The *i-* of the alternative language name shows it to be an inappropriate

form, for it is the male gender prefix: *ítèsò* 'Teso man', *átèsò* 'Teso woman'.

MASAI, *Otuho* and their relatives form the *Lotuko-Maa* division of Eastern Nilotic. Masai is spoken in an extensive, sparsely populated territory in central Kenya and Tanzania. Masai and *Samburu*, its northern dialect in Kenya, are mutually intelligible but have usually been treated as different languages. *Otuho* or Lotuko is a language of southern Sudan.

KALENJIN is the usual name now for the *Nandi* group of dialects, which includes Nandi itself, *Kipsigis*, *Päkot* and others. The group includes Sebei, pastoralist neighbours of the agricultural Gisu and thus included in the local term *baMasaba*, 'people who live on Mount Elgon' (see also LUYIA).



Karamojong praise poetry of the Jie

On ritual and other occasions, the praise songs of the grandfathers' generation were recorded, one of which began:

Nyamonía a ngikosowa
awatar angítome –
ioye, nyengori.

The forest of buffaloes:
the elephants were standing there,
O, the dark grey ones.

and the other:

Toremo, nyetome, i toremo.
Atome ayong.

Spear the elephant, spear it.
I am the elephant.

John Lamphear, *The traditional history of the Jie of Uganda* (Oxford: Clarendon Press, 1976) pp. 41–2

TURKIC LANGUAGES

Like others of the ALTAIC family, Turkic languages originated in central and northern Asia, where, as their recorded history dawns about 1,500 years ago, they were spoken by already widespread communities of nomadic horse-breeders, shepherds and metal-workers.

Their first steppe empire was established, in a revolt against the little-known Juan-Juan, in 552–5. It was centred in western Mongolia, and its diplomatic contacts extended from Byzantium to China. Turkic languages, all quite closely related to one another, are now spoken across a wide swathe of Russia and central Asia, even extending to the Near East and south-eastern Europe – as a result of a series of epic migrations and conquests. Turkic speakers did not always initiate these wanderings, but their horsemen were typically the majority among the warriors. Thus their languages remain as evidence of the great empires, while MONGOLIAN LANGUAGES are far less widespread.

Islam had begun to find adherents among Turkic speakers by the 11th century. Thus Turkic languages today are spoken chiefly by Muslims – though this does not apply to Chuvash or Yakut. Most were once written in the Arabic script. However, the Cyrillic alphabet, long used for Chuvash under Russian influence, has also been used for all the other Turkic languages of the old Soviet Union. Turkish, Azeri and Uzbek have adopted the Latin alphabet. Uighur, one of the most significant minority languages of China, is written in its own script, a descendant of the ancient Syriac alphabet which came to western China with the Manichaean religion thirteen hundred years ago.

Even while they were developing along separate lines, the Turkic languages were mostly alike in admitting numerous important loanwords from Arabic and Persian – words so basic to every language that they are seldom borrowed from one to another. Arabic *wa* ‘and’ is found in Azeri *va*. The Persian subordinating conjunction *ke* re-

appears in Azeri, Turkish and Uzbek *ki*, though the logic of the surrounding sentence is quite different in Persian and in these Turkic languages.

Traditional literature in Turkic languages is notable for epics of heroic adventure, for shorter romantic tales (*hikaye*), and, equally, for brief love poems of haunting beauty. Oral epic poetry has been collected from many Turkic peoples in the 20th century, and it is still to be heard in some places.

The geography of Turkic conquest is reflected not least in Turkic proverbs. Of a suburb of Istanbul: *Kasımpaşa, eli maşa*, ‘The men of Kasımpaşa often carry knives’. Of two towns in Xinjiang: *Sayramdın uğrı, Kuçadın güi*, ‘Sayram for its thieves, Kuça for its cuckolds’.

Examples from *Philologiae turicae fundamenta* vol. 2 ed. Pertev Naili Boratav and others (Wiesbaden: Steiner, 1965) p. 69

The spread of Turkic languages

The oldest Turkic migration that has left linguistic traces is that of the Bulgars, feared in early medieval Europe: their remnant, now peacefully settled near Kazan’ (in part of medieval Great Bulgaria), speak CHUVASH. Although unrecorded in history, the original split of early Chuvash from proto-Turkic must have come two thousand years ago or more.

Then proto-Turkic began to divide into dialects. To the north Altai, Khakas and Tuva speakers (see map 2) have probably migrated little, while the related YAKUT spread gradually further north-eastwards into the Siberian forest and tundra. To the south the Uighur settled and ruled the cities of the Silk Road. This is the origin of UIGHUR and UZBEK, the Eastern or Chagatai group of Turkic languages.

With later empires came the spread of noma-

dic Turkic speakers of the central Asian mountains. Beginning from what are now Kazakhstan and Kyrgyzstan, this Western or Kipchak group spread westwards with the conquering horsemen. Thus the group includes KAZAKH with Karakalpak and Nogai, KYRGYZ, KUMYK with Karachai and Balkar, TATAR and BASHKIR.

South-westwards the Seljuk Turks extended their rule across western Asia and beyond. In their wake, the Southern, Turkmen or Oghuz group of Turkic languages, TURKMEN, AZERI and TURKISH gradually spread across parts of Iran, Azerbaijan, Turkey and the Balkans. Salar, a language of 55,000 speakers in north-western China, also belongs to this group.

The Northern Turkic languages

Altai with 50,000 speakers, Khakas with 60,000 and Tuva with 200,000 – and the minority Shor and Tofa – are all languages of Asiatic Russia. The three peoples were converted to Buddhism through Mongolia and have traditionally been under Mongolian cultural influence. Until recently Mongolian was the written language that they used. These three languages are now written in Cyrillic script.

Tuva or Tuvinian was the language of Tannu Tuva, a state that remained precariously independent between 1921 and 1944 but was then incorporated in the Soviet Union. Buddhism, 'feudalism' and traditional agriculture were gradually undermined, leaving Tuva with severe social problems of youth crime, alcohol and drug abuse. There are also speakers across the border in Mongolia. Tuva's Buddhism came by way of Mongolia, and the language has many Mongolian loanwords. The Mongolians call this people *Tsaatan* 'reindeer herders'.

Khakas is a group of Turkic dialects spoken by disparate minorities in the self-governing republic of Khakassia in Russian Siberia – and also spoken by the 'Yellow Uighurs' in the Chinese province of Gansu. The term Khakas was invented in early Soviet times.

Altai (Oïrot) is spoken in the Mountain Altai Region. A small language, it has the distinction of sharing its name with the very diverse linguis-

tic family of which the Turkic languages form one branch. A significant part of the Altai country is contaminated by radiation as a result of the Semipalatinsk nuclear test in 1949: the population here is in decline.

The First Türk Emperors

The 'Orkhon inscriptions', discovered in the Orkhon valley in Mongolia and in the upper Yenisei valley in Siberia, are the oldest full-length texts in a Turkic language. They are major historical sources for the early central Asian empires. They show justifiable pride in the achievements of the ruling family. The rhythmic style displays a haunting parallelism – repetition with variations – which is typical of oral and traditional literature in many languages. The inscriptions are written in a 'Runic' script, so called because it looks rather like the runes of early Scandinavia.

When the blue sky high above and the brown earth down below had been created, between the two were created the sons of men; and above the sons of men stood my ancestors, the Kaghans Bumin and Ishtemi. Having become masters of the Türk people, they established and ruled its empire and fixed the law of the country. Many were their enemies in the four corners of the world, but they campaigned against them. They conquered and pacified many nations in the four corners of the world: they made them bow their heads and bend their knees.

These were wise kaghans, these were brave Kaghans, and all their officers were wise and brave, and all the nobles and all the people were just. Thus it was that they were able to master so great an empire, and to govern it, and to uphold the law.

8th-century Orkhon inscription. Translation after Denis Sinor, *Inner Asia: history, civilization, languages* (Bloomington: Indiana University, 1969) p. 103

Language of the Huns

Little is known of the language spoken by the briefly powerful and destructive Huns, who burst into European and Chinese history in the 4th century AD. The Buddhist monk Fotudeng spoke a few words of oracular advice in Hunnish to the north-eastern Hun monarch in 329. The monk's words were recorded in Chinese script:

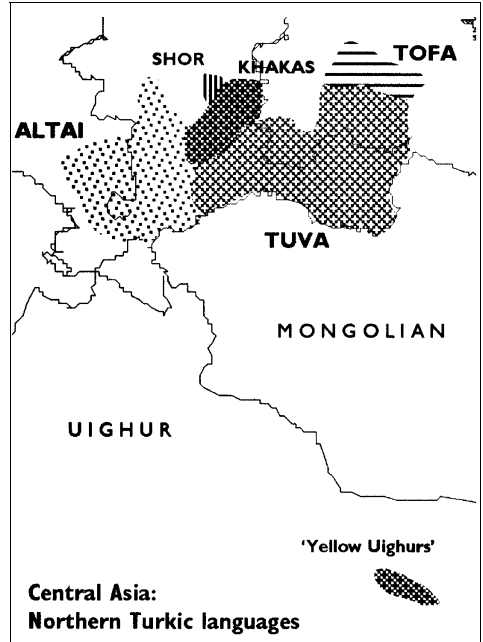
秀支替戾岡、僕谷劬禿當

On Fotudeng's historic advice the monarch defeated a rival warlord and established himself as Chinese emperor, founder of the Later Zhou dynasty. But what was the advice? In 4th-century Chinese these characters were pronounced: Syog tieg t'iei liäd käng b'uok kuk g'iw t'uk täng. (See B. Karlgren, *Grammata Serica*. Stockholm, 1940.)

Louis Bazin was able to interpret this puzzling text as a rhyming couplet in an early form of Turkic, several hundred years older than any other now known:

Süg Tägti idqang,	Send out your army,
boqughigh tutqang.	Capture the warlord!

After L. Bazin, 'Un texte proto-Turc du I^{er} siècle: le distique Hiong-nou du *Tsin-chou*' in *Oriens* vol. 1 (1948) pp. 208–19



Cyrillic for Russian and Cyrillic for Turkic languages

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 ш щ ъ ы ь 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 s h s h c h " ' e y u y a

To the letters of the standard Cyrillic alphabet, shown above in its Russian form, the different Turkic languages make a wide variety of additions:

Ӓ ӓ	ӓ Chuvash
Гъ гъ	gh Kumyk, Karachai, Balkar
Гь гь	h Kumyk
Ғ ғ	gh Khakas, Bashkir, Kazakh, Karakalpak, Uighur
Ӣ ӣ	gh Yakut
ДЖ дж	j Balkar
Дь дь	j Yakut
Ӗ ӗ	ə Chuvash
Жж	j Tatar, Uighur, Turkmen
Ӣ ӣ	ǯ Bashkir
Ӣ ӣ	ə Khakas, Kazakh
Ј ј	j Altai
Къ къ	q Kumyk, Karachai, Balkar
Қ қ	q Kazakh, Karakalpak, Uighur
Җ җ	q Bashkir
НГ нг	ng Kumyk, Balkar, Karakalpak
Нь нь	ng Khakas, Karachai, Nogai
НӢ нӣ	ny Yakut
Нң	ng Yakut, Tuva, Tatar, Bashkir, Kazakh, Kyrgyz, Uighur, Turkmen
Нн	ng Altai
Обь обь	ö Kumyk, Nogai
Ӗ ӗ	ö Khakas, Altai
Ө ө	ö Yakut, Tuva, Tatar, Bashkir, Kazakh, Karakalpak, Kyrgyz, Uighur, Turkmen
Ҫ ҫ	sy Chuvash, th Bashkir
Ӧ ӧ	w Karachai
Уь уь	ü Kumyk, Nogai
Ӱ ӱ	ü Chuvash, Khakas, Altai
Ү ү	ü Yakut, Tuva, Tatar, Bashkir, Kazakh, Kyrgyz, Uighur, Turkmen
Үү	ö Kazakh
Һ һ	h Yakut, Tatar, Bashkir, Kazakh, Uighur
Ң ң	j Khakas
Ә ә	ä Tatar, Bashkir, Kazakh, Uighur, Turkmen

Based on Nicholas Poppe, *Introduction to Altaic linguistics* (Wiesbaden: Harrassowitz, 1965) pp. 53–5

TURKISH

PERHAPS 50,000,000 SPEAKERS

Turkey, Northern Cyprus

One of the TURKIC LANGUAGES, national language of Turkey and of Northern Cyprus, Turkish is also spoken by shrinking communities in other countries that were once part of the Ottoman Empire.

Turkish is most closely related to Azeri and Turkmen (for a table of numerals see AZERI). All three languages were established as the result of a series of conquests and migrations of nomadic Turkic peoples, principally the Oghuz, from northern central Asia, beginning in the 6th century and largely completed by the 13th. The main languages of what is now Turkey had until then been Greek, Armenian and Kurdish. The long political and cultural dominance of Turkish, established by the 13th century, meant that it gradually became the language of the vast majority in the country; for all that, the other three languages were still spoken by millions at the beginning of the 20th century. Massacres and mass migrations have now transformed the picture: only the Kurdish minority is still numerically significant.

The name *Turkish* is used both for the language of Turkey itself (sometimes called *Osmanli Turkish*) and for the whole group (for which *Turkic* is preferred in this book). *Osmanli* is the name of the Ottoman dynasty established by 'Osmān I, who died in 1324.

The first Anatolian dynasty of Turkic origin was that of the Seljuks. They had previously ruled in Iran and absorbed the Islamic religion and Persian culture – which were inherited from them by the Turkish-speaking Ottomans, who succeeded to power in the 13th century. Thus Turkish is more pervasively influenced by Per-

sian (and through Persian by Arabic) than are the other Turkic languages. The language of older poetry can be so heavily laden with Arabic and Persian that it may not appear at first sight to be Turkish at all. The genres and the rhetoric of this literature are all Persian-derived.

Older in origin than this rich classical Turkish literature of the Ottoman period, the traditional tales of *Dede Korkut* are known from two manuscripts of the 16th century in a dialect close to Turkish or Azeri. Related tales in other Turkic languages have been collected by scholars of oral literature. Although superficially set in Anatolia at the time of the Turkish conquests, many features of the manuscript tales belong to the pre-Islamic legendary history of the Oghuz. They are 'in prose interspersed with rhythmic, alliterative, and assonant or rhyming passages of *soylama*, "declamation". The level of the language fluctuates, now highly poetic and dignified, now racy and colloquial' (G. Lewis, *The book of Dede Korkut* (Harmondsworth, 1974) p. 14).

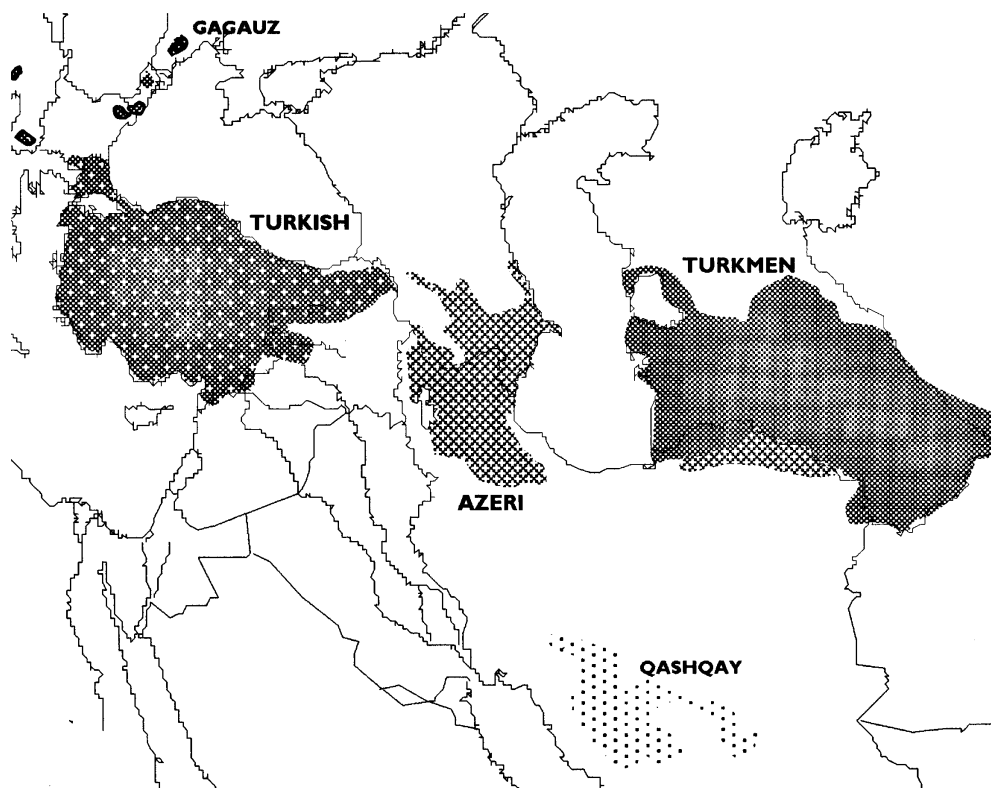
In the 19th century Turkish literature began to look to France for its inspiration. The Revolution of 1908 and the reforms of Atatürk (see box) furthered the change to a secular, western-influenced, nationalist culture. Research on the early history of Turkic peoples and their languages led to attempts to 'purify' Turkish, eliminating Ottoman words borrowed from Arabic and Persian and replacing them with rediscovered ancient Turkic words. Some of these had actually survived in peripheral dialects, such as Cypriot, but the main effect of the reform was to put a distance between the new written language and the speech of every day.

The geography of Turkish

Turkey is by no means homogeneous ethnically or linguistically, but the extreme nationalism of Turkish political life makes research on minorities difficult. The Yörük and Türkmen of western and central Anatolia are traditionally nomadic pastoralists. The Tahtacı are foresters, the Abdal are musicians and tinkers. *Abdal düğünden, çocuk oyundan usanmaz* is the proverb: 'A child never tires of playing, nor an Abdal of weddings.' All four groups speak dialects of Turkish, though the Abdal also have a secret language of their own. The Alevi (Kızılbaş) religion of the Türkmen, with their tribe and clan organisation and their traditions of migration from central Asia long ago, distinguishes them from the sedentary populations of Anatolia who have gradually adopted their language.

Until the early years of this century there was a large population of Christian Turkish speakers of Armenian origin in parts of Asia Minor. Their literature, in a form of Turkish with Armenian loanwords, was written and printed in the Armenian alphabet. Armenian oral poets of this region were often skilled both in Armenian and in Turkish. This population was largely killed in the genocide of 1913. There were also Turkish-speaking Christians of Greek origin, *Karamanlı-lar*, and thus a Turkish literature in the Greek alphabet. Most of these were moved to Greece in the population exchange of 1924, and their special form of Turkish is now almost extinct.

Also just outside the borders of 20th-century Turkey, a significant Turkish minority population (120,000 in 1974) had long been settled in Cyprus, where the majority was Greek. Since 1974 there has been an enforced division of the



island on religious and linguistic lines. In Turkish Northern Cyprus the government has attempted to impose standard Turkish as the national language: this has met with some resistance from the inhabitants, who are, however, now almost outnumbered by new immigrants from Turkey.

Over a hundred thousand 'Meskhetian' Turks, from the Turkish border region of Georgia, were exiled to Uzbekistan during the Second World War, apparently owing to Stalin's personal distrust of Caucasian Muslims. Their numbers have doubled in exile and their perceived prosperity has excited ethnic violence in which many have died. They remain unwelcome in independent Georgia, but some have resettled near their former home, in Azerbaijan, in lands from which LEZGHIANs have been uprooted.

Under the Ottoman Empire, Turkish, the language of government and culture, spread widely in south-eastern Europe, particularly among town-dwellers. In the 20th century as many as 1,100,000 'Rumelian' Turkish speakers have been expelled from Greece, Bulgaria and the former Yugoslavia, most of these fleeing to Turkey. A minority still remains in Europe.

Gagauz is the name of the Turkic language – with many Slavonic words in its vocabulary – spoken by communities of Greek Orthodox Christians of Turkic descent who live in several Balkan countries. The largest group, numbering well over a hundred thousand, is in south-western Moldova, where they recently claimed the position of a self-governing republic. Others live near Varna in Bulgaria.

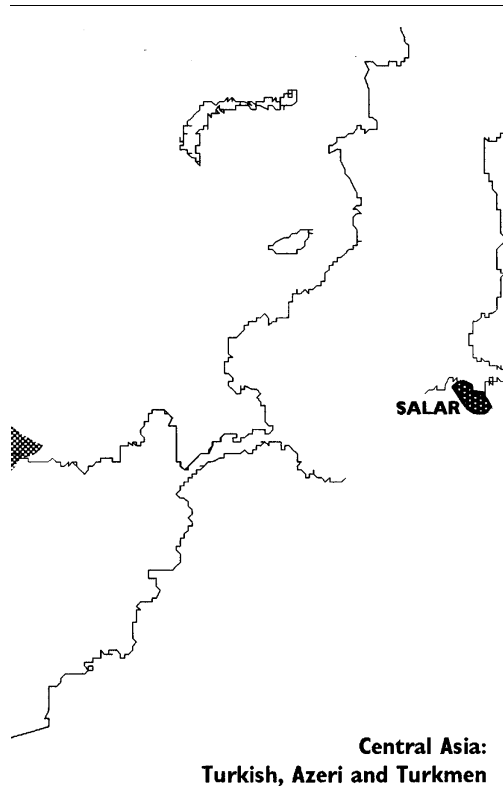
There are some more or less stable Turkish-speaking minority populations elsewhere, including perhaps 90,000 Turkish Cypriots in Britain. They cluster in the Green Lanes and Islington districts of London, and are large enough to be a political force, regularly canvassed for support for the shaky international status of Northern Cyprus. In Germany the number of Turkish speakers is much greater but many are short-term migrant workers.

Türkiyede altmışaltı buçuk millet var: 'There are sixty-six and a half nations in Turkey', some Turks say. The half-nation is the Gypsies (a half more than the status allowed them in many clichés elsewhere).

A Turkish proverb claims to characterise the two minorities seen as most dangerous to the comfortable order of things, *Çingene çalar, Kürt oynar*: 'If the Gypsy plays, the Kurd will dance.' The *Mıtrp*, a Gypsy caste of musicians, actually do live in close contact with the Kurds of Turkey and Iraq and are bilingual in Kurdish.

Examples from Peter Alford
Andrews and others,
Ethnic

groups in the Republic of Turkey
(Wiesbaden: Steiner, 1989)



Vowel harmony

A highly distinctive feature of practically all the Turkic languages is vowel harmony. In Turkish, for example, vowels can be classified according to three features:

	Unrounded		Rounded	
	High	Low	High	Low
Back	ɪ	a	u	o
Front	i	e	ü	ö

- 1. If the first vowel of a word is a back vowel, the others will also be back vowels.
- 2. If the first vowel is a front vowel, the others will also be front vowels.
- 3. Unrounded vowels are followed by unrounded.
- 4. Rounded vowels are followed by low unrounded or high rounded.

Frequently the principle is applied even to loan-words. The football term ‘penalty’ appears in Turkish not as *penalti* but as *penaltı*, the front *i* changing to back *ı* under the influence of the preceding back *a*. The French *épaulette* becomes *apolet*. The standard Turkish for ‘bus’ is *otobüs*, a phonetic spelling of the French *auto-bus*, but in uneducated speech the pronunciations *otobus* and even *otobos* are heard, as the front *ü* does not come naturally after two back *os*. Similarly, the French *vapeur*, ‘steamship’, has become *vapur*, though the Turkish phonetic spelling would be *vapör*.

G. L. Lewis, *Teach yourself Turkish* (London: English Universities Press, 1953) p. 20

The Turkish alphabet

By a 1928 decree of Mustafa Kemal Atatürk, Turkish exchanged its old Arabic script for a new Latin alphabet of 29 letters.

A B C Ç D E F G Ğ H I İ J K L
M N O Ö P R S Ş T U Ü V Y Z
a b c ç d e f g ğ h i j k l
m n o ö p r s ş t u ü v y z

Turkmen, Azeri and Turkish

Owing to continuing migration, speakers of these three very similar languages are inter-mixed.

TURKMEN is the national language of Turkmenistan.

AZERI is spoken in northern Iran and in the independent republic of Azerbaijan. There are also minorities in Georgia and Turkey. Azeri is an important lingua franca in the Russian republic of Dagestan.

Qāshqāy has perhaps 100,000 speakers in the Persian province of Fars.

Turkish, the national language of Turkey and of Northern Cyprus, is spoken by significant minorities in Bulgaria, Macedonia, Serbia and Bosnia.

Gagauz has over 100,000 speakers in south-western Moldova.

TURKMEN

4,000,000 SPEAKERS

Turkmenistan, Uzbekistan, Iran, Iraq, Afghanistan

One of the TURKIC LANGUAGES, Turkmen is the national language of the formerly Soviet republic of Turkmenistan.

The majority language of the region was probably formerly Iranian. Turkic speakers began to migrate here from the north in the 6th to 10th centuries. Oghuz Turks, from Mongolia, may have formed the majority, but by the 11th century the new name *Turkmen* was being used by Arabic authors. Islam became the majority religion here in the 11th century. The continuing movement south-westwards of Oghuz and Seljuk Turks resulted eventually in the settlement not only of modern Turkmenistan but also, much further to the west, of Azerbaijan and Turkey, whose languages are closely related to

Turkmen (see map at TURKISH and table of numerals at AZERI).

Political power in the region was long disputed between the Golden Horde, established to the north, and Persia to the south. Russia conquered Turkmenistan in the 1880s.

Until the 20th century Turkmen was not a written language, but there is a corpus of Turkmen oral epic poetry, including versions of the *Story of Dede Korkut* also known from a Turkish manuscript. Other stories dealt with the medieval wars of the Turks as reflected in popular tradition. In Soviet times Turkmen was written successively in Arabic, Latin and Cyrillic scripts. A new Latin alphabet, based on that used for Turkish, has been adopted since independence in 1991.

Bardı ärän qonuq bulup qutqá saqár,
qaldı **alğ oyuk** körüp ävní yqár.

Gone are the fine men who think
it lucky to welcome a guest;
Still around are the bad ones who strike
camp when they see you in the distance.

Two pessimistic lines mark the traditional beginning of Turkmen literature. Taken from the collected poems of the 11th-century author Maǧm ūd Kāšǧarī, they include words specific to the dialect that would later split into Turkmen, Azeri and Turkish.

After Johannes Benzing, 'Die türkménische Literatur' in *Philologiae turcicae fundamenta* vol. 2 ed. Pertev Naili Boratav and others (Wiesbaden: Steiner, 1965) p. 724

UDMURT

550,000 SPEAKERS

Russia

One of the URALIC LANGUAGES, Udmurt is spoken in Udmurtia, in eastern European Russia, south of the related KOMI, in the valleys of the rivers Vyatka and Kama (see map at KOMI).

Udmurt is the speakers' own name for themselves and their language, and is officially used in Russian too. By most other outsiders they are called *Votyak* (German *Wotjak*).

Of all the Uralic languages Udmurt may be the one that has migrated least. The Bronze Age Turbino culture (second millennium BC) and the Iron Age Ananino culture (first millennium BC), as archaeologists call them, may mark the location of a proto-Permian language, ancestor of Udmurt and Komi. Both centre on the valley of the Kama, a middle Volga tributary. Udmurt speakers simply stayed put, sometimes dominated by external powers such as the Tatars and the Russians.

The language was first recorded in the 18th century. Udmurt religious books multiplied in the 19th, though Christian Russia never entirely succeeded in eradicating the pagan religion of the Udmurt. However, the area over which Udmurt is spoken has gradually shrunk under the pressure of Russian expansion and Russian-led development.

'The Votyaks have attracted the attention of ethnologists because of their extensive sexual promiscuity' is probably the most interesting sentence in Björn Collinder's *An introduction to the Uralic languages* (Berkeley: University of California Press, 1965).

A nation of poets

'Folklore collectors remarked that every Votyak was able to improvise. There were occasions that simply obliged a Votyak man or woman to improvise a song: a young man joining the army, a young woman going to marry, a housewife receiving guests or seeing them out of her home, all alike practised the art of improvisation.'

Péter Domokos, 'Finno-Ugrian folk poetry' in
Ancient cultures of the Uralian peoples
(Budapest: Corvina, 1976) p. 295

The first Udmurt textbook

Odyk, kyk, kvin', nil', vit',
kuat', sizim, kiamys, ukmys, das

The first ten numerals in Udmurt – from the above double-page spread in *Sochineniya prinadlezhashcheniya k grammatike Votskago* (Writings on Udmurt grammar), published in St Petersburg in 1775 and probably compiled under the supervision of Archbishop Venyamin Putsek-Grigorovich of Kazan', an accomplished linguist. In this textbook Udmurt is in the left column, Russian in the right, both in Cyrillic script.

UIGHUR

6,750,000 SPEAKERS

China

One of the **TURKIC LANGUAGES**, Uighur is spoken by one of the most important national minorities of China, occupying the Xinjiang ('New Frontier') Autonomous Region. There is a minority of Uighur speakers in three former Soviet republics of central Asia.

The Uighur Empire was established by this nomadic Turkic people in western Mongolia in 745 AD, in succession to the earliest Türk empire whose history is recorded in the Orkhon runic inscriptions (see box at **TURKIC LANGUAGES**). In 762 it adopted the syncretistic religion of the prophet Mani, brought by **SOGDIAN** missionaries, and thus became the world's only known Manichaean state.

The Uighurs were driven from power by the Kyrgyz (whose territory then lay to their north) and fled to the lands that lie on either side of the Silk Road as it crosses the heights of central Asia on its way from China to the West. Here speakers of Uighur, and its offshoot **UZBEK**, settled and prospered, as they ruled the valleys and oases through which the inland trade in silk and other luxuries had to pass. Buddhism and Nestorian Christianity competed with Manichaeism; eventually the Uighurs were converted to Sunni Islam in the 11th century.

In 1755 the Uighurs came under Chinese rule. They were the major participants in the 'Mohammedan uprising' in western China in 1862–77. In spite of further rebellions, leading to Soviet interest in the region in the 1930s, it remained part of China. As recently as the 1950s Uighur adopted numerous Russian loanwords; since then, naturally enough, Chinese has been the major influence.

Earlier forms of the language of the Uighurs,

mostly written in a script similar to that of Mongolian, are conventionally called Ancient Turkic (9th and 10th centuries), Karakhanid (11th to 13th centuries) and post-Karakhanid (13th and 14th centuries). Buddhist texts survive from the earlier period, translated from Sanskrit via Tocharian or Sogdian; there are also Manichaean and Nestorian texts, and translations from Chinese such as an adaptation of the fortune-telling *I Ching*.

Uighur script

The traditional Uighur form of writing was based on an early medieval Sogdian script, ultimately derived from Aramaic. Its transmission to central Asia followed the route of the Silk Road – and accompanied the texts and teachings of Manichaeism. In turn, the Mongolians borrowed Uighur script and applied it to their own language. Uighur script differed from Sogdian in one essential way: like Chinese it is written from top to bottom, not from right to left.

Uighur is now normally written in Arabic script. In the 1950s and 1960s the Chinese government attempted to effect a change to the Latin alphabet. For a while this was taught in all schools. It now appears to have been abandoned, and Arabic script is used again in the local press.

Uighur and Uzbek

First recorded from western Mongolia, Uighur has for many centuries been the major language of the central Asian region now known as Xinjiang (Sinkiang) in China. The modern literary

language is based on the dialect of the regional capital, Urumchi. Uighur dialects of Xinjiang can be divided into three groups: Central, Lobnor and Khotan.

The Uighur spoken by about 200,000 people in Kazakhstan, Kyrgyzstan and Uzbekistan is written in Cyrillic script, and is sometimes classed as the *Ilī* dialect.

In Soviet times standard written Uzbek was at first based on the dialect of the city of Turkestan – in Kazakhstan, north of the Uzbek border. This had undergone relatively little Persian influence. Now, however, the more cosmopolitan dialect of Tashkent is paramount.

Numerals in Uighur and Uzbek

Uighur		Uzbek
bir	1	bir
ikki	2	ikki
üch	3	uch
töt	4	türt
bäsh	5	bish
altä	6	olti
yättä	7	yetti
säkkiz	8	sakkiz
toqquz	9	tûqqiz
on	10	ün

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

He knew all the languages and scripts

‘A Persian historian recorded a legend of the origin of the Uighur empire. At the confluence of two rivers, Selenge and Tughla, stood a mound. A light which descended from heaven made this mound pregnant. Day by day the mound grew bigger: finally, “as with pregnant women at the time of their delivery”, a door opened and inside were five cells, “like tents”, in each of which sat a baby boy. The Uighurs chose as their leader, from among the five boys, Buku Khan, for he was superior to the others in beauty, strength of mind and judgment – and he knew all the tongues and writings of the different peoples.’

From ‘Ala al-Din ‘Ata-Malik Juvaini, *The history of the world conqueror* translated from the text of Mirza Muhammad Qazvini by J.A. Boyle (Manchester: Manchester University Press, 1958) vol. 1 pp. 55–6

UKRAINIAN

45,000,000 SPEAKERS

Ukraine

One of the Eastern SLAVONIC LANGUAGES, Ukrainian has a longer history than Russian. Formerly the second largest language of the Soviet Union, it is now the national language of a major independent state in eastern Europe, with its capital at Kiev (for map see RUSSIAN).

In Russian *Ukraina* means 'borderland' and referred to the eastern Ukraine, settled by the Cossack borderers. Until the 20th century Ukrainian speakers generally called themselves *Rusyny*, meaning 'Russians' (the translation often used abroad was 'Ruthenians'). Historically Kiev was indeed the centre of the land of *Rus'*.

Ukrainian is probably spoken on the territory where, in prehistoric times, the proto-Slavonic dialects first developed and began to spread. Its recorded history begins in the century after the official conversion of the principality of Kiev to Christianity in 988.

The Bible translations and religious texts that Byzantine missionaries brought to Kiev were in OLD SLAVONIC, the written language that had already been used to take the Christian religion to other Slavonic-speaking peoples. Old Slavonic continued to be the single significant written language of Ukraine for several centuries: it is the language of the *Russian Primary Chronicle* and other major early texts. However, Ukrainian is to be seen in many of these texts. It is there both in the 'mistakes' made by religious writers while writing or copying Old Slavonic, and in the vernacular words that were necessarily introduced into chronicles and other documents when discussing local matters for which no Biblical word could be found. At this stage 'Old

Ukrainian' is not to be distinguished from *Old Russian*, which is the name often given to the vernacular element in these manuscripts.

Mongolian and Tatar warriors of the Golden Horde destroyed Kiev in 1240. Moscow, too, was to be subject to the Horde, while most of Ukraine passed to the Lithuanians (whose written language at this time was an early form of BELO-RUSSIAN). From this point onwards the histories of Ukrainian and Russian diverge.

Poland, already ruler of Galicia, annexed the remainder of Ukraine in the 16th century. Latin (the old official language of Poland) and Polish itself now had an official role in Ukraine. The deciding factor in the country's subsequent history was the arrival of the Cossacks, armed adventurers and vagabonds of miscellaneous origins who settled the empty lands of eastern Ukraine. Their 18th-century leader, Bohdan Khmelnytsky, invited Russian help against Polish domination. Under Catherine the Great the Cossacks eventually moved on, to settle the Russian frontiers further east, but Ukraine remained in Russian hands (except for its westernmost region, eastern Galicia, which belonged to Austria-Hungary).

In the 17th century the Belorussian language of former Lithuanian rule had continued to be used, even by the Cossacks, and had gradually adopted Ukrainian features: it was called *prostaya mova*, 'common tongue'. Under Russian rule, however, Ukrainian came to be considered no more than a local dialect of Russian (often called *Little Russian*). While authors such as Tarash Shevchenko and Pan'ko Kulish, inspired by European Romanticism, were developing a language and a literature based on popular speech

and folklore, the cultivation of Ukrainian was officially frowned on. An 1876 decree forbade the printing or importing of Ukrainian books. The 19th-century revival of Ukrainian poetry and historiography was thus conducted partly underground, or from neighbouring Galicia where Ukrainian nationalism was able to flourish. The Galician dialect thus became a major influence on standard Ukrainian.

Briefly independent in 1918–19, Ukraine was reconquered and declared a Soviet Republic. Its western and southern provinces were divided among Poland, Czechoslovakia and Romania – Galicia, formerly Austrian, going to Poland – but were repossessed in 1944. The old linguistic attitudes were not wholly abandoned under Soviet rule. Russian had priority in education and employment. Ukrainian began to lose its status as the official language of Soviet Ukraine. Many Ukrainians speak Russian fluently: there is also a very large Russian minority in eastern Ukraine and in the major cities.

To the surprise of many Russians, Ukraine asserted its independence in 1991. Its traditional links with central Europe, broken under Soviet rule, are now being cultivated once more. Priority is given to relations with the United States and Canada, where there are strong Ukrainian minorities. Canada numbers 600,000 of a total of

four million Ukrainian émigrés worldwide. The descendants of the million Ukrainians who migrated to central Asia and Siberia under Soviet rule are, in many cases, now returning.

In many ways still close to Russian, Ukrainian is a language with its own character – a language in which one can speak of Сокира така тупа що як за мамою кинути би не був гріх, ‘an axe so dull that it would not be a sin to throw it at your mother’.

In Britain there are about 20,000 Ukrainians. Most of them, young men arriving at the end of the Second World War, married women of other immigrant communities with strong religious beliefs (especially Italian, Austrian and Irish). The Ukrainian community clusters in Lancashire, Yorkshire and the Midlands. While modern Ukrainian of the Ukraine has been increasingly influenced by Russian, the Ukrainian of the emigrants and their children incorporates more and more English words and English turns of phrase.

Ukrainian is written in the Cyrillic alphabet. It can be distinguished at a glance from Russian (see table there) by its use of three additional vowel symbols, Є є for *ye*, І і for *i*, and Ї ї for *yî*. An additional consonant, Ґ ґ for *g*, also occurs. For a table of Ukrainian numerals see BELORUSSIAN.

URALIC LANGUAGES

A family of languages of northern Europe and western Siberia. There are in total only 24,000,000 speakers of Uralic languages, but the family includes three national languages, ESTONIAN, FINNISH and HUNGARIAN, as well as several important minorities of European Russia (KOMI, MARI, MORDVIN, UDMURT) and the language of the SAMI or Lapps of northern Scandinavia.

The first written records of Uralic languages date back only to the 13th century. Their earlier history can be traced through notes in historical writings of other nations (especially Romans, Greeks and Chinese) and before that by linguistic reconstruction, which includes the tracing of loanwords between Uralic and other languages. Alongside this goes archaeological exploration of the prehistoric cultures of north-eastern Europe.

Proto-Uralic dialects ancestral to all the modern languages may have been spoken on both sides of the central and northern Ural Mountains – the traditional dividing line between Europe and Asia – in the sixth millennium BC or even before. From this linguistic community the SAMOYEDIC LANGUAGES must have separated first, for these are most different from the rest. Their speakers have perhaps always been foragers and hunters in the forests and tundra of Siberia.

Before the Ugric dialects (with later Hungarian) had separated from the Finno-Permian dialects (with later Finnish), proto-Finno-Ugric

borrowed some significant words from proto-Indo-Iranian: the two early languages must, at this time, have been close enough for regular contact. Examples include proto-Indo-Iranian *septa* ‘seven’, modern Hungarian *hét*; *shata* ‘hundred’, Finnish *sata*, Hungarian *száz*; *sharva* ‘horn’, Finnish *sarvi*, Hungarian *szarv*; *orbho* ‘orphan’, Finnish *orpo*, Hungarian *árva*.

The Ural Mountains themselves may have formed the dividing line between the proto-Ugric and proto-Finno-Permian dialects, a division which perhaps became established in the third millennium BC. Ugric languages now consist of Khanty and Mansi, still spoken east of the Urals, and Hungarian, whose speakers finally settled in central Europe after centuries of migration.

The Finno-Permian peoples, perhaps always settled on the European side of the Urals, may be tentatively identified with a series of Neolithic, Bronze Age and Iron Age cultures of the upper Volga and its tributary the Kama – and soon spreading to the neighbourhood of the eastern shores of the Baltic. Linguistically, the first to become separated were the most eastern group, the Permian dialects that were to become modern Udmurt and Komi. The remainder (Mari, Mordvin, Sami and the proto-Finnic group that gave rise to modern Estonian and Finnish) began to separate into distinct languages in the course of the first millennium BC.

Proto-Uralic trees

Tree names shared among the modern Uralic languages help to locate the habitat of the speakers of proto-Uralic about eight thousand years ago, and of proto-Finno-Ugric in the next two millennia.

<i>Picea obovata</i> , spruce	proto-Uralic <i>kowese</i>	Finnish <i>kuusi</i>
<i>Pinus sibirica</i> , cembra pine	proto-Uralic <i>sikse</i>	Komi <i>sus-</i>
<i>Abies sibirica</i> , Siberian fir	proto-Uralic <i>nyulka</i>	Mari <i>nulgo</i>
<i>Betula spp.</i> , birch	proto-Uralic <i>kojwa</i>	Finnish <i>koivu</i>
<i>Populus spp.</i> , poplar	proto-Uralic <i>poje</i>	Mordvin <i>poj</i>
<i>Salix spp.</i> , willow	proto-Uralic <i>paje</i>	Hungarian <i>fagyal</i>
<i>Pinus silvestris</i> , fir	proto-Uralic <i>juwe</i> ;	Mansi <i>jiw</i>
	proto-Finno-Ugric <i>penye</i>	Hungarian <i>fenyo</i>

<i>Larix sibirica</i> , larch	proto-Finno-Ugric <i>nyänge</i>	Komi <i>nyia</i>
<i>Ulmus spp.</i> , elm	proto-Finno-Ugric <i>syala</i>	Hungarian <i>szil</i>

Three trees are crucial. Cembra pine and Siberian fir were slowly spreading across the Urals westwards in proto-Uralic times; elm was meanwhile spreading eastwards from central Europe, and reached the north-western Ural foothills. The region where proto-Uralic dialects were spoken most probably included the relatively narrow zone, west of the Ural watershed, where the two trees met.

URDU

40,000,000 SPEAKERS

India, Pakistan

One of the INDO-ARYAN LANGUAGES, Urdu is the twin of HINDI. It has the same origin in the regional language of the country around Delhi. Culturally, the two languages are a world apart.

Urdū is in full *Zabān-i-urdū*, 'language of the camp' – a Persian phrase that incorporates the Turkish word *ordu*. Thus 'Urdu' is the same in origin as *horde*, see box at UZBEK. The spoken Urdu of the 19th century, one of the major languages of British India, was then called *Hindustani* – the lingua franca of the subcontinent whose Persian name is *Hindūstān*, 'country of the Hindus'.

Urdu, in origin the speech of the northern Indian Muslim courts and cities, spread as a lingua franca in India wherever Mughal influence was felt. Evidently well known and serviceable in both north and south, it was much favoured during the early expansion of British rule.

The earliest Urdu poetry, of the 16th and 17th centuries, comes from the Muslim courts of the south of India, particularly Hyderabad. At the beginning of the 19th century, the British Fort William College encouraged the development of a new literary standard on the basis of the Urdu of Delhi, a literary language intended to supplant the Braj form of Hindi. But this policy overlooked the fact that Islam was a minority religion in India, and Urdu vocabulary and style, under Persian influence, had drifted away from its popular base. Its script, too, was not ideal for an Indo-Aryan language and unsuitable for typesetting.

When, at independence, India split on reli-

gious lines, Urdu, which had the best-developed cultural tradition among languages of Indian Muslims, took its place as the sole official language of Pakistan. It is thus widely spoken there as a second language, but it is the mother tongue of only a minority, numbering about 8,000,000 – and many of these are emigrants or children of emigrants from North India. Their main centre is Karachi. Urdu functions as the literary language of the numerically dominant PANJABI and Lahnda speakers of Pakistan.

In India, *Dakhinī Urdū* (Urdu 'of the Deccan, of the south') still centres on Hyderabad, Bijapur, Gulbarga and other mainly Muslim towns of the Deccan plateau. Urdu is still widely spoken in the big northern Indian cities. The total number of speakers in India may be as many as 32,000,000 – a figure is difficult to give, since they may well be competent in Urdu and Hindi equally.

The history of Urdu as a lingua franca lives on in its use in pidginised form as a trading language in great cities such as Calcutta and Bombay – both of which owe their early growth to their status as centres of British rule. Naturally influenced by the majority languages (Bengali and Marathi respectively), *Bazār Hindustānī* in these cities serves for communication among those who do not themselves speak Bengali or Marathi.

Outside the subcontinent, Urdu is the cultural language of many emigrant communities of Indian Muslims, particularly Panjabi and Gujarati speakers.

Urdu is written in a script based on Arabic, with added letters as used for Persian, and some variants specific to Urdu. It is usually printed from calligraphy. Its typical sloping style cannot be satisfactorily imitated with movable type. One or two specially designed word-processing programs are now able to generate good written Urdu.

Numerals in Hindi and Urdu		
	Hindi	Urdu
1	ek	اِک
2	do	دو
3	tin	تین
4	cār	چار
5	pāñc	پانچ
6	cha	छ
7	sāt	सात
8	āṭh	आठ
9	nau	नौ
10	das	दस

Urdu in print

بیابان عاشقان کو ملک اسکندر برابر ہی
 ہر اک گوہر انچھو کا بخت کی اختر برابر ہی

To a lover, the desert equals Alexander's kingdom,
 And each pearl of a tear is a lucky star.

Lines 1–2 of a ghazal by Wali Dakhani (1668–1743): after John A. Haywood, 'Wali Dakhani and the development of Dakhini Urdu Sufi Poetry' in *Acta orientalia* vol. 28 (1964–5) pp. 153–74

UZBEK

16,000,000 SPEAKERS

Uzbekistan, Afghanistan

In terms of number of speakers, Uzbek is the second largest of the TURKIC LANGUAGES, after Turkish itself, and it was the third largest language of the old Soviet Union after Russian and Ukrainian. The Uzbek capital, Tashkent, is the greatest metropolis of central Asia (see map at UIGHUR).

Uzbak Khān (1312–40), of the Golden Horde, became a Muslim, and his name was applied to the Muslim Turks who owed allegiance to the Horde. Their descendants moved eastwards, towards central Asia, in the fifteenth century, and established khanates in Kokand, Khiva and Bokhara. The city-dwellers, their subjects, were partly PERSIAN-speaking Tajiks: others spoke a Turkic language, a variant of early Uighur. There has been centuries-long interchange and intermarriage between the two. The modern form of this Turkic language is called *Özbäk* (*Uzbek* in most foreign languages) after the far-ranging conqueror and his nomadic followers.

Chagatai was the older written language of the Uzbek khanates and the Golden Horde. It takes its name from one of the sons of Genghis Khan, who ruled in central Asia in the 13th century.

Used by Timur (ruled 1369–1407) and his descendants, notably Sultan Husain Baykara (ruled 1469–1506), it was at Husain's court that Chagatai emerged as a literary medium – a Turkic language competing for the very first time with Arabic and Persian – in the hands of the poet Alisher Navoi (1440–1501). Chagatai is now sometimes called 'Old Uzbek', though Uz-

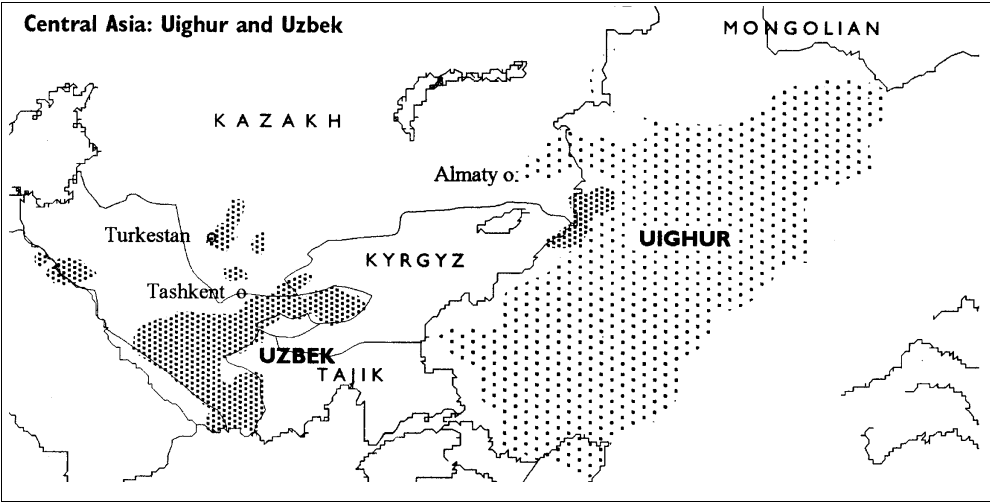
bek is only one of the several modern Turkic languages spoken where Chagatai once held sway.

The three Uzbek khanates were incorporated in the Russian Empire between 1865 and 1876. Russian influence on the language – alongside that of the Islamic missionary TATARS – now began to overlay that of Arabic and Persian. The 'local customary tongues' (essentially Uzbek and KYRGYZ) were declared equal with Russian in 1918, but until the 1920s Uzbek was scarcely identified or named as a separate language.

Its status was assured with the creation of the Uzbek Republic, within the Soviet Union, in 1924. There was a movement to 'cleanse' Uzbek of Arabic and Persian loanwords – but Russian loanwords multiplied meanwhile.

The first wholly Uzbek newspaper, *Turkiston Viloyatining Gazeti*, had appeared in 1893. Supplanting the Arabic script inherited from Chagatai writing, a Latin alphabet was used from 1927, to be replaced in turn by Cyrillic in 1940. In the last few years the Latin alphabet has been officially readopted in independent Uzbekistan, but its widespread use lies somewhere in the future. Everywhere, indeed, one sees the catch-phrase in Cyrillic script, Узбекистан – келажачи буюк давлам, 'Uzbekistan, a state with a great future'.

The region once ruled by the Uzbek is now divided between Uzbekistan, Afghanistan and Kazakhstan. There are at least a million Uzbek speakers in Afghanistan, where it is recognised as a national language and used in broadcasting. Numbers there grew as a result of emigration from Uzbekistan in Soviet times.



One of the main areas where Persian (Tajik) influence is felt is the tendency for the vowel system to shift from the typical ten-vowel system of most Turkic languages, with distinctive front-rounded and back-unrounded vowels (*ö ü, ə ɨ*), to a six-vowel system practically identical to that of Tajik. The process is not complete, and it has not occurred in all the dialects, but it is far

advanced in the city dialects of Tashkent and Samarkand.

For a table of numerals see **UIGHUR**.

The Golden Horde was not a band of nomadic warriors. It was the camp – Russian *orda*, Turkish *ordu* – which was their mobile capital. The language name **URDU** has the same origin.

A riddle

Ol ne dur kim şahddin tatlıq erür hammārğa,
İhtiyār etsām satarnı kirmägäy bāzārğa?

What is sweeter than honey to the eyes,
Yet much as I wanted to sell it
I could not get it to market?

The answer? Sleep.

Uvaysi (18th-century woman poet)

The Chagatai alphabet

ي و ه ن م ل گ ک ق ف غ ع ظ ط ض ص ش س ز ز ر ذ د ج خ ح ث ت ب ب ا
a b p t s j h kh č d z r z ž s š ž ž ' gh f q k g l m n h v y

In the Arabic script as used for Chagatai, the two semivowel signs (the last two letters of the alphabet) were also used for vowels: و *v* served for *o u ö ü*; ي *y* served for *i ı*.

VENDA

750,000 SPEAKERS

South Africa

Venda or *tshiVenda* is one of the BANTU LANGUAGES and one of the eleven official languages of South Africa. It is rather isolated linguistically, having no close links either with the Nguni group (Zulu and others) or the Sotho group or the Shona group of Bantu languages, which surround it (see map at SOTHO).

In the 18th century, groups from north of the Limpopo – the *Singo* lineage – are said to have crossed to the south and established their overlordship in the Venda country, where they remain powerful. This explains why Venda has quite numerous Shona loanwords. Unlike its neighbours, it has never undergone Zulu influence.

Venda has a special *Musanda* vocabulary which

must be used at court, avoiding taboo expressions: thus, if a chief is the subject of the sentence, one does not say – *vhulaha* ‘kill’ but instead – *ponda*. This is a Shona loanword: in the Zezuru dialect of Shona it means ‘execute, murder’.

In the 19th century Venda speakers attracted the attention of the Berlin Missionary Society. The written form of Venda (at first called *Ba-suetla*) was thus the creation of German-speaking missionaries. That resulted in some unexpected spelling conventions, such as *dzh* for the sound that in neighbouring languages is written *j*.

The first ten numerals in Venda are: *-thithi*, *-vhili*, *-raru*, *-na*, *-tanu*, *-tanu na* *-thili*, *-tanu na* *-vhili*, *-tanu na* *-raru*, *-tanu na* *-na*, *fumi*.

VIETNAMESE

55,000,000 SPEAKERS

Vietnam

Now generally agreed to be one of the AUSTRASIATIC LANGUAGES, Vietnamese is so different from its relatives that quite contradictory views on its origin have been expressed. It is the official language of Vietnam.

In recent centuries the cultural centre had been the old royal city of Hue, capital of Annam. The language has therefore sometimes been called *Annamese* or *Annamite*. Vietnam, the 'Viet country', is an inclusive term covering also the southern region of Cochinchina and the northern region of Tonkin. Vietnamese is the majority language in all three.

For over two thousand years Chinese culture has exerted a pervasive influence on that of Vietnam. According to traditional history, China ruled the kingdom of Annam, which included Tonkin, from the 2nd century BC until 968, when Dinh Bo Lanh established himself as an independent monarch. Over the following centuries Vietnamese rule extended southwards, where the kingdoms of Champa (see CHAM) and Funan had once held sway.

Roman Catholic missionaries were involved in Vietnam from the 17th century, in competition with Buddhism, Confucianism and the still-prominent survivals of local religion. Their great contribution to its culture was the *quoc ngu* script, a special version of the Latin alphabet suited to the sounds and especially the tones of Vietnamese. Until the early 20th century this coexisted with *nom*, the Vietnamese adaptation of Chinese script: but it was encouraged by the French administration, it promoted the spread of literacy and thus of new political ideas, and it is now universally used.

Of Vietnamese literature before the 19th century a great deal is 'Sino-Vietnamese' – not only dependent on Chinese models but written in more or less pure Chinese. This includes the history *Dai Viet su ky* 'Records of Great Viet', compiled in the 13th century and revised later. Here the succession of Vietnamese emperors is traced back to 2879 BC. Literature in true Vietnamese, still full of Chinese words but closer to its real time and place, is rich in poetry and legend. The greatest work of Vietnamese literature is *Kim Van Kieu*, the 'Tale of Kieu', a verse romance by Nguyen-Du (1765–1820).

Until the French conquest, which began in 1862, printing in Vietnam was by means of woodblocks. Western-style printing in *quoc ngu* was introduced at once and a newspaper press developed in the next decades. French, Japanese and American domination ended in thirty years of war, culminating in the reunification of independent Vietnam in 1975.

The standard dialect of Vietnamese, represented in the script, is that of Hanoi. This has six tones: thus the script, too, differentiates between six tones, though speakers in most of Vietnam recognise only five. As can be seen from the box, tones are essential in Vietnamese in distinguishing the meaning of otherwise identical words.

Chinese words now make up as much as 60 per cent of the vocabulary of written Vietnamese. Many of these loans arrived before the 10th century, as is evident when their Vietnamese pronunciation is compared with modern Chinese – though the logic is obscured by the fact that Chinese in Vietnam has, in any case, continued to be pronounced in the medieval

WA

PROBABLY WELL OVER 1,000,000 SPEAKERS

Burma, China

Wa consists of a group of related AUSTRALASIAN LANGUAGES, the main body of whose speakers live on the mountainous borderland between Burma and China. In Burma the region is part of Shan State, but has always been difficult country for the Burmese authorities and a traditional centre of Communist insurgency. Speakers in China belong to the Va and Bulang national minorities.

The legendary centre of the Wa country is Lake Nawngkhio, high in the mountains on the China–Burma frontier. The older history of the language and its speakers is unknown, but Austroasiatic speech may well have a history of several millennia in this region. The population of the Wa country is high and cultivation of the mountain slopes is intense. Their most saleable product is opium, of which this is one of the world's main sources.

Facts on the Wa country, including its population, are still hard to assemble. Outsiders have not found it easy to get to grips with Wa political philosophy. The British, for example, who annexed the region in the 1890s, could not understand why the village rulers (often given the Shan title *Sawbwa*) would not identify themselves to strangers. Sometimes they disappeared; sometimes they were actually there among the villagers, but unidentifiable. This explains why the first entry in the Wa wordlist published by Lieutenant Daly in 1891 is 'Where is the Sawbwa?'

A further problem has been the prevalence of headhunting (see box). During a China–Burma border delineation in 1900 Wa warriors captured

two British heads in a daring raid near a Chinese market town. These two heads were still objects of worship sixty years later.

Apart from their mountain heartland, Wa speakers are to be found in scattered communities in mountainous parts of Xishuangbanna, of the southern Shan State (especially the former state of Kengtung) and of north-western Thailand. For those who have adopted Buddhism and valley agriculture, Shan and Lanna Thai become the adopted languages: it is likely enough that Wa was once the main language of Kengtung and other neighbouring states before the spread of Tai speech.

Wa is the Shan name for this people and is the most widely familiar. *Va* is the Chinese form. *Lawa* is the usual name for the Wa languages of Thailand and their speakers.

The term *La* is also used, specifically for Wa speakers who have adopted Buddhism and other cultural features associated with Tai populations: those who had not were traditionally called *Wa hai* by Shans, *Wild Wa* in English.

The first serious notes on Wa languages were made by French and British explorers in the early 1890s. Wa languages are largely monosyllabic, like their Tai and Tibeto-Burman neighbours, but unlike some other Austroasiatic languages. They have no tones, but they tend to have rich and complicated vowel systems: Chinese researchers have counted fifty 'vocalic nuclei' (vowels and diphthongs) in the Paraok dialect. In the numerals table, the accents ´ and ` make the distinction between 'tense' and 'lax' vowels, one that is

paralleled (though with different details and different names) in Mon, Khmer and other Austroasiatic languages.

Wa shows the influence of neighbouring Tai languages, in which some speakers are bilingual. Loans from Shan are found in central Wa dialects such as Paraok; Lawa dialects have borrowed from Lanna Thai; Bulang dialects have borrowed from Lü, the Yunnanese variant of Lanna Thai.

There has been little writing or publishing in Wa languages, and no generally accepted orthography. But the American Baptist Mission Press, in Rangoon, brought out translations of the Gospels of Matthew and John in 1934–5, and some books in Wa have appeared in China at intervals since the 1950s.

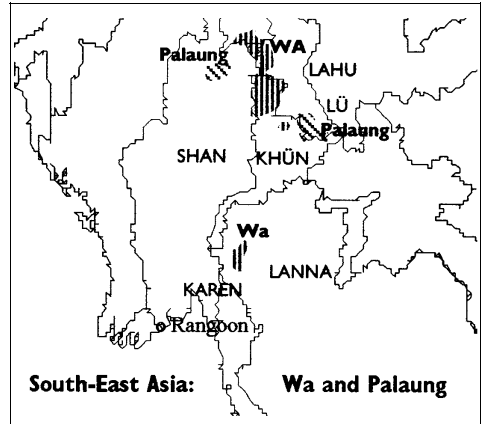
Exploring the Wa country

'The country is described as very difficult and almost pathless; it is never entered by caravans; fear of the Wa Hai and Muhsö Hai ['Wild LAHU'] is so deep and general that guides could not have been obtained . . . No Shans, except those living in the immediate vicinity and who supply the Was with salt and other necessities, would dare to pass through this region.'

H. Daly, *The Northern Trans-Salween States and the Chinese Border* (Rangoon, 1891)

'[We] passed another avenue of posts, about fifty of them, under big trees, about a third of which contained heads. Any number of small spirit shrines under archways of greenery. Then came upon evidence without question that the head-hunting was in full swing: a human body right across the path, beheaded and with the hands and feet cut off . . . Endless bother with guides. Had no sooner started than the guide refused to go beyond the upper villages. He solemnly unfastened a piece of rag, and taking out the two-anna bits that had been given him, handed them back, and then just legged it.'

Sir George Scott's Diary, February 1893 (India Office Library and Archives)



Wa and Palaung

Wa and Palaung form two groups that make up a single branch, 'Palaungic' or Palaung-Wa, of the Austroasiatic language family. Geographically it is centred in the Shan State and western Yunnan. Speakers of these languages are the modern representatives of a population that has an older history in the mountains of inner south-east Asia than the speakers of Tai or Sino-Tibetan languages.

The Wa or 'Waic' languages include *Central Wa* dialects, Paraok, Avüa', La and others; *Phalok*, formerly called Khalo or Mae Rim Lawa; *Lawa* and its dialects; and the dialects of *Bulang* (or Blang or Samtao), counted as a separate Chinese national minority.

The Palaung or Palaung-Rumai group, with perhaps 500,000 speakers in total, includes several languages of Burma and China. Speakers in China belong to the *De'ang* (formerly Bēnglóng) national minority. The major language of the group is *Ta-ang* or Palaung, spoken in the former 'Shan State' of Tawngpeng and by minorities in Hsipaw, Hsenwi and Möngmit states, all now part of Shan State, Burma. Minor Palaung languages include *Rumai*; *Riang* and *Yinchia* (also called Black Karen, Striped Karen); *Palê*, including Da-ang and Na-ang; *Ka-ang*, *Ra-ang*.

Numerals in Wa languages and Palaung				
	Kawa	Lawa	Bulang	Ta-ang (Palaung)
1	tìʻ	thiʻ	ktìʻ	ũ, hlĕh
2	rá	la	lǎʻál	ār
3	láy	laʻua	lǎʻýy	u-āī
4	pón	paun	pún	phōn
5	phúan	phon	phón	phən
6	liah	lēs	lĕh	tōr
7	ʻaliah	ʻa-lēs	harrĕh	pūr
8	ntáiʻ	sateʻ	sətíʻ	tā
9	ntím	sataim	sətím	thīm
10	káo	kau	kul	kör
G. Diffloth, 'The Wa languages' in <i>Linguistics of the Tibeto-Burman area</i> vol. 5 no. 2 (1980) pp. 1–182 and other sources				

WARAY-WARAY

2,400,000 SPEAKERS

Philippines

One of the Bisayan languages, like CEBUANO (see map there), and thus a member of the larger family of AUSTRONESIAN LANGUAGES, Waray-waray forms a dialect group including *Samar-Leyte*, *Northern Samar*, and *Gubat* or *Southern Sorsogon*. They are spoken in the islands of Samar and eastern Leyte and part of the Sorsogon district at the southern extremity of Luzon.

The first ten numerals in Waray-waray are: *'usa*, *duha*, *tulu*, *'upat*, *lima*, *'unum*, *pitu*, *walu*, *siyam*, *napulu*'.

This dialect group of the two eastern islands of the Visayas chain has no generally agreed name. Speakers often call their language *Binisayaq*, 'Visayan', or else adopt the regional nickname *Waray-waray*, which comes from their local word *waray* 'there isn't any'. The names *Samarëño* and *Samar-Leyte* are also used.

WELSH

PERHAPS 500,000 SPEAKERS

United Kingdom

Welsh is one of the two surviving CELTIC LANGUAGES (see map there) that descend from the speech of southern Britain at the time of the Roman conquest: the other is Breton. Although Latin is seen in Wales on stone inscriptions of all kinds that date from the four centuries of Roman rule, Celtic probably remained the everyday language for a large proportion of the population. So it is that after the Anglo-Saxon conquests had spread over what is now England in the 5th and 6th centuries, no trace remained of any native Latin-speaking communities. Celtic speech, alongside religious and literary Latin, lived on and flourished beyond the bounds of the initial Anglo-Saxon expansion in the western part of Britain, the country now known to us as Wales.

Wealas, modern *Welsh*, is in origin simply the Anglo-Saxon term for 'foreigner'. The Welsh name for themselves – equally unspecific – is *Cymry*, 'fellow-countrymen': the language is known as *Cymraeg*.

Britain, Latin *Brittania*, *Britannia*, is itself a Celtic word in origin, as seen in Welsh *Ynys Prydein* 'the island of Britain'.

Taliesin's poetry tells of a 6th-century king Urien of Rheged in southern Scotland; *Y Gododdin* of Aneirin describes a disastrous battle between Celts and Northumbrians at Catterick in about 600. These two heroic poems, wherever they were composed and at whatever later time they were written down, mark the beginning of recorded Welsh literature. Welsh became the language of great poetry of many kinds. The prose tales usually called the 'Mabinogion' include the earliest known versions of the tales of

King Arthur, the 'once and future king', legendary hero of Celtic resistance against the Saxons.

Welsh was the official language of the independent principality of Wales, but that succumbed to the English in 1282. Over the next centuries Wales became increasingly integrated with England, administratively at least, and English gained ground. Ironically it was under a Welsh dynasty – the Tudor Kings of England, in the 16th century – that this development was first enforced by law. The Act of Union of 1536 disqualified Welsh speakers from official employment and established English as the language of the courts. In the 19th and early 20th centuries, while schoolteachers punished and humiliated children who spoke Welsh, the Baptists and other Nonconformist Christian sects helped to keep Welsh literacy alive. Determined resistance led to the gradual easing of official discrimination between 1944, when the first publicly funded Welsh-language schools were established, and 1973, when the first court case was heard entirely in Welsh.

At the beginning of the 20th century half the population of Wales (about 1,000,000 people) could speak Welsh; in 1981 only about 20 per cent knew the language, though in numerical terms that still meant 500,000 people. Few if any speakers are now monolingual – nearly all use English as well as Welsh. There may be 60 per cent or more of Welsh speakers in most of rural western and northern Wales, but in the capital, Cardiff, and its neighbourhood, there are fewer than 10 per cent.

Welsh literature goes back to the 8th century. In the Middle Welsh period, 1150–1400, the language of the court poets formed a standard.

The translation of the Bible in 1588 (into a form of speech which was influenced by this medieval poetry) and the use of biblical language by preachers meant that a standard language became familiar to more people than ever before. Nowadays there are local weekly newspapers, radio, television, and some literary and educational publishing. Education in Welsh is available up to university level.

There is a considerable Welsh-speaking community in London. In the 19th century a Welsh colony was established at Chubut and Puerto Madryn in Argentina. Here Welsh was recognised for official use until the 1930s, but was then replaced with Spanish. Only older people now speak Welsh.

Modern standard literary Welsh is a conservative form of the language: this makes it more difficult to learn, but those who can read it can also read the early classics. *Cymraeg Byw* or 'Living Welsh', a new standard forming a compromise between literary Welsh and the colloquial dialects, was introduced into schools in 1964.

Northern dialects of Welsh are relatively conservative; the south-eastern dialects are structurally furthest from standard literary Welsh. There are some differences in vocabulary between north and south: 'road', northern *ffordd*,

lôn, southern *heol*; 'oven', northern *popty*, southern *ffwrn*; 'milk', northern *llefrith*, southern *llaeth*; 'cake', northern *teisen*, southern *cacen*.

Cacen 'cake' is one of the many English loanwords in Welsh, like *sir* 'shire, county', *taten* 'potato', *sosban* 'saucepan', *zô* 'zoo'. But much of the vocabulary goes back to the Celtic and Indo-European roots of the language: *dant* 'tooth', *môr* 'sea'.

English shares much with Welsh, in its sound pattern, its idioms and its vocabulary. Phrases such as *Sut rydych chi?* 'How are you?' are exactly mirrored in the two languages. Welsh loanwords in English include *pert*, which means 'pretty' in Welsh. Few English linguists know Welsh, so the similarities tend to be overlooked or played down.

Welsh looks difficult to the unfamiliar eye because of some unusual spelling conventions: *w* and *y* are common vowels; *ll* is an unvoiced *l*; *dd* is a voiced *th* sound. To use a Welsh dictionary, one must allow for the so-called mutations of initial consonants in normal speech and writing. The dictionary form *Cymru* 'Wales' can be found in ordinary text, in various surroundings, as *Cymru*, *Chymru*, *Gymru*, *Nghymru*: thus *yng Nghymru* 'in Wales'. For a table of numerals see BRETON.

Latin loanwords in Welsh

Welsh naturally has many Latin loanwords, dating from the time of the Roman Empire, when early Welsh and Latin coexisted for four hundred years: *lleidr* 'thief'; *sebon* 'soap'; *meddyg* 'physician'; *gwin* 'wine'. No wonder that the names of the days of the week are very like those in the Romance languages. In both sets, five of the names commemorate ancient Roman gods.

Welsh	French	
Dydd Llun	lundi	Latin <i>Dies lunae</i> 'Moon's day'
Dydd Mawrth	mardi	Latin <i>Dies Martis</i> 'Mars's day'
Dydd Mercher	mercredi	Latin <i>Dies Mercurii</i> 'Mercury's day'
Dydd Iau	jeudi	Latin <i>Dies Jovis</i> 'Jupiter's day'
Dydd Gwener	vendredi	Latin <i>Dies Veneris</i> 'Venus's day'
Dydd Sadwrn	samedi	Latin <i>Dies Saturni</i> 'Saturn's day'
Dydd Sul	Dimanche	The French originates as <i>Dies dominica</i> , 'Lord's day'; the Welsh, like English, as 'Sun's day', Latin <i>Dies Solis</i>

6th-century Wales: Christian hymns in competition with bardic poetry

Arrecto aurium auscultantur captu non Dei laudes canora Christi tironum voce suaviter modulante neumaque ecclesiasticae melodiae, sed propriae, quae nihil sunt, furciferorum referto mendaciis simulque spumanti flegmate proximos quosque roscidaturo praeconum ore ritu bacchantium concrepante.

With ears pricked up, you attend not to the praises of God, as the tuneful voice of the apprentices of Christ rings sweetly, nor to the breath of religious melody, but to praises of yourself, which are nothing, from the mouth of convicts, stuffed with lies and sure to bedew bystanders with foaming spittle, yelling like priests of Bacchus.

Gildas, *The Ruin of Britain* 34

This Latin lament by a 6th-century abbot is addressed to the Welsh prince Maelgwn.

WOLAYTTA

PERHAPS 1,000,000 SPEAKERS OF OMETO DIALECTS

Ethiopia

Wolaytta is the most important representative of a close-knit group of dialects of southern Ethiopia – to some or all of which names such as *Ometo* and *Welamo* have been given. They belong to the OMOTIC LANGUAGES. Wolaytta is now one of the official literary languages of Ethiopia.

Speakers call themselves *Welamo*. This word, which may be spelt *Walamo*, is often used as a name for the language, but the local name is *Wolaytta*. Variant forms include *Welaita*, *Waratta*; the official term in Amharic is *Wolami-nya*. *Ometo* means ‘people of the Omo river’ – thus it is identical in origin with the modern name of the larger language group, *Omotic*.

Wolaytta and the Ometo dialects (see map at GONGA) are the speech of small once-independent kingdoms which were conquered by Menelik of Ethiopia in 1894. The first brief wordlist of an Ometo dialect was published by the explorer Charles Tilstone Beke in 1846.

In these dialects – unlike the Gonga languages – neither vowel length nor tone appears to be a significant feature of the sound pattern. The first ten numerals in the Kullo dialect of the town of Jimma are: *ita*, *naa* (or *laa*), *hezu*, *oyda*, *icesh*, *osuphuna*, *laphuna*, *hosphuna*, *uduphuna*, *tamma*.

WOLOF

2,000,000 SPEAKERS

Senegal, Gambia

One of the Atlantic group of NIGER-CONGO LANGUAGES (see map at FULANI), Wolof has reached the status of de facto national language of Senegal, more widely broadcast on radio and television than any other except French.

Wolof is the speakers' own name for their language. The French spelling *Ouolof* is often found. Speakers call their region *Dyolof*: hence *Jolof*, occasionally used in English and French as a name for the language.

Wolof is the first language of most of north-western Senegal, including the capital, Dakar, itself, and the Atlantic coast: as such it is one of the six official regional languages of the country, with two million speakers. To the east and south it is a second language, known to perhaps another four million people – to the great majority of towns-people a rapidly growing number of villagers.

As a coastal people, the Wolof ethnic group, the largest in Senegal, were the first to make significant contact with Europeans. Already in the 16th century Portuguese traders found Wolof interpreters almost as useful to them in this whole region as MANDEKAN speakers. In modern Senegal it became the most widespread language of trade. It is now essential for those looking for work outside their own districts and for all who migrate to a city. Wolof is more and more the language that children learn first, especially those whose parents are from different ethnic backgrounds.

In the Gambia Wolof is spoken by about 100,000 people along the north bank of the river and in the capital, Bathurst. On the south bank KRIO is still used, though it is giving way to Wolof.

Numerals in Wolof, DIOLA and FULANI

	Wolof	Diola	Fulani
1	bèn	-əkon	go'ó
2	ñār	-gaba	dīdī
3	ñèt	-fègir	tati
4	ñènt	-bākir	nayi
5	jūrom	futək	jowi
6	jūrom-bèn	futək di -əkon	jeego'ó
7	jūrom-ñār	futək di -gaba	jeedidī
8	jūrom-ñèt	futək di -fègir	jeetati
9	jūrom-ñènt	futək di -bākir	jeenay
10	fuk	unyən	sappo

In all three languages the numerals '6' to '9' are formed as '5+1', '5+2' etc. In Diola, the basic four numerals take a prefix matching that of the noun: for example, *si-jamən futək di si-gaba* 'seven goats'.

XHOSA

6,900,000 SPEAKERS

South Africa

One of the BANTU LANGUAGES, Xhosa is spoken by the second largest language community of South Africa, after Zulu and ahead of Afrikaans. With Swazi, Ndebele and ZULU (see map there) it is a member of the Nguni language group. Long-standing Xhosa interaction with Khoe and other KHOISAN LANGUAGES is signposted by the fact that most Xhosa place names and many personal names are of Khoe origin.

Xhosa used to be called *Kaffir* or *Kaffrarian* (*Caffre* in Portuguese) – a term that equally covered Zulu and several other Bantu languages.

Xhosa shares its older history with ZULU, but – lying at the western end of the ‘Nguni’ dialect spectrum – Xhosa dialects were most exposed to influence from Khoisan languages, from Afrikaans and from English.

Linguists believe that it was as late as the 18th century that Xhosa completed the development of its unusual sound pattern. This was when Khoe and Xhosa speakers began to merge their social structures, the Khoekhoe becoming members of Xhosa lineages and the Xhosa merging into Khoe chiefdoms. There had probably been interaction between the two groups for much longer than this; now, at any rate, great numbers of Khoe loanwords were adopted into Xhosa, such as *irhamba* ‘puff-adder’, and with them came the click consonants so typical of the so-called Bushman languages of southern Africa and so rare elsewhere. An astonishing total of twenty-one (some say twenty-five) consonants, including fifteen clicks, were added to the Xhosa sound system at this period.

The bitter and damaging ‘Kaffir Wars’ of the

late 18th and early 19th centuries were sparked by competition for land between Xhosa speakers and Europeans. European interest in the Xhosa language is first evident late in this period. It was only in 1826 that John Bennie, of the Glasgow Missionary Society, produced the first serious dictionary, *A systematic vocabulary of the Kaffrarian language*. John W. Appleyard, a Wesleyan missionary, published an important grammar in 1850.

A second social upheaval has been the labour migration of Xhosa-speaking men within South Africa, focused on Johannesburg and the neighbouring settlements, and intensifying in the 1960s. ‘The Transkei’, one of the homelands of the late apartheid period in South Africa, was intended as a Xhosa reservation.

As a majority language in its region, Xhosa now functions as a lingua franca for speakers of several smaller languages. There is a good deal of multilingualism between Xhosa on the one hand and Zulu, southern Sotho and English on the other. Nowadays Xhosa naturally incorporates English loanwords: *ibhasi* ‘bus’, *ititshala* ‘teacher’.

Isikhwetha, a secret language, is still used by young men undergoing traditional Xhosa initiation rituals. *Hlonipha* is the women’s language of ‘respect’ or ‘avoidance’.

Hlonipha, the avoidance language

Hlonipha means ‘respectful avoidance’ of certain expressions and forms of behaviour, an avoidance that is particularly expected of women who are married or engaged to be married. To linguists, *hlonipha* is particularly the form of language that is used to avoid taboo expressions.

Numerals in Xhosa and Zulu

Xhosa		Zulu
-nye	1	-nye
-bini	2	-bili
-thathu	3	-tatu
-ne	4	-ne
-hlanu	5	-hlanu
-thandathu	6	isitupa
isixhenxe	7	isikhombisa
isibozo	8	isishiyagalombili
ilithoba	9	isishiyagalolunye
ilishumi	10	ishumi

Expressing high numbers in standard Xhosa and Zulu requires many syllables. For ‘29 trees’ the Xhosa is *imithi engamashumi amabini an-ethoba*; the Zulu equivalent is *imithi engama-shumi amabili nesishiyagalolunye*. In the modern colloquial languages, English loanwords are usually used for high numbers.

In the traditional rules as they have existed among Xhosa and Zulu speakers, women must not pronounce the names of their fathers-in-law, their mothers-in-law nor some other male in-laws. They must not even use syllables from those names: ‘Thus, a woman whose father-in-law is named Bongani must avoid the name itself and the syllables *bo* and *nga* – wherever they occur in speech . . . the effect on each individual woman’s speech may be dramatic.

‘A variety of linguistic mechanisms is used to achieve avoidance, including consonant substitution (e.g. *ulunya* ‘cruelty’ becoming *uluchya*), ellipsis (e.g. *umkhono* ‘foreleg’ becoming *um’ono*), synonymy (if *kufa* ‘die’ is to be avoided, *kushona* ‘set, die’ used in its place), derivation (if *imbuti* ‘goat’ is to be avoided, *inkhuleko* ‘thing for tethering’ invented in its place) as well as neologism, archaicism and borrowing’ (R. K. Herbert).

Nowadays borrowing is often from English. In the past, frequent intermarriage between Xhosa and Khoisan speakers would have led to the borrowing of sounds and words from Khoisan languages into *hlonipha* speech. This is probably how the rich range of Khoisan consonants, including clicks, came to be adopted into Xhosa and the other Nguni languages. In turn, Southern SOTHO shows heavy borrowing from Zulu (and has adopted one of the Khoisan clicks), and this can be attributed to Zulu-Sotho intermarriage in the 19th century.

Hlonipha does not fit well with modern lifestyles. It can make it impossible to fill in forms correctly. Women often have to abandon it at their place of work. ‘Many schoolchildren consider the whole concept a joke’ (R. Finlayson). It seems likely that the practice will not last through many more generations.

Based on papers by Robert K. Herbert and R. Finlayson in *Language and social history: studies in South African sociolinguistics* ed. Rajend Mesthrie (Cape Town: David Philip, 1995)

YAKUT

300,000 SPEAKERS

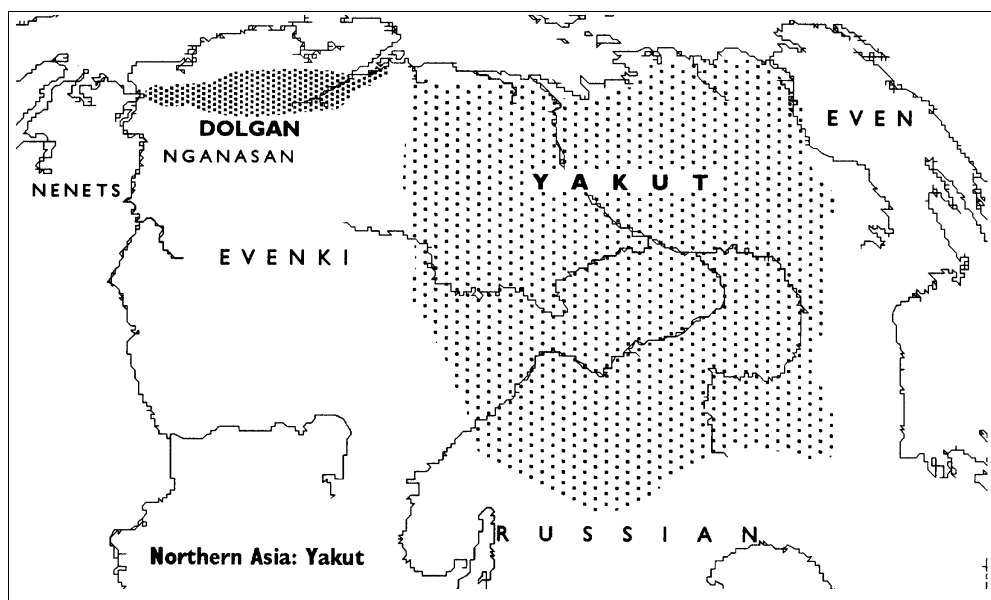
Russia

One of the TURKIC LANGUAGES, Yakut (or Sakha) has been geographically separated from its relatives for many centuries. The Yakut, in their vast Siberian heartland, have a very different environment from that of other Turkic language speakers, and have developed a different way of life, very like that of neighbours who are speakers of TUNGUSIC and PALAEO-SIBERIAN LANGUAGES.

To themselves, Yakut are *Sakha*, 'people of the edge'. They did indeed occupy the very edge of the habitable world as known to early Turkic speakers. In Buryat Mongol this same word appears as *Yakhuud*, from which Russians learnt to call the people and their language *Yakut*.

It may have been the Mongolian expansion of the 13th century that separated Yakut speakers from the Turkic regions to the south-east. Their legends tell of an original home on the shores of Lake Baikal. Their territory, formerly Yakutia, is now called Sakha. Fourteen times the size of Britain, Sakha has a population of around 500,000, just over half of whom are Yakut speakers. Russian speakers account for nearly all the rest. The region was already being explored by Russian traders in the 17th century: its modern capital, Yakutsk, was founded by them as a fort and trading post in 1632.

The first alphabet for Yakut, an adaptation of the older Cyrillic alphabet, was devised by Russian missionaries in the early 19th century. A Latin alphabet with a great many additional let-



ters was used 1922–39, when the modern Cyrillic alphabet – again with extra letters – was introduced. It is only in the 20th century that Yakut has been regularly used in education and the press.

Yakut is close enough to its relatives to be easy for Turkish speakers to learn, though they have to cope with consonant assimilation that multiplies the possible forms of suffixes. The plural suffix, which may be *-ler* or *-lar* in Turkish, has sixteen forms in Yakut: *-lar -ler -lor -lör -nar -ner -nor -nör -tar -ter -tor -tör -dar -der -dor -dör*.

The first ten numerals are: *biit, ikki, üs, tört, bies, alta, sette, arıis, toγus, uon*.

Yakut and its neighbours

The basin of the great River Lena forms most of Sakha, a self-governing republic of Russia. Per-

mafrost makes it impractical to build railways, metalled roads or heavy buildings. Rivers, ice-covered most of the year, are navigable only for a short summer season. Traditionally Yakuts are herders of cattle, horses and (in the further north) reindeer. They are also prominent in the fur trade. The Russian speakers who once threatened to overwhelm Yakutia are now fewer and less well paid.

In the far north-west are 3,000 speakers of the Dolgan language. They are ethnically distinct from the Yakut, but their speech is clearly a divergent dialect of Yakut which appears to have a strong admixture of some Tungusic language. These reinder herders may be descendants of Tungusic speakers who, at some time in the past, took to speaking Yakut.

Epic hero of the north

Yakut oral literature is rich, varied and rooted in the Siberian way of life. Prose tales, short poems and epics have been recorded. In poetry, consonant alliteration is added to the vowel assonance that is built into the structure of all Turkic languages. An epic hero introduces himself:

Хардарылаах айаннаах, халыан сырылаах, хара тыа хайдан түспүтүн курдук хангыл хара аттаах хаан айыы сизнэ Халыадымар Бэргэн – дизн киһибин.	Khardarılaakh aiannaakh, khalian sırılaakh, khara tia khaidan tüspütün kurduk khangıl khara attaakh khaan aiıı siene Khalyajimar Bergen dien kihibin.	Facing a desolate road, Riding a wild-stepping Forest-black Mountain-offspring Untamed raven-black horse, Inheriting blood-sin, I am the man called Khalyajimar Bergen.
--	---	--

From John R. Krueger, *Yakut manual* (Bloomington: Indiana University, 1962) p. 225

YAO (AFRICA)

1,200,000 SPEAKERS

Malawi, Tanzania, Mozambique

Yao, one of the BANTU LANGUAGES, is spoken in northern Mozambique and in neighbouring districts of Tanzania and Malawi (for map see MAKUA).

Its speakers traditionally give their ancestral home as a hill called Yao, near Muembe, in north-eastern Mozambique. In Nyanja, Yao speakers are called *Achawa*, a word derived from *Chao*, the singular form of *Yao*.

Traditionally a farming and cattle-keeping people, Yao speakers controlled trade routes between the Lake Nyasa region and the Indian Ocean. For many centuries they dealt with Swahili-speaking traders at the coast, and travelled far inland. In the 19th century they were still prominent in the slave trade. At that time they were expanding their territory in what is now southern Malawi, where Yao speakers formed a ruling class and raided for slaves. The trade was outlawed in 1896, but the language had spread rapidly over the area that they ruled.

By the later 19th century many of the Yao had adopted Islam under Swahili influence. Swahili was the written language of the religious schools.

Unlike its coastal relatives, Makua and Makonde, Yao is unmistakably a tonal language. Linguists have distinguished up to five tones, although, after allowing for variation due to phonetic context, these can be reduced to two. The tones of Yao seem to have no relation to the reconstructed tones of proto-Bantu, and may have originated quite recently.

There are no great dialect differences in Yao, in spite of its wide extent. Since the Yao-speaking country has never been politically united, this geographical uniformity must be a sign of

frequent travel and good communications among Yao speakers in the recent past.

The first ten numerals in Yao are *-mo*, *-wili*, *-tatu*, *mcheche*, *msano*, *msano na -mo*, *msano na -wili*, *msano na -tatu*, *msano na mcheche*, *likumi*.

'A man who is going to tell a story says *Tele!* "Full!" and the people answer *Lokote, lokote, kaselo ndi! Lakata!* "Pick it up, the basket is full, heaped up to overflowing!" When he has finished the audience says *Ajokole chitolo, chitakununga!* "Take the rat off the fire or it will smell!"'

Meredith Sanderson, *A Yao grammar*
(London: SPCK, 1922) pp. 118–19

Yao traditions

'The yearly initiation ceremonies, *unyago*, are the principal cultural institution of the Yao. That for the boys, *lupanda*, is the most important. It is held at the end of the dry season, before the bush is burned. Each initiate is accompanied by a sponsor, *nkamusi*, who may be a brother, uncle or friend. Before they set out for the place appointed to be *lupanda*, they all assemble before the chief to be anointed with millet-flour specially prepared by his head wife. The word *lupanda* refers to a forked stick over which the sacrificial flour is poured. This stick is planted at the place of assembly, the *masakasa*. The next day the initiates go to a spot in the bush, where the circumcision is performed, and remain in the bush school, *ndagala*, for about two months, until the wounds are completely healed.

'During these weeks the boys are subjected to rigorous discipline and instructed in various skills

and tribal lore. On the last night before the return to the village the initiates undergo a ceremonial bathing, each boy being carried to the stream by a woman, who thereafter is addressed as “elder sister”. Finally, with new names, they march to the chief’s village to be redeemed by him from the master of ceremonies, *m’micira*, with gifts of cloth. After initiation it is a deadly insult to call a boy by his child-name.

‘Most Yao men and women act as sponsors of initiates at least once, and all participate yearly in the great initiation festival for boys and girls. Consequently as they repeat each year the lessons of their own initiation, the ceremonies are the main vehicle of Yao culture.’

Mary Tew, *Peoples of the Lake Nyasa region*
(London: Oxford University Press, 1950)
pp. 19–21, abridged

YAO LANGUAGES

PERHAPS 900,000 SPEAKERS

China and south-east Asian countries

Yao and Nu, with MIAO (see map there), belong to the small Miao-Yao language family, which some linguists consider to be a component of the wider grouping of AUSTRO-TAI LANGUAGES.

Yao is the official term for both languages in China. Speakers call themselves *Mien*; *Man* is the term used in Vietnam.

Speakers of the two Yao languages live in the hills of ‘Lingnan’, southern China. Little is known of their history, as with many other peoples of this region of mountains and valleys, but they are said to have moved southwards from Hunan, under pressure from expanding Chinese rule, in the 12th and 13th centuries. This was possibly the first move in a gradual spread southwards which certainly had reached northern Vietnam by the 17th century. The migration trend may have speeded up in this century, so that some groups are now living as far south as Thailand. The majority remains, however, in the Chinese provinces of Guangxi and Guangdong. Here Yao speakers tend to be hill farmers, while most valley rice-growers speak Chinese or Tai languages.

Though quite distinct from Chinese in their origin, Yao and Nu show very heavy Chinese influence from the centuries of symbiosis be-

tween the two languages. Many Yao speakers in China are now bilingual in Chinese. Yao and Cantonese, the Chinese language that is most widely spoken in Yao areas, show interesting similarities in their sound pattern. Both have a similar range of syllable-final consonants, *-m*, *-n*, *-ng*, *-p*, *-t*, *-k*, and a distinction of vowel length.

Numerals in Yao, Nu and Miao			
	Yao (Mien)	Nu	Miao
1	yat	i ^{1a}	i
2	yi ¹	aw ¹	i ¹
3	po ¹	pe ¹	pu ¹
4	pyey ¹	pla ¹	prey ¹
5	pya ¹	pru ¹	pra ¹
6	ču ⁷	taw ⁵	to ⁵
7	syē ⁶	syong ⁶	čiong ⁶
8	cet ⁸	yu ⁸	ji ⁸
9	do ²	caw ²	čio ²
10	tsyop ⁸	caw ⁸	ku ⁸

The Yao languages are highly tonal: one analysis of Nu recognises 11 tones, numbered 1 to 8 and (the highest) 1a to 3a. In Nu the numerals ‘9’ and ‘10’ differ only in their tone: the same is true of ‘1’ and ‘2’ in Miao.

YI

SEVERAL MILLION SPEAKERS

China

With Burmese, which has had a very different history, Yi is one of the two major members of the Burmese-Lolo group of SINO-TIBETAN LANGUAGES (see map at BURMESE).

Lolo is a derogatory term in Chinese, used for speakers of Yi and related languages of Yunnan: westerners used it too, often including the AKHA under this name. The less loaded, but equally unspecific, Chinese term *Yi*, 'hill people', is now preferred by Yi speakers themselves. The group of dialects to which Yi belongs has been called 'Northern Loloish' and 'Nasoid' by linguists, *Nosu* and *Mosu* being among the most common of Yi speakers' traditional names for their own people.

Yi speakers are historically a fiercely independent people. They are found in Chinese records over nearly two thousand years, and sources differ over the extent to which, even now, the 'Independent Yi' of the Cool Mountain, Liangshan in southern Sichuan, have submitted to Chinese control. Yi speakers are traditionally raiders and farmers, producing buckwheat for subsistence and opium for profit.

They distinguished, within their stratified society, 'Black Lolo' – rulers and conquerors, who belonged to the true clan structure – from 'White Lolo', serfs and the descendants of slaves, of varied origin. These again were distinct from slaves. Many slaves and White Lolo had certainly been Chinese speakers; Miao were also enslaved. Thus, in the past, the Yi language spread widely through the practice of conquest and enslavement.

Not all 'White Lolo' are of local origin. The British adventurer Donald Brooke was foolish enough to lead a party into Liangshan in 1909. He was killed, and the rest of his party was enslaved.

Nowadays, with the political and social ascendancy of Chinese, the position is probably reversed. The slaves have been freed, and Yi itself may be a shrinking language as former Yi speakers become Chinese and adopt Chinese speech. Estimates of current speakers of Yi vary from 2,000,000 to 5,500,000.

Most Yi dialects have ten vowels, three tones and a distinction of vowel register, which may be 'laryngealised' or clear. The language has borrowings from Chinese, from Pali (by way of Burmese and Shan) and a few from European languages: the word for 'soap' is Portuguese.

The numerals in Yi or Lolo script

The Yi pictographic writing system (or syllabary, as some would describe it) was used by priests, *pimu*, to record rituals and magical and medical prescriptions. It is 'the same regardless of dialect and thus predates dialect differentiation', according to David Bradley's argument: at any rate its history goes back several hundred years. Although some characters look like Chinese, it is in reality completely independent of the Chinese script. Its application varied: in some districts it was written horizontally, in some vertically like Chinese. Its range of characters varied from place to place, up to around 8,000.

Yi numerals in script

1
 2
 3
 4
 5
 6
 7
 8
 9
 10

After some decades of discouragement, China's minority policy eventually permitted the adaptation of Yi script for modern uses. It has been reborn as a carefully designed true syllabary of 819 characters – but few learn it.

For the pronunciation of the numerals, see table at LAHU.

YIDDISH

PERHAPS 2,000,000 SPEAKERS

*United States, Israel, Russia,
Ukraine and many other countries*

Yiddish is one of the GERMANIC LANGUAGES – a language closely related to GERMAN, spoken by Jews. The Yiddish language and its culture have suffered more than any other from 20th-century barbarism. About three-quarters of its speakers, well over five million people, were killed in German-occupied Europe between about 1940 and 1945.

How do we explain Yiddish as a separate, German-like language spoken over the same territory where German is spoken?

The origin of Yiddish can be traced to the Rhineland cities of Germany in the early Middle Ages – for Yiddish shows clear links to the old German dialects of the middle Rhine. From their ancient settlements here, German-speaking Jews gradually spread eastwards and south-eastwards, beginning as early as the 10th century, across a vast area of central Europe.

It has been supposed that Jews in medieval Germany initially spoke German no different from that of other inhabitants, and that Yiddish gradually became a distinct language because of the separateness, partly compulsory, of Jewish communities in medieval German cities; because of their independent culture and religion, rooted in their religious languages, Hebrew and Aramaic; and also because, as they spread eastwards across central Europe and into Russia, and as they began to leave Germany itself, Yiddish speakers were eventually no longer surrounded by German speakers.

It is not entirely a false picture: but the origins of Yiddish are more complex, and older, than this. The Jewish communities of the Rhineland were, in the 10th-century context, part of a

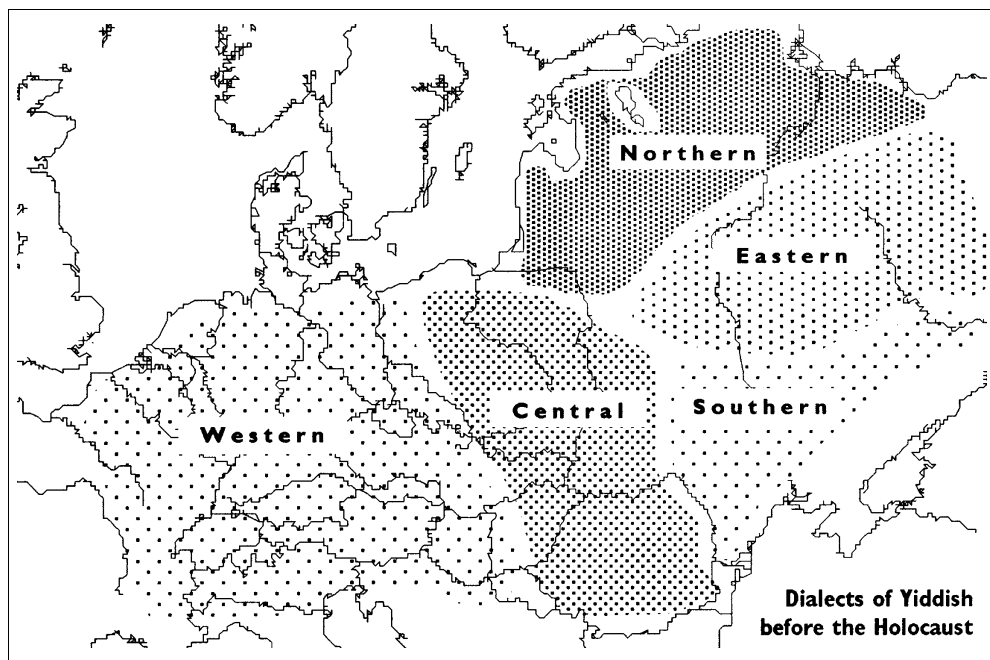
culture region extending not eastwards into Germany but westwards across most of France; thus medieval French, as well as Aramaic, Hebrew and medieval German, had its part in earliest Yiddish, which has been aptly described (by Max Weinreich, *History of the Yiddish language*, 1980) as a 'fusion language'.

Yidish is the regular equivalent, in Yiddish, of the German word *Jüdisch* 'Jewish'. The term is first found in print in 1597. Among Jews the language has, just as appropriately, been called *Taytsch*, the Yiddish equivalent of *Deutsch* 'German'. In Hebrew terms it may be regarded as the language of the *Ashkenazim*, the 'people of Ashkenazi', which is the medieval Hebrew name for Germany.

Some German words in Hebrew script are found in 12th-century Jewish manuscripts, but the first real texts in a language that can be identified as Yiddish date from the 14th century.

By the 18th century Yiddish-speaking Jewish settlements, most of them in cities, existed from eastern France and north Italy eastwards as far as the Baltic states, Ukraine, Moldavia and the Crimea. The majority, probably, was in the largely German-speaking Holy Roman Empire and Austrian Empire, but a considerable minority was to be found living under various governments to the east. Major cities of Yiddish-speaking settlement included Krakow, Wroclaw, Warsaw, Vilnius, Lvov, Chernovtsy, Odessa and Kiev.

Jews had not been allowed to settle in the old Russian Empire: however, as Russia annexed Ukraine, Belorussia, Lithuania, eastern Poland and the Khanate of the Crimea, mostly in the



18th century, it also annexed a large number of Jewish citizens. By the beginning of the 20th century they numbered over five million. Nearly all of these were speakers of Yiddish, and most of them lived in separate communities, in urban ghettos and rural shtetls. The Yiddish-speaking population of Belorussia was particularly large: later, in Soviet times, the Belorussian coat of arms would bear the words 'Proletarians of all countries, unite!' in Belorussian, Russian, Polish and Yiddish. Even in the 1990s there are Yiddish radio broadcasts in independent Belarus.

Yiddish literature has been of world importance only since the 18th century, when there were written the most lively and readable of all the texts that have ever been called 'mystical', the tales of Rabbi Nachman of Bratslav.

The diaspora of the 19th and 20th centuries

In the 19th century a westward migration gathered pace, speeded by increases in anti-Jewish activity. Preferred destinations were western Europe, the United States, Argentina and other Latin American countries. Already by 1900 the United States could be regarded as the centre of

Yiddish and its culture. There was, and still is, a very large Yiddish-speaking population in New York. The East End of London – Aldgate, Whitechapel, Spitalfields – was another major Yiddish-speaking community. In the early 20th century migration to Israel became an option, one that has continued to attract large numbers of Yiddish speakers.

Meanwhile, towards the end of the 19th century there was a growing awareness of Yiddish as a language. The first World Congress of Yiddish was held in Czernowitz in 1908.

Most of the German and central European speakers of Yiddish who had not emigrated were killed in the early 1940s. Very few indeed are now to be found in Germany, Poland, the Czech Republic, Slovakia or Hungary.

Most remaining Belorussian, Ukrainian and Lithuanian Jews were also killed during the German occupation of eastern Europe. Of the surviving Jews of the Soviet Union, whose numbers gradually declined through emigration and assimilation, only about a sixth declared their first language as Yiddish in the 1979 census. There are now perhaps 80,000 speakers in Russia, 80,000 in Ukraine and 10,000 in Belarus. The

‘Jewish Autonomous Region’ of Birobijan, established in eastern Siberia in 1934, has only about 7,000 Jewish inhabitants – though much higher estimates have been published.

Since the Second World War, the general role of Yiddish has been as the mother tongue of refugees from Jewish communities all over eastern Europe. But in this role it will not survive long. Yiddish is generally heavily discouraged in Jewish education, which favours linguistic and cultural assimilation – to Hebrew in Israel, to national languages elsewhere.

But Yiddish has a special importance as the language of fundamentalist communities of Ashkenazi Jews of eastern European origin, concentrated in New York and Israel. Their children are still brought up with Yiddish as their mother tongue, and modern Hebrew is avoided. For this community, Hebrew is the language of religious texts but Yiddish is the language of exposition and of festivity. For them, the old-established Yiddish newspapers – the weeklies *Forverts* and *Der Yid* in New York, the daily *Letste Nayer* in Israel – retain their importance.

Yiddish is almost purely German in its structure. Its vocabulary comes largely from German but also from the other languages spoken by Jews. From Hebrew come numerous terms for religious concepts and tradition. Other loans are drawn from Aramaic, from medieval French, Provençal and Italian, and from the Slavonic languages. Modern Yiddish is rich in English and Russian loanwords. In return, colloquial English borrows freely from Yiddish: *kosher*, *schmaltz*.

Yiddish is traditionally written in Hebrew script – the feature that most obviously distinguishes it from its close relative German. A new standardised orthography was agreed in 1937. The vowels are written fully in native German words, while loanwords from Hebrew are written with

their usual Hebrew spelling in which most vowels are unmarked. A now-standard transliteration into the Latin alphabet, based on the Lithuanian pronunciation of Yiddish, is quite often used.

Yiddish must be seen now as a threatened language. It retains official status in Russia and Belarus, but has none in Israel. As linguistic assimilation proceeds it is likely to give way to Hebrew there, to Spanish in Argentina, and to English in the United States.

Yiddish happens to share with English the alternate forms *a*, *an* for the indefinite article: *a boyom* (German *ein Baum*) ‘a tree’; *an oyg* (German *ein Auge*) ‘an eye’.

The pre-1945 dialects of Yiddish

After the destruction of central European Jewry in the early 1940s, the former regional dialects of Yiddish scarcely exist in their original locations. The old dialect boundaries can be linked to late medieval and early modern political and cultural frontiers; but they remained fluid and relatively insignificant in the Yiddish context, owing to continuing frequent migration, travel and inter-marriage among Jewish communities.

Numerals in Yiddish		
eyns	1	אײנס
zvay	2	צוויי
drey	3	דריי
fir	4	פיר
finf	5	פינף
zcks	6	זעקס
zibn	7	זיבן
akht	8	אכט
neyn	9	ניין
tsen	10	צען

Was that a party

Oy, iz dos geven a simkhe!
Fish un fleysh gebrotn,
gut gefefert, feyn gezaltsn,
Punkt vi s'iz getrotn . . .

Oh, was that a party!
Baked fish and roast meat,
Well spiced, finely salted,
Just at its best . . .

Refrain of a popular song

Mayrev-Yidish 'Western Yiddish' or *Oyberlendish* 'upland language', the dialect of German-speaking lands, Bohemia, northern Italy and Hungary, was already in decline by the early 20th century, because Jews in these areas – especially in Germany itself – were becoming

rapidly less isolated, and preferred to adopt local standard languages.

Mizrakh-Yidish or 'Eastern Yiddish' can be regarded as a grouping of three dialects: *Polnish*, *Litvak* and *Galitsianish-Ukraynishe* Yiddish, the latter also called *Interlendish* 'lowland language'.

Yiddish alphabet: Hebrew and Latin

א ב ג ד ה ו ז ט י כ ל מ נ ס ע פ צ ק ר ש ת װ ן ן ן ן ן
a b g d h u z t i y k h l m n s e p f t s k r s h t v z h t s h e y a y o y

YORUBA

20,000,000 SPEAKERS

Nigeria, Benin

Yoruba is one of the four national languages of Nigeria, alongside Hausa, Igbo and English. It belongs to the family of NIGER-CONGO LANGUAGES.

Yoruba was originally an outsiders' name for the language and people, but it has long been widely accepted. Early reports refer to *Hio*, *Eyo* and *Oyo*. *Aku* was the name given to their language by the once-influential community of speakers at Freetown, Sierra Leone. Speakers of Yoruba dialects in Benin are called *Nago* (a name used by the Fon or Ewe) or Anago or Nagot; in Togo they are called *Ana*.

The 'Defoid' languages, the immediate group to which Yoruba belongs, are spoken in adjacent parts of south-western Nigeria and Benin, and are probably long established there. Yoruba itself extends into Benin and Togo, and Yoruba traders took the language northwards and eastwards into north-eastern Ghana and the middle Niger valley.

European knowledge of Yoruba speakers and their language came late – not before the early 19th century. At that period, however, many Yoruba were already being sent to the Americas as slaves. Missionary activity began soon afterwards, with the help of freed Yoruba-speaking slaves who had resettled in Freetown. The first publications in Yoruba – brief teaching booklets – were produced by John Raban in 1830–2, but the main figure in early Yoruba literacy was the linguist Samuel Crowther, whose *Yoruba vocabulary* appeared in 1843. The centre of activity moved from Freetown to the Yoruba country around 1850; it was also about then that the

orthography of Yoruba was fixed in its modern form. *Iwe Irohin*, the first vernacular periodical printed in West Africa, appeared at Abeokuta in 1859–67.

Yoruba now has a substantial press and publishing industry. As the local language of Lagos and its hinterland its importance is likely to remain high. Standard Yoruba, much used in the media, is now known and used by speakers of many neighbouring languages, including its relatives Itsekiri and Igala.

Islam is an important religion in the Yoruba country: it came from the north, under Hausa influence, and Yoruba has borrowed many cultural words from Hausa (some originally Arabic: *fitila* 'lamp'). Modern Yoruba includes numerous English loanwords. Yoruba language and culture, shared by a high proportion of slaves of the late 18th and early 19th centuries, have been important in Brazil, Cuba and other American countries.

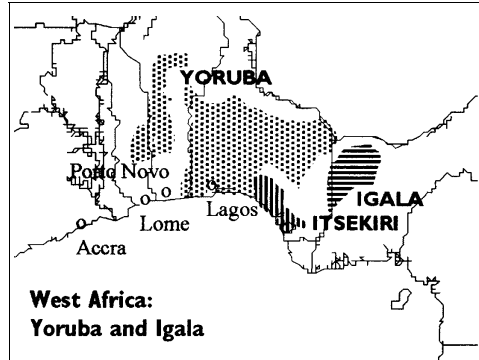
Yoruba origins

'All Yoruba regard the town of Ife as their place of origin and the *Oni* of Ife has been generally accepted as their ritual leader. The present royal lineages of the various Yoruba chiefdoms trace their descent from *Oduduwa* or *Odua*, the deified culture hero whom Ife mythology credits with the creation of the earth, and from whose seven grandsons the various branches of the Yoruba-speaking peoples claim descent.'

Daryll Forde, *The Yoruba-speaking peoples of southwestern Nigeria* (London: International African Institute, 1951) p. 4

Samuel Crowther

Born around 1806, brought up in Yoruba country, sold as a slave, Samuel (Ajai) Crowther was freed and settled in Freetown, Sierra Leone, where he studied at the Christian Institution in 1827. A natural linguist, he took part in the Niger Expedition of 1841–2: his *Journal of an expedition up the Niger and Tshadda rivers* was published in 1855. He studied Igbo, Tiv, Yoruba and other languages of Nigeria, wrote and translated in several of them, and was the first Christian bishop of West African origin. He died in 1891.



Dialects of Yoruba

Yoruba has three level tones, high, mid and low. There are no falling or rising tones. The first twenty numerals are *òkàn, méjì, métà, mérín, mārún, mēfà, mēje, mejo, mēsán, mēwá; mòkànlá, mejílá, métàlá, mérinlá, mārúnlá, mērindílógún, mētàdílógún, méjìdílógún, mòkàndílógún, ogún*. The numerals from sixteen to twenty are ‘counted down’ – ‘four less than twenty’, ‘three less than twenty’, and so on.

Standard Yoruba, close but not identical to the dialect of Oyo, is used in schools, in literature and in contacts between speakers of different dialects. Yoruba dialects extend into Benin and Togo. In Benin the language is called *Nago*, in Togo *Ana*.

Itsekiri or Isekiri or Jekri, with half a million speakers around Warri and Sapele, is sometimes considered a dialect of Yoruba. It is heavily influenced by EDO.

Igala, another member of the Defoid group, has 800,000 speakers.

It is dark

Ailélé, igi oko parere;

akope ẹruju ko ma kiyesi igba . . .

It is dark: the trees in the farm are silent;

One who taps palm-wine must look to his climbing rope.

Two oracular lines from *Sixteen cowries: Yoruba divination from Africa to the New World* ed. William Bascom (Bloomington: Indiana University Press, 1980) pp. 116–17

YUCATEC

500,000 SPEAKERS

Mexico, Belize

One of the MAYAN LANGUAGES (see map and table of numerals there), Yucatec is sometimes called *Maya*, for this is the language of Yucatán, and thus one of the languages directly associated with the monuments of Maya civilisation.

The first Spanish expedition to this region, in 1517, took two Indians prisoner and renamed them Melchior and Julian. The Governor of New Spain understood the Indians to be saying that their country was called 'Yucatán'. In fact they were saying that they grew *yuca*, 'cassava', to make bread.

Maya culture was in fact multilingual. Over more than two thousand years, Mayan speech had spread northwards and north-westwards from the Guatemalan highlands into the lowlands of south-eastern Mexico. It had already diverged into languages: Tzeltal, Tzotzil, Chontal, Chol and – rather different from these – Yucatec. As lowland Maya culture flowered, between 100 BC and AD 950, these languages were influenced more and more by one another.

Melchior and Julian, Yucatec speakers, were easily able to make themselves understood in the Tabasco district, where the local language was Chontal but Yucatec was well known. They were the Spaniards' first interpreters on the mainland of Mexico – along with a Carib woman from Jamaica who had been shipwrecked off Yucatan and had learnt the language of her new home.

The Lowland Maya languages were also heavily influenced by the language of the Olmecs, bearers of an older civilisation in central Mexico.

The Olmec language has not been deciphered, but it is clear that it belonged to the Mixe-Zoque group: one reason for believing this is that so many Mixe-Zoque loanwords of all kinds are to be found in Yucatec and other Mayan languages. The word for 'rabbit' was borrowed into Yucatec from Olmec (proto-Mixe-Zoque **kaya* 'rabbit') not because of the everyday importance of rabbits, but because it was a calendrical term: the Mexican calendar was shared and adopted among many cultures, remaining generally the responsibility of a priestly caste. Other Yucatec borrowings from Mixe-Zoque include pMZ **pom* 'incense', **cima* 'calabash', **kakawa* 'cacao' (see box), **kuku* 'turtle-dove'.

Yucatec also borrowed terms from other Mexican languages, including Zapotec (*pi'kku* 'dog' and *mani* 'deer') and Nahuatl. In its turn, Yucatec or another Mayan language has transmitted culture terms to languages of further south: for example, Mayan *cjenek* 'bean' is to be found in Xinca, Lenca and Jicaque.

Yucatec is first recorded – in short texts of historical and religious significance – in hieroglyphic stone inscriptions on Maya sculptures and buildings from the pre-Columbian period. The same hieroglyphic script was later used to write book texts, on long strips of bark paper, folded concertina fashion. Three, *Codex Dresden*, *Codex Madrid* and *Codex Paris*, written in the 14th or 15th centuries, survive to this day.

More traditional Yucatec literature was written down, in the three centuries that followed the Spanish conquest of Mexico, in the Latin alphabet that the Spaniards had introduced. The so-called *Books of Chilam Balam* are the greatest Yucatec texts. Chilam Balam was said to have

lived in the city of Mani before the Spaniards came, and to have prophesied a new religion. Copies of the *Book* – whatever its original form – appear to have been taken to various priestly centres, and in each place local additions were made. The *Books* that now survive include Spanish texts translated into Yucatec, as well as the history of Maya migrations and conquests, time charts and calendars, prophecy, ritual, astrology and medicine. Equally compelling is the *Book of the Songs of Dz'itbalché*, an 18th-century manuscript, containing lyric poetry and prayers: notable are two long poems or liturgies narrating the sacrifice of a captive.

Maya hieroglyphs

Central American hieroglyphic writing reached its highest elaboration among the Maya. Recent research, following a breakthrough in the 1950s

by Yuri V. Knorosov, has shown that the language could be fully written in a partly phonetic, partly ideographic system. Earlier specialists, such as Eric Thompson, had believed that many short Maya inscriptions on dishes and bowls were meaningless decoration. They are now known to include dedications, artistic information, and even 'recipes'. The first recipe to be deciphered was that for cacao – drinking chocolate.

David Stuart showed that this consists of a drawing of a fish, preceded by a comb-like sign, the syllable *ka-*, and followed by *-w*. The fish symbol turned out to be another version of the comb (the comb is really a fish's fin) so that he was able to read the whole word as *ka-ka-w* 'cacao'.

After Sophie and Michael Coe, *The true history of chocolate* (New York: Thames and Hudson, 1996)

ZANDE

1,200,000 SPEAKERS

Congo (Kinshasa), Central African Republic, Sudan

Zande is one of the most easterly of the Ubangi group of NIGER-CONGO LANGUAGES.

Zande is the name of the language. *Azande*, with the plural *a-* prefix, is the name of the people.

Zande, BANDA, GBAYA and Ngbandi (ancestor of SANGO) are believed to have come to the south-western Sudan and Central African Republic as a result of a migration from the west, over three thousand years ago. The early speakers of Zande, the Avongura, were to be found mostly in Sudan until the 18th century.

The warlike Avongura then moved south-eastwards and invaded the country between the Uele and Mbomu rivers around 1800, eventually conquering all the land from Bangassou (now in the Central African Republic) eastwards to Maridi (Sudan). Although they cannot have been more than a small minority over this huge area, their language, now called Zande, soon became dominant. Succeeding generations of conquered peoples began to consider it their

mother tongue and to consider themselves to be Azande.

The Azande attracted 19th-century Arabic-speaking slave traders but retained some political unity until eventually conquered by the French, Egyptians and British after 1899. Their lands were carved up by the colonial powers. The language remains widely spoken, though over a sparsely populated country (see map at SANGO).

Zande is a tonal language – the two tone levels can also be heard on the slit drum, *gugu*, traditionally used for signalling. In its many verb tenses, formed by infixes between the personal prefix and the verb root, Zande resembles Bantu languages. There are four noun genders: masculine, feminine, animal and inanimate.

Arabic influence is demonstrated in loanwords – including *babur* ‘steamboat’, itself an Arabic loan from French *vapeur*.

The first ten numerals in Zande are: *sa, ue, biata, biama, bisue, bisue bati sa, bisue bati ue, bisue bati biata, bisue bati biama, bawe*.

Two proverbs

Kperende na kpari tiru na sueru	The cricket could cry for itself with its wings
	[I can manage by myself]
Badia gbuku na dika ko ku gira rago	It was the owl's friend who kept him back till sunrise
	[Choose your companions carefully]

E. C. Gore, *A Zande grammar* (London: Sheldon Press, 1926)

ZAPOTEC

PERHAPS 450,000 SPEAKERS

Mexico

Zapotec is one of the Otomanguean family of AMERIND LANGUAGES, like MIXTEC (see map there). It is many centuries since Zapotec was the speech of a single political entity – if it ever was – and Zapotec is now a group of dialects, some of which are so different that they are not mutually intelligible.

The historic centre of Zapotec culture and language was the hilltop site of Monte Albán, above Oaxaca. Its heyday was AD 600 to 900, but the culture can be traced back with no significant break to 1200 BC. In the later pre-Spanish period, Mixtec speakers expanded eastwards, driving the Zapotec from Monte Albán and its neighbourhood.

The Zapotec, in turn, spread eastwards: for example, it is clear that the speakers of 'Isthmus Zapotec' migrated, not long before the Spanish arrival, from the Zaachila district (south-west of Oaxaca) to the Isthmus of Tehuantepec. According to glottochronology, the controversial statistical technique that attempts to measure the chronological distance between languages, the split between Zaachila Zapotec and Isthmus Zapotec took place 560 years ago.

Earlier Monte Albán inscriptions already show the existence of a writing system. The ritual calendar, of 260 days, stands out as clearly identifiable: the day signs in the script can be

identified from their shapes. There are many short inscriptions on various subjects, not yet fully identified: signs include human and animal heads and hands. Details of dress, in the figures depicted in the pictographs, seem to be important.

Among later inscriptions, the typical ones consist of four phrases: an upside-down human head with death-eyes and with detailed, varying head-dress; a 'mountain' with details on it; dates, which can be fully deciphered; and a fourth phrase, seemingly more miscellaneous. Alfonso Caso, who has worked on these texts, suggests that they are reports of conquests.

By contrast with Nahuatl and Mixtec, there are no early Zapotec manuscripts using the pictographs. So no direct help is available for those who try to decipher the Monte Albán pictographic inscriptions. The first Zapotec grammar, written in Latin script, was by J. de Córdova: entitled *Arte de la lengua zapoteca*, it appeared in 1578.

The importance of Zapotec civilisation in the Mexican context is demonstrated by the existence of Zapotec loanwords in Yucatec (Zapotec *pi'kku* 'dog' and *mani* 'horse, originally deer') and Huastec (Zapotec *pi'kku* 'dog', *picjinja* 'deer, large animal', *taa* 'woven mat', *pisjiicju* 'coat-imundi').

ZHUANG

PERHAPS 11,000,000 SPEAKERS

China, Vietnam

Zhuang is the name for two neighbouring TAI LANGUAGES. The Zhuang count as China's largest minority group, with a population of 13,300,000 (but many of them, probably, speak Chinese rather than Zhuang). Zhuang is officially considered one of the 'major minority' languages of China, alongside Tibetan, Mongolian, Uighur and Korean. Why has hardly anyone outside China heard of it?

The answer is that Zhuang have not, historically, been anxious to project a distinct ethnic identity. They wanted to be Chinese – and Chinese they have largely become.

The older transliteration of the name is *Chuang*. I have also seen *Juang*. The language and its speakers have in the past been called *T'u-jen*, 'people of the soil'.

Southern China is the region from which, it is now supposed, Thai languages originally spread. Thus it is likely that dialects ancestral to Zhuang have been spoken here for over a thousand years, and probably for much longer than that.

Most Zhuang speakers have been bilingual in Chinese for two generations or more. Many have family names which could be argued to be Chinese, and have developed Chinese family histories to go with them. They have continued to speak Zhuang at home and in their own rural communities. Under the Nationalist government this inconveniently large linguistic minority was studiously ignored. Communist China has followed precisely the opposite policy. The large Guangxi Zhuang Autonomous Region was created in 1958, and a Latin orthography was devised in which

books and magazines are regularly published (most of them translations from Chinese). Local radio stations broadcast in Zhuang, and films are dubbed in the language. What was once a little-known language is now widely seen in print, and much studied by Chinese linguists.

Until these recent changes the only way of writing Zhuang was in a local modified form of the Chinese script. This is at least as old as the 18th century, the date of the earliest preserved specimen: how much further it may go back is unknown, but it did spread widely, a similar system being known in THO. It was not used in print, but only for personal notes, trade accounts and letters. There was no standardisation, and it would be difficult to read a text by an unknown writer: 'Zhuang texts are often a farra-go of nonce creations and individual and regional variations' (Ramsey). If Zhuang speakers wrote for general information, they wrote in Chinese. The most widespread use of this older Zhuang script was, however, to record the texts of Zhuang songs – here the exact sounds had to be transmitted, so simply writing in Chinese would not do.

The northern and southern dialect groups of Zhuang are so different from one another that they are considered to be two separate languages. Northern Zhuang has six tones in open syllables, two in closed syllables (those ending in *-p*, *-t*, *-k*). Like Chinese, Zhuang is essentially a monosyllabic language. For a table of numerals see TAI LANGUAGES.

Based on S. Robert Ramsey, *The languages of China* (Princeton: Princeton University Press, 1987) pp. 234–43 and other sources

Central and Northern Tai languages

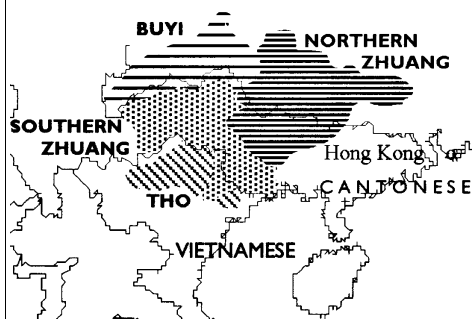
In one accepted classification, Buyi and Northern Zhuang are Northern Tai languages; Southern Zhuang and Tho, with some minor languages, are Central Tai.

BUYI is spoken by a minority population of over 2,000,000 in south-western Guizhou province and around the city of Guiyang. There are a few Buyi-speaking communities in Yunnan, and similar dialects (*Dioi, Giay, Yay*) have been found in Vietnam and Laos.

Northern Zhuang, with perhaps 7,000,000 speakers, includes the dialect of *Wuming*, in central Guangxi, on which standard written and broadcast Zhuang is now based.

Southern Zhuang, with perhaps 4,000,000 speakers, consists of the Tai dialects of south-western Guangxi, close to the Vietnamese border. *Nung, Nung-an* and *Lunghow* are names that linguists have given in the past to varieties of southern Zhuang: 'Nung' is parti-

South-East Asia:
Central and Northern Tai languages



cularly used of the language as spoken in northern Vietnam, where its speakers form an official nationality.

THO is the major Tai language of the northern border regions of Vietnam, where there are about 1,000,000 speakers, forming an official nationality of Vietnam.

ZULU

8,800,000 SPEAKERS

South Africa, Lesotho

One of the BANTU LANGUAGES, Zulu is spoken by the largest language community of South Africa. There are also 200,000 speakers in Lesotho.

Some linguists prefer the prefixed form *isi-Zulu*, literally 'Zulu in language or culture'.

The language group that includes Zulu, Xhosa, Ndebele and Swazi was once called *Kaffir*, Portuguese *Caffre*, a vague word that sometimes denoted all the Bantu languages of southern Africa (it is the Arabic word for 'pagan'). Zulu itself was sometimes called *Zulu-Kaffir*. The group is now usually named *Nguni*, a modern reapplication of a term that once denoted a single constituent element within the Zulu-speaking population. Customarily the variant form *Ngoni* is used for the people who migrated northwards (see map) in the mid 19th century.

It is likely that speakers of languages ancestral to these modern ones were already in Natal in the 9th century, the likely date of the earliest settlements found by archaeologists that are typical of modern 'Nguni' peoples – a central cattle byre surrounded by a row of huts. This marks the beginning of the Later Iron Age in southern Africa.

By the 19th century Nguni language speakers were politically dominant over much of modern South Africa, and at times well beyond its modern borders. Sotho, Pedi (Northern Sotho) and Kgatla (Tswana) speakers had come to share the typical cattle-keeping culture of the Nguni. Parallel with this development, Nguni influence can be traced on their languages.

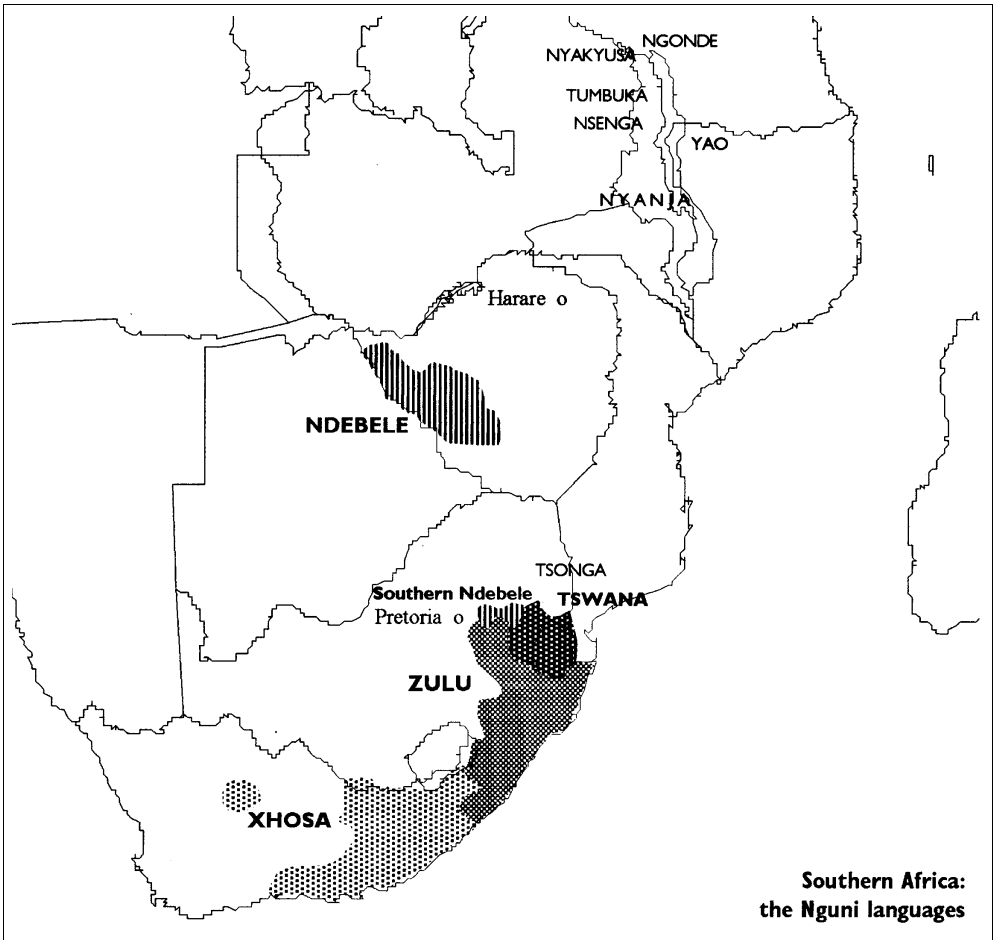
Zulu had previously been influenced in its

turn. Its click consonants originate in loan-words from Khoe or some other KHOISAN language, evidently the result of deep and long-lasting interaction with its speakers. Pastoralism and an associated cattle cult were clearly shared between the two language groups: Zulu words for 'cow', 'sheep' and 'milk' are Khoisan in origin. Seventeen consonants were added to Zulu in the course of this interaction. Zulu (like Xhosa) now has a sound pattern that is most unusual for a Bantu language. In Zulu spelling the three clicks borrowed from Khoisan languages appear as *c* (dental), *q* (palatal), *x* (lateral).

The Zulu empire of Shaka, and the great upheavals of that time, mark a break between earlier and later identifications of peoples – and therefore of languages – in what is now South Africa. Before Shaka's time, the 'Zulu' were simply one clan among many. After the *difaqane* 'forced migrations' of the 1820s and 1830s, the rebellions and famines and population movements in what is now eastern South Africa, a new linguistic map begins to emerge.

Two contemporaries, the Wesleyan John Appleyard and the American missionary Lewis Grout, were the first to make linguistic comparisons of the languages now called Nguni. Grout's work appeared in a paper in volume 1 of the *Journal of the Oriental Society* in 1849: he was also the author of a detailed grammar of Zulu published in 1859.

Both Zulu and Xhosa are official languages in South Africa, and there is television and radio broadcasting in these languages. For the *hloni-pha*, the women's 'avoidance language' of Zulu, and for Zulu numerals, see boxes at XHOSA.



Language of white rule

In the late 19th century a pidgin form of Zulu developed in which Europeans and Indians were able to communicate with Africans. Originating along the Natal coast, 'Fanakalo' spread as far as Rhodesia (modern Zambia and Zimbabwe) with the pioneers who founded Salisbury (Harare) in 1890.

As the mines of the Witwatersrand developed and drew labour from as far away as Malawi and Tanzania, Fanakalo became the lingua franca among migrant workers of different mother tongues, as well as the language of communication between bosses and workers. Employers held classes to teach it formally to new arrivals; but its use is now in steep decline. In the early 20th century it played the same role in the mines of

Zambia and Shaba. There it has now been replaced as an informal lingua franca by TOWN BEMBA or has given way to SWAHILI, languages which have now no overtones of African subjection.

Enza fana-ga-lo, 'do it like this', was how a master began his instructions to a servant. But this language had many forms and names: *Kitchen Kafir*, as used between mistress and cook; *Garden Kafir*; *Mine Kafir* as the working language in the gold and diamond mines; *Pidgin Bantu*; *Basic Bantu*. Africans often called it *silungu-boi*, from Zulu *isilungu* 'white language' and English *boy* 'servant', or else *isilololo*, the 'lo-lo-lo language' because *lo* 'the, this, that' is heard so often in it. In Zambia, it has been called *cikabanga*.

About three-quarters of the words of Fanakalo are Zulu, and about a fifth are English.

Some linguists once thought that Fanakalo would grow into a creole and become the majority language of the black population of South Africa. With hindsight, this was never likely. It was too closely identified with a set of social circumstances: it was the servant-to-master, employee-to-employer language. It was the whites who stuck to Fanakalo so tenaciously: some of them thought it insolent for blacks to use English. Despised by the Africans who used it, Fanakalo never became anybody's mother tongue.

The Nguni languages

Zulu, *XHOSA*, *SWAZI* and Southern *NDEBELE* form a dialect continuum and a close-knit subgroup within the South-eastern group of Bantu. The subgroup is usually known as *Nguni*. *Southern* or *Transvaal Ndebele* has 800,000 speakers. *Swazi* has 1,500,000 speakers in South Africa and Swaziland. *Xhosa* has 6,900,000 speakers in South Africa.

The peoples called *Ngoni*, now to be found in Mozambique, Malawi, Zambia and Tanzania, originate from a series of epic migrations northwards in the 1820s by Zulu and other Nguni warriors looking for freedom from the rule of Shaka. As they travelled they found young men to incorporate in their 'army', and many married local women: thus, even as they settled and (by 1850) founded petty kingdoms, their language was usually on the way to obsolescence – though it lasted long enough to help European explorers, such as David Livingstone, to make themselves understood in the course of very long journeys in south-eastern Africa. In these regions the spoken languages now are *TSONGA*, *TUMBUKA*, *Nsenga*, *NYANJA*, *Nyakyusa*, *Ngonde* and *YAO*. The true Ngoni or Zulu language does not survive there (except, it is said, in ritual use among Tumbuka- and Nsenga-speaking Ngoni).

The exception is *Northern Ndebele*. This language survived and spread, now numbering 750,000 speakers around Bulawayo in Zimbabwe.

The overtones of Fanakalo

'The "blacks" have to learn it from the "whites" or from their servants – often a somewhat painful process. It is not surprising, therefore, that amongst Africans it is now regarded as a slave-driving jargon, while to many Europeans it is a means of "keeping the Kafir in his place". It denies Africans their tribal status while refusing to admit that they are fit to speak English or Afrikaans.'

Irvine Richardson, 'Some observations on the status of Town Bemba' in *African language studies* vol. 2 (1961) pp. 25–36

English loanwords and Zulu noun prefixes

The *TSONGA* word *xitimela* looks a Bantu word, with a class 7 prefix, yet it is a loan from English *steamer* via Zulu. English and Afrikaans words with initial *s* + consonant are invariably remorphologized and assigned to the seventh class in Zulu, e.g. *isitolo* 'store', *isipunu* 'spoon', *isipinashi* 'spinach', *isipanji* 'sponge'. When travelling northwards into other Bantu languages, these words retained their allegiance to the seventh noun class, and changed the prefix according to the prevailing sound laws, so that the *SHONA* word for 'store' became *chitoro*, no longer recognizable as an English 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, abridged

Iscamtho

Zulu Yini u-zonda izi-nsizwa na?
Iscamtho Why u-zunda ama-jents?

'Why do you hate young men?'
Jents is borrowed from English *gents*.

Iscamtho, language of rebellion

A jargon of urban youth in the black communities of the Johannesburg region, Iscamtho has grown out of *Shalambombo*, the criminal argot of two gang networks, the mainly Zulu amaLaita and the Sotho amaRussia, both of which operated in Johannesburg between 1890 and the 1930s. There are both Zulu-based and Sotho-based versions of the language. It first became established in the largely Zulu squatter communities of Orlando and Pimville, and the Sotho-speaking Eastern Native Township, Newclare and Moroka Emergency Camp.

Iscamtho is a mixture of Zulu and Sotho, in varying proportions, with English and other languages. In comparison with its Bantu origi-

nals it is characterised by vowel loss: *iskule* for Zulu *isikule* 'school'; *Iscamtho*, the language name, for what in Zulu would have been *isiqamtho*, derived from the verb *ukuqamunda* 'talk volubly'.

Among some groups Iscamtho has become an essential marker of status and of masculinity. Fewer women than men use it: those women who do, risk being seen as *isfebe* 'prostitutes' or *i-tiye* 'tea, i.e. shared refreshment'. Iscamtho tends to carry greater prestige than the AFRIKAANS-based Tsotsitaal.

Based on papers by D. K. Ntshangase and others in *Language and social history: studies in South African sociolinguistics* ed. Rajend Mesthrie (Cape Town: David Philip, 1995) with other sources

Glossary

The facts of language are hard to pin into definitions. The words ‘typically’, ‘usually’ should be read into many of the definitions below. Examples in this glossary, unless otherwise stated, are English words or sounds.

accent: may be used to denote a *dialect* that only differs from one’s own in its sound pattern (a ‘British accent’, a ‘West Country accent’). May also be used for symbols that modify letters of the alphabet (thus *é* may be called ‘e with an acute accent’): the technical term for these symbols is *diacritical mark*

accusative: in languages in which nouns have alternative forms (*cases*) depending on their function in the sentence, this is a name for the form that serves as the *object*

affix: a *prefix* or *suffix*

agglutinative: a type of language in which verbs and nouns have a sequence of separately identifiable affixes marking their relationship to one another within a sentence. See *typology* and, for more information, *ALTAIC LANGUAGES*

alphabet: a form of writing based on the principle of a single sign for each successive distinct sound or *phoneme* of a language. Some people define the scripts of *AMHARIC*, *HINDI* and *KOREAN* (and their relatives) as *syllabaries*, because in them each syllable forms a block. They are more often classed as alphabets, because symbols for each phoneme can be identified within the block

alveolar: a consonant formed by the alveolar ridge and the tip of the tongue (for example, *t, d*)

augmentative: a word with an affix indicating largeness, importance

back vowels: formed at the back of the mouth (for example, *o, u*)

bilingual: a bilingual person can communicate in two languages with equal facility; a bilingual text is written in two languages to be read by speakers of either

calque: a word or phrase built on the model of one in another language (for example, French *gratte-ciel* for skyscraper)

case: a series of alternative forms of a noun serving to mark its function within a sentence. Different names have been chosen by grammarians working on different languages, but *nominative*, *accusative* and *possessive* often occur

compound: a word formed from two or more separately identifiable words (*birdsong*)

concord: marking of more than one word in a sentence, in a pattern that depends on their relationship with one another. See *KONGO* for more information

consonant: in spoken language, *vowels* alternate with consonants or consonant groups. Consonants are the sounds formed by the momentary narrowing or stopping of the air stream

creole: a language of mixed origins (such as a *pidgin*) which has begun to be used as a *mother tongue*

dental: a consonant formed by the teeth and the tip of the tongue (for example, the sound *th*)

diacritical mark: a symbol (or ‘accent’) used to modify the letters of an alphabet

dialect: local, relatively uniform, varieties of speech. ‘A group of dialects that are mutually comprehensible’ is one definition of a *language*

diminutive: a word with an affix indicating smallness, unimportance

diphthong: a vowel whose pronunciation begins at one position in the mouth and ends at another (*eye, oh, I* are among English words that consist of single diphthongs)

family: a grouping of languages descended from a single earlier language. In this book, *family* is used for the largest such groupings that can be demonstrated; *branch*, *group*, *sub-group* are used for subdivisions of these. For very large groupings the terms *stock* and *phylum* have been adopted by some linguists

feminine: see *gender*

fricative: a consonant creating turbulence but not stopping the air stream (for example, *f*, *v*, *th*, *sh*)

front vowels: formed at the front of the mouth (*e*, *i*)

fusional: a type of language in which words mark their relationship to one another within a sentence by regular modifications which cannot be identified as units separate from the word root. See *typology* and, for more information, ALTAIC LANGUAGES

future: a verb *tense* marking the action as taking place in the future

gender: in many languages, nouns are grouped into two or more *noun classes* which show some parallelism with the male/female classification that we apply to human beings and to animals. This kind of grammatical classification is called 'gender': the most commonly named classes are *masculine*, *feminine*, *neuter*

genetic relationship: languages which can be shown to have differentiated from the same earlier language are said to be *related*. In linguistic terms this is often called a genetic relationship, though it has nothing to do with genes and human relationships

glottochronology: calculation of the elapsed time since two languages diverged, by counting the number of words in a standard list that are still in use in both. The technique has not been shown to work accurately

grammar: the set of rules, adopted more or less unconsciously by any speaker, that produces sentences in a particular language

guṇa: in *fusional* languages, a modification made to the vowel of a word root, in order to form a derived word, may be called by the Sanskrit grammarians' name 'guṇa'. The name for a second modification of the same vowel is *vr̥ddhi*. An example from Sanskrit: root *vid-* 'know', guṇa form *veda* 'knowledge, lore',

vr̥ddhi form *vaidya* 'one who knows, doctor'.

See also MARATHI

ideographic script: a writing system based on the principle of a single sign ('ideogram') for each successive idea – this may mean, roughly but not exactly, one sign per word

infinitive: a verb form which is undefined as to *person* and *number*

intransitive: a verb which, when it forms part of a sentence, has no specified *object*

isolate: a language which cannot be shown to belong to any *family*

isolating: a type of language in which words do not have alternate forms marking their relationship within a sentence. Typically, such relationships are indicated instead by the order of words in the sentence. See *typology*

labial: a consonant formed at the lips (for example, *p*, *b*, *f*, *v*, *w*)

language: may be defined as a group of *dialects* that are mutually comprehensible; may also mean a *standard language* that is more widely understood and accepted than any local dialect

lingua franca: a language used for communication among speakers of several different mother tongues, often in the course of long distance travel and trade. The original Lingua Franca was used in the harbours and seaways of the medieval Mediterranean (see ROMANCE LANGUAGES)

living language: a language which is still used for everyday communication, or which some people still living learnt as their *mother tongue*

loanword: a word borrowed from another language

logogram: a written word borrowed as a whole into the writing system of another language. Some SUMERIAN logograms were borrowed into AKKADIAN and HITTITE writing: the result is that decipherers do not know the Akkadian and Hittite pronunciation for those words. The curly E is a logogram in English: it is the initial letter of Latin *libra*, but it is read as *pound*

masculine: see *gender*

monolingual: able to communicate in one language only

mood: marking of a verb to indicate the cer-

- tainty or otherwise of the activity that it denotes. In English, words such as *may*, *might*, *would* are used for this purpose
- mother tongue:** the dialect or language that a child learns first
- multilingual:** able to communicate in many languages
- national language:** a language that is widely used for communication within a nation state. Its status may be defined legally as 'national language' or '*official language*'. Some governments draw a distinction between these two concepts, but this book does not
- neuter:** neither masculine nor feminine. See *gender*
- nominative:** in languages in which nouns have alternative forms (*cases*) depending on their function in the sentence, this is the usual name for the form that serves as the *subject*
- noun:** a word denoting an object or concept; see *part of speech*
- noun classes:** groupings that regulate the forms of nouns and the forms taken by other words to which these nouns are attached in a sentence. *Gender* is one type of noun classification. Bantu languages (for examples see LOZI and LUBA) have elaborate noun class systems
- number:** marking, typically of verbs and nouns, to indicate the number of items to which they refer. See *singular* and *plural*
- object:** in *syntax*, the recipient of the activity that is denoted by a *transitive* verb. For an example see box at OCCITAN
- official language:** a language which may be used in official contexts, such as parliaments, courts, schools and government offices. Its status is usually defined legally as 'official language' or '*national language*'. Some governments draw a distinction between these two concepts, but this book does not
- parts of speech:** a classification of words according to their potential functions in any sentence. The classification may vary from language to language but will include *verbs* and *nouns*
- past:** a verb *tense* marking the action as having already taken place
- person:** marking, typically of a verb, to indicate the speaker's relationship to it. In the grammar of most European languages, distinctions are made that identify at least *first person singular* ('I', the speaker is the subject), *first person plural* ('we', the speaker belongs to the group that is the subject), *second person* ('you', the speaker is addressing the subject), *third person* ('he, she, it, they', the subject is neither speaking nor being addressed)
- phonemes:** the distinctive sounds of which the sound pattern of any language is made up
- phylum:** see *family*
- pidgin:** a simplified language developed in regular limited contact between people of different mother tongues, and combining elements from these
- plural:** in languages in which a grammatical distinction of *number* is marked, the usual distinction is between *singular* (one) and plural (more than one). English makes this distinction: *dog* (singular); *dogs* (plural)
- possessive:** in languages in which nouns have alternative forms (*cases*) depending on their function in the sentence, this is the usual name for the form that denotes a possessor or owner. In English, nouns have possessive cases, singular and plural; they sound identical but the distinction is marked by the position of an apostrophe: *the dog's bowl* (the bowl belongs to one dog); *the dogs' bowl* (the bowl belongs to more than one dog)
- prefix:** an addition to the *root* or stem, at the beginning of a word
- present:** a verb *tense* marking the action as taking place at the moment of speech
- proto-language:** a language postulated as the ancestor of one or more groups of later languages: thus proto-Indo-European is defined as the early language from which all the modern INDO-EUROPEAN LANGUAGES have descended. Often there is no other evidence for a proto-language than the modern languages themselves; it can be reconstructed, up to a point, by comparing the modern forms
- reconstruction:** words and forms of an earlier language can be reconstructed by comparing the later languages that descend from it. In linguists' jargon, such reconstructions are called 'starred forms' because they are marked

with an asterisk to indicate that there is no textual evidence for them

related: see *genetic relationship*

retroflex: a consonant formed by turning the tip of the tongue upwards

root: the basic form of a word from which other forms derive. The concept applies to word forms in *agglutinative* and *fusional* languages but not in *isolating* languages. See AFROASIATIC LANGUAGES for more information

rounded: a sound formed by rounding the lips (for example, *o*, *u*, and German *ö*, *ü*)

rule: the elements that make up the grammar of a language

script: a writing system – an *alphabet*, *syllabary* or *ideographic script*

sign languages: of all the languages not dealt with in this book, sign languages are perhaps the most significant culturally. Their principal use is for communication among those unable to speak or to hear speech. Their grammar has a complexity similar to that of spoken languages, and they change historically in similar ways

singular: in languages in which a grammatical distinction of *number* is marked, the usual distinction is between singular (one) and plural (more than one). See example at *plural*

sociolect: a variety of a language that is typical not of a geographical region (see *dialect*) but of a social class or group

standard language: the variety or *dialect* of a language that is generally accepted in formal speech and in the media, and is usually taught in schools

stock: see *family*

stop: a consonant formed by blocking the air stream momentarily (for example, *p*, *b*, *t*, *d*, *ch*, *k*, *g*)

subject: in *syntax*, the performer of the activity that is denoted by the *verb*. For an example see box at OCCITAN

suffix: an addition to the *root* or stem, at the end of a word

syllabary: a form of writing based on the principle of a sign for each successive syllable of a language

syllable: the unit of sound formed by a vowel,

any preceding consonants, and sometimes following consonants

syntax: study and description of the structure of sentences

tense: the traditional grammatical term for verb markings to indicate the time, relative to the moment of speech, at which the activity takes place. In many languages, *past*, *present* and *future* tenses are distinguished, but the detailed use of tenses varies a great deal from one language to another

tonal: a language in which *tones* are essential constituents of the sound pattern of each word

tone: the relative pitch and sound quality of vowels

transitive: a verb which, when it forms part of a grammatical sentence, has a specified *object*

typology: a classification of languages depending on whether alternate forms of words are used to mark relationships within a sentence. The three classes are *isolating*, *fusional* and *agglutinative*. For more information see ALTAIC LANGUAGES

unrounded: a sound formed without rounding the lips (for example, *a*, *e*, *i*)

unvoiced: a sound spoken without vibration of the vocal cords. In English, *p t k f sh ch h* are unvoiced

velar: a consonant formed between the velum and the back of the tongue (for example, *k*, *g*)

verb: in *syntax* and as a *part of speech*, a verb denotes an activity, and may define the relation of other words in the sentence to that activity

voiced: a sound spoken with vibration of the vocal cords. In English, *b d g v z r w y* and all vowels are voiced

vowel: in spoken language, vowels alternate with *consonants* or consonant groups. Vowels are the continuous sounds formed when the air stream is open; they are differentiated by the shaping of the mouth and lips

vowel harmony: a rule existing in many languages by which vowels within a word must resemble one another, for example being all *front* or all *back* vowels, all *rounded* or all *unrounded*

vɾddhi: see *guṇa*

Index

- Abaza (Caucasian) **1**
Abenaki (Algonquian) 17
Abkhaz (Caucasian) **1**, 136, 216, 419
Abor-Miri-Dafila (North Assam) 562
Abron (Akan) 10
Abung (Austronesian) 50
Abyssinia 23
Abzakh (Circassian) 1
Abzhui (Caucasian) 1
Achehnese **2**, 48, 49, 50, 118, 119
Acholi (Luo) 370, 454
Ada (Ga) 208
Adamawa-Ubangi (Niger-Congo) 449-50
Adhola (Nilo-Saharan) 454
Adi (North Assam) 562
Adja (Gde) 186
Adongo (Luo) 370
Adyge (Circassian) 1, 136
Aeolic 232
Afar (Cushitic) **3**, 142, 143
Afghanistan 65, 266, 267, 525
 Arabic 2-8
 Brahui 96
 Dari 494-6
 Kurdish 340
 Moghol 426
 Pashto 492-3
 Turkmen 655
 Uzbek 665
Afrikaans **4-5**, 159, 169, 554
Afroasiatic **6-7**, 88, 117, 142, 165, 475, 477, 546, 570
Agam (Malayic) 391
Agar (Dinka) 149
Agew (Cushitic) 142
Agona (Akan) 10
Aguaruna (Amerind) 21
Agul (Caucasian) 109, 110
Agw (Cushitic) 143
Ahanta (Niger-Congo) 452
Ahikuntakayas (Sinhala) 560
Ahom **8**, 41, 553, 605
Ainu (*Japan*) 288
Ajuran (Oromo) 478
Aka (Pygmy) 68
Akan **9-10**, 185, 243, 286, 331, 450
Akassa (Ijo) 263
Akha *see* Hani
Akhty (Lezghian) 354
Akhvakh (Caucasian) 53, 109
Akkadian **6**, **11-12**, 156, 253, 495, 496, 546, 587
Aklanon (Bisayan) 50, 112
Akuapem (Akan) 9, 10
Akusha (Dargwa) 148
Akwapim (Akan) 10
Akyem (Akan) 10
Alaba (Cushitic) 143
Alan (Ossete) 479
Alas (Batak) 79
Alaska 171, 181, 276-7, 434, 525
Albania 39, 232, 519, 525
Albanian **13-15**
Alemannic 222
Aleut 180-1
Algeria 26, 89, 295, 511, 607, 609, 614
Algonquian 16-17, **16-17**, 21, 22, 139
Alor Island 50
Altai (Turkic) 648, 649
Altaic **18**, 134, 287, 328, 424, 641, 647
Alur (Kuo) 370, 454
Ambon Malay 390, 392
Ambulas (Papuan) 491
Amdo (Tibetan) 627
American English 5, 77, 170-1, 179
Amerind 16, **19-22**, 34, 57, 123, 125, 410, 421, 436, 470, 505, 644, 697
Amharic **23-4**, 184, 370, 629
Amis (Formosan) 197
Ana (Niger-Congo) 451
Anatolian 252
Anatri (Chuvash) 135
Ancheya (Nyanja) 467
Andalusian (Spanish) 580, 582
Andaman Islands 46
Andi (Caucasian) 53, 109
Andoni (Ibibio) 164
Anfillo (Gonga) 228
Angami (Naga) 334, 335
Angas (Chadic) 243
Angkola-Mandailing (Batak) 79
Anglo (Gbe) 186
Anglo-Norman 513
Anglo-Romani 516
Anglo-Saxon 166
Angola 70, 71, 72, 73, 315, 501, 525
 Chokwe 132, 133
 Lozi 363-4

- Lunda 367-8
 Luvale 373
 Mbundu 412-13
 Annang (Ibibio) 164
 Antillais 203
 Antsukh (Avar) 53
 Anyi (Baule) 9, 80, 450
 Anyin (Niger-Congo) 450
 Anywa (Luo) 369, 370
 Ao (Konyak) 334
 Arabic **25-30**, 455
 Afar 3
 Azerbaijan 59
 Berber 88, 89
 Cambodia 118
 Chechnya 121
 Dargwa 148
 Fulani 207
 Javanese 290
 Kabyle 295
 Lezghian 354
 Luri 371
 Maba 377
 Malagasy 387
 Maltese 394
 Mandekan 395, 397
 Mozarabic 430
 Persian 495
 Portuguese 500
 Saharan Africa 302
 Somali 570-1
 Songhay 572
 Soninke 574
 Spanish 579
 Sundanese 589
 Susu 590
 Swahili 591-2, 593
 Tamazight 609
 Tashelhet 614
 Temne 619
 Tigrinya 629
 Zande 696
 Aragonese (Spanish) 513, 580, 582
 Arakanese 105
 Aramaic 12, 25, **31-3**, 426, 494, 657, 688
 Arapaho (Algonquian) 17
 Araucanian 21, **34**, 506, 507
 Arawakan (Amerind) 21
 Archi (Caucasian) 109, 110
 Ardhamagadhi **35**, 265
Argentina 581, 582, 583
 Araucanian 34
 Italian 284, 582
 Quechua 505
 Welsh 675
Armenia 272, 339
 Armenian **36-8**, 109, 275, 515, 652
 Aromunian **39-40**, 514, 519, 564
Aruba 489
 Arusi (Oromo) 478
 Asante (Akan) 10
 Ashkhar (Caucasian) 1
 Aslian (Austroasiatic) 45
 Asmat (Papuan) 491
Assam 8, 293, 605
 Assamese **41-2**, 87, 94, 266, 334, 399
 Assaorta (Cushitic) 3
 Assyrian 12, 32
 Asu (Bantu) 70, 593
 Athapaskan (Na-Dené) 434, 435
 Atikamek (Cree) 16, 17
 Atlantic (Niger-Congo) 448-9
 Atoni (Austronesian) 50
 Atsi (Burmese-Lolo) 293
 Atsina (Algonquian) 17
 Attic 232, 233
 Attié (Niger-Congo) 450
 Aukaans (Bush Negro) 178
 Aukstaiciai (Lithuanian) 64
Australia **43-4**, 173-4, 179, 384, 533
 Papua New Guinea 633
Austria 144, 219, 549, 567
 Austroasiatic **45-6**, 52, 311, 313, 431, 541, 668, 670
 Austro-Tai 47, **52**, 296, 416, 605
 Avar (Caucasian) **53-4**, 109, 110
 Avestan **55-6**, 278, 279, 494
 Awaba 43
 Awadhi (Hindi) 249, 250
 Awsanian (South Arabian) 29
 Awutu (Niger-Congo) 450
 Aymara (Amerind) 21, **57**, 506, 507
 Aymellel (Semitic) 24
 Azari (Iranian) 59, 226, 496
Azerbaijan 37, 38, 53, 58, 110, 648, 653, 654
 Gilaki 226
 Kurdish 340
 Lezghian 354
 Russian 526
 Azeri 29, **58-60**, 109, 110, 226, 337, 354, 496, 647,
 648, 651, 654
Azores 501
 Aztec *see* Nahuatl

 Baba Malay 391
 Babu English 178
Babylon 12, 31
 Babylonian Aramaic 32
 Bactrian 278, 569
 Badaga (Kannada) 300
 Bade (Chadic) 243
 Badinani (Kurdish) 340
 Baga (Temne) 619
 Bagirmi (Nilo-Saharan) 302
 Bagulal (Caucasian) 53, 109
Bahamas 176
 Bahnar (Austroasiatic) 45, 46
Bahrain 27
 Bai **61**, 562, 606
 Bajalani (Gurani) 340
 Bajau (Austronesian) 50, 51
 Bakhtiar (Luri) 372
 Bakoko (Bantu) 70
 Balanta (Niger-Congo) 450
Bali 50

- Balinese 50, **62-3**, 290, 380, 545
 Balkar (Turkic) 337-8, 648
 Balti (Tibetan) 628
 Baltic **64**, 224, 274, 275, 353, 360, 530, 564
 Balto-Finnic 182
 Balto-Slavic 564
 Baluchi **65-6**, 279
 Bamana (Senufo) 452
 Bambara (Mandekan) 395, 396, 397, 398
 Bamendjou (Niger-Congo) 450
 Bamun (Niger-Congo) 450
 Banaphari (Hindi) 249
 Banda (Niger-Congo) **67**, 215, 450, 537, 696
 Bangala (Bantu) 71
 Banggai (Austronesian) 50
 Bangi (Bantu) 358
Bangladesh 42, 46, 86-7, 137, 541
 Bangubangu (Bantu) 70, 365, 366
 Banjar (Malayic) 390, 391, 392
 Banjari (*India*) 266
 Barbadian Creole 176
 Bari (Nilotic) 454, 455, 645
 Bariba (Niger-Congo) 450
 Baric (Tibeto-Burman) 562
 Barikanci (Hausa pidgin) 243
 Barito 257, 387, 388
 Barracoon (Swahili pidgin) 592-3
 Basaa (Bantu) 70
 Bashgali (Nuristani) 267, 268
 Bashkir **75**, 324, 658
 Basketto (Omotic) 228
 Basque **76-7**, 88, 199
 Bassa (Kru) 332, 449, 450
 Batak (Austronesian) 49, 50, **78-9**
 Bats (Nakh) 122
 Bauan (Fijian) 191
 Baule (Niger-Congo) 9, **80**, 450, 451
 Bazaar Malay 390
 Beach-la-mar (English pidgin) 93, 179, 633
 Beja (Cushitic) 3, **81**, 142, 143
Belarus 498, 525, 689, 690
 see also Belorussian
Belgium 158, 159, 201, 522
Belize 21, 176, 411, 694
 Belorussian **82-3**, 361, 528, 564
 see also Belarus
 Bemba 70, **84-5**, 701
 Bembe of Congo (Bantu) 70
 Bena (Bantu) 70
 Bengali 41, 42, **86-7**, 90, 266
 Arabic 29
 Khasi 311
 Magahi 381
 Oriya 476
 Benguela Mbundu (Bantu) 72, 412, 413
Benin 185, 206, 242, 427, 449, 450, 451, 452, 455, 572, 692
 Benue-Congo 68, 450, 464
 Bera (Bantu) 70
 Berber 6, 7, 25, 27, 29, **88-9**, 295, 579, 607, 608, 614
 Berbice (Dutch) 160
 Berta (Nilo-Saharan) 454
 Bété (Kru) 332, 449, 450
 Beti (Bantu) 188
 Bezgheta (Caucasian) 53, 109
 Bhaṭṭeālī (Dogri) 487
 Bhatrī (Assamese) 42
 Bhattiānī (Panjabi) 487
 Bhīlī (*Gujarat*) 238
 Bhojpuri **90-1**, 176, 191, 250, 266, 383
 Bhumij (Austroasiatic) 45, 46, 431
Bhutan 161, 626, 627
 Biak (Austronesian) 50
 Bihari *see* Bhojpuri; Magahi; Maithili
 Bikol (Austronesian) 49, 50, **92**, 601
 Bilin (Cushitic) 143
 Bima (Austronesian) 50
 Bimoba (Niger-Congo) 451
 Bingkokak (Austronesian) 50
Bioko Island 69, 188
 Bira (Bantu) 70
 Bird's Head (Papuan) 50, 621
 Birifor (Niger-Congo) 450
 Birom (Niger-Congo) 450
 Bisa (Bemba) 71, 85
 Bisayan 111, 247
 Bishnupriya Manipuri (Bengali) 86, 87
 Bislama **93**, 179
 Blaauw (Austronesian) 50
 Black Carib 21
 Black English 171, 176
 Blackfoot (Algonquian) 16, 17
 Black Tai 349, 605
 Bobangi (Bantu) 358
 Bobo Fing (Niger-Congo) 450
 Bobo Wule (Niger-Congo) 450
 Bodie 562
 Bodo-Garo 41, **94-5**, 562
 Bohairic (Coptic) 137
Bohemia 144, 222, 691
 Bokmål (Norwegian) 459
 Boko (Niger-Congo) 450
 Bolaang Mongondow (Austronesian) 50
Bolivia 57, 505-7, 581
Bonaire 489
 Boomu (Niger-Congo) 450
 Bor (Dinka) 149
 Borena (Oromo) 478
 Borgu (Niger-Congo) 450
Borneo 50, 51, 99, 111, 112, 539
 Boşa (Gypsy) 517
 Bosnian **548-50**
 Botlikh (Caucasian) 53, 109
Botswana 637
 Brahui 65, **96**, 155, 156, 341
 Braj (Hindi) 249, 250, 509
Brazil 284, 500, 501, 582, 644
 Breton **97-8**, 113, 115, 199
Britain
 Celtic 113, 114
 Gujarati 237
 Italian 284
 Nyanja 466
 Oriya 476

- Panjabi 487
 Polish 498
 Romani 516
 Sylheti 86
 Tonga 634
 Turkish Cypriots 653
 Ukrainian 660
 Yiddish 689
see also English
 Brong (Akan) 10
 Bru (Austroasiatic) 45, 46
Brunei 390
 Brythonic 114, 115
 Bubi (Bantu) 188
 Budukh (Caucasian) 109, 110
 Bugis (Austronesian) 49, 50, **99-100**, 384
 Bukusu (Bantu) 70, 375, 376
 Bulang (Wa) 671, 672
Bulgaria 473, 474, 525
 Gagauz 653
 Macedonian 378
 Tosk 15
 Bulgarian 14, **101-2**, 134, 550, 564, 568
 Buli (Niger-Congo) 450
 Bullom (Niger-Congo) 450, 619
 Bulu (Bantu) 71, **187-9**
 Bumthang (Tibetan) 161, 627
 Bunak (Timor-Alor-Pantar) 392
 Bura (Chadic) 243
 Bura Mabang *see* Maba
Burkina Faso 89, 206, 449, 450, 451, 452, 455
 Mandekan 396, 397
 Moore 427
 Songhay 572
 Tamasheq 607
Burma 539
 Hani 241
 Jinghpaw 292-3
 Karen 304-5
 Khün 346
 Kuki-Chin 333-5
 Lisu 359
 Mon 423
 Palaung 46
 Shan 552-3
 Tibeto-Burman 562
 Wa 670-2
 Burmese 61, **103-6**, 293, 423, 562
 Burmese-Lolo 104-5, 344, 562, 686
 Burun (Luo) 370
Burundi 72, 158, 522, 593
 Burushaski (*Kashmir*) 155, 156
 Buryat (Mongolian) 425, 426
 Busa (Niger-Congo) 450
 Bushman 315, 316
 Bush Negro 177-8
 Butler-English 178
Buton Island 51
 Buyi **107**, 605, 699
 Bwaanaba (Kiribati) 407
 Bwamu (Niger-Congo) 450
 Bzhedugh (Circassian) 1
 Bzyp' (Caucasian) 1
 Cafreal de Sena (Sena) 547
 Caipira (Portuguese creole) 501
 Cajun (French) 202
 Čakavian (Serbo-Croat) 549
 Cakchiquel (Mayan) 21, 410, 411, 437
 Caló (Romani) 516-17, 583
Cambodia 50, 539
 Cham 118-19
 Khmer 46, 313-14
 Malay 391
 Stieng 46
Cameroun 68, 69, 70, 71, 117, 450, 451, 452, 454
 Duala 157
 Ewondo (Yaoundé) 187, 188
 Fulani 207
 Gbaya 215
 German 221
 Hausa 242
 Kamtok 299
 Campidanese (Sardinian) 285
Canada
 Algonquian 16-17
 Basque pidgin 76-7
 Canadian English 170
 Cree 17, 139-40
 French 202
 Gaelic 209
 Inuit 276-7
 Laurentian 123
 Sinhala 559
Canary Islands 88
 Cantonese *see* Yue (Cantonese)
 Cape Bushman 316
Cape Verde Islands 141, 501
 Capiznon (Bisayan) 50, 112, 247
 Car (Austroasiatic) 45, 46
 Carib (Amerind) 21
Caribbean 91, 160, 171, 176, 203-4, 489, 582
 Carrier (Athapaskan) 434
 Caspian 226
 Castilian 513, 580
 Catalan **108**, 285, 469, 513, 514, 543, 580-2
 Caucasian 1, 36, 53, **109-10**, 216, 354, 479
Caucasus 121-2, 136, 525
 Cayuga (Iroquoian) 123
 Cebaara (Senufo) 450, 452
 Cebuano (Austronesian) 49, 50, **111-12**
Celebes see Sulawesi
 Celtic 97, **113-16**, 198, 209, 224, 275, 280, 674
 English 168-9
 Latin 350
Central African Republic 67, 68, 71, 243
 Gbaya 215
 Lingala 358
 Niger-Congo 450, 451, 452
 Sango 536-7
 Zande 696
 Cewa (Nyanja) 467, 555
 Chabacano (Spanish creole) 583
Chad 26, 28, 206, 242, 453, 454, 455, 456

- Kanuri 302
 Maba 377
 Sara 542
 Chadic 6, 7, 28, **117**, 242, 243
 Chagai-Kharani (Baluchi) 65
 Chagatai (Uzbek) 308, 309, 665, 666
 Chagga (Bantu) 70, 415
 Chakma (*India; Bangladesh*) 42
 Cham (Austronesian) 2, 29, 49, 50, **118-19**, 391
 Chamalal (Caucasian) 53, 109
 Chamitic *see* Afroasiatic
 Chamorro 49, **120**
Channel Islands 201-2
 Chaoui (Berber) 89
 Chaplino (Eskimo) 181
 Charoda (Avar) 53
 Chasu (Bantu) 593
 Chechen 29, **121-2**
Chechnya 526
 Cherkess *see* Circassian
 Cherokee 21, **123-4**
 Chewa (Nyanja) 467
 Cheyenne (Algonquian) 16, 17
 Chhattisgarhi **125**, 250, 266
 Ch'iang (Tibeto-Burman) 562
 Chibchan (Amerind) 21
 Chicano (Spanish American argot) 583
 Chichewa (Bantu) 72
Chile 34, 284, 581
 Chimbu (Papuan) 491
 Chin *see* Kuki-Chin
China 47, 525
 Aceh-Chamic 118
 Bai 61
 Buyi 107
 Dong 154
 Hani 241
 Jinghpaw 293
 Kadai 296
 Kazakh 308
 Khotanese 317
 Korean 329
 Lahu 344
 Li 355-6
 Lisu 359
 Lü 346
 Manchu 641-3
 Miao 416-17
 Mongolian 424-5
 Naxi 441
 Palaung 46
 Salar 648
 Sariqoli 279
 Shan 552-3
 Tai 605-6
 Tangut 613
 Tibetan 626
 Tibeto-Burman 562
 Tocharian 632
 Tujia 105
 Uighur 657-8
 Wa 670-2
 Yao 685
 Yi 686-7
 Zhuang 698-9
 see also Chinese
 China Coast Pidgin 178
 Chinantec (Otomanguan) 422
 Chinese 52, **126-31**, 289, 561
 Fiji 191
 Japanese 287
 Korean 329
 Malaysia 390
 Persian 495
 Taiwan 196
 Viet-Muong 45
 Vietnamese 668-9
 see also *China*
 Chinook Jargon (Amerind) 20, 21
 Chipewyan (Atheaskan) 434
 Chiriguano (Amerind) 21
 Chisena (Sena) 547
 Chokwe (Bantu) 70, **132-3**, 367, 373, 412
 Chol (Mayan) 410, 694
 Choni (Tibetan) 627
 Chontal (Mayan) 410, 694
 Chopi (Bantu) 70
 Chorasmian 278, 569
 Chru (Aceh-Chamic) 119
 Chukchi (Chukotko-Kamchatkan) 481
 Chukotko-Kamchatkan 481
 Church Slavonic 82, 473
 Chuvash 75, **134-5**, 404, 647
 Chwabo (Bantu) 70, 386, 466
 ciLuba (Bantu) 365
 Circassian 1, **136**
 Ciro (Luo) 370
 Citak (Papuan) 491
 Cockney 171
 Cocoliche (Italian-Spanish pidgin) 582
Colombia 21, 505, 581
 Comanche (Uto-Aztecan) 438, 439
Comores 70, 204
 Comorian (Bantu) 70, 594-5
Congo (Brazzaville) 69, 71, 73, 188, 322, 357
Congo (Kinshasa) 69, 70, 71, 72, 73, 453, 454, 455
 Banda 67
 Bemba 85
 Chokwe 132-3
 Kituba 322
 Lingala 357-8
 Luba 365-6
 Lunda 367-8
 Luvale 373
 Ngbandi 537
 Swahili 592
 Zande 696
 Cook Islands Maori 401
 Coorg (Kannada) 301
 Coptic 25, 27, **137-8**, 165
 Cora (*Mexico*) 437
 Cornish 97, 113, 115
 Corsican 285, 543-4
 Cree (Algonquian) 17, **139-40**

- Crete* 230
 Crimean Gothic 229
 Crioulo **141**, 501
Croatia 474
 Croatian **548-50**, 564, 567, 568
 Cuabo (Bantu) 70, 386
Cuba 525, 581, 582
 Cuicatec (Mixtec) 422
 Cuna 21
Curacao 160, 489
 Cushitic 3, 6, 7, 23, **142-3**, 477, 570
 Cutchi-Swahili 592
 Cuyonon (Bisayan) 112
Cyprus 28, 230, 233, 651, 652-3
 Czech **144-5**, 564, 566
Czechoslovakia, former 222, 525

 Dagaari (Niger-Congo) 449, 450
Dagestan 53, 58, 59, 148, 354
 Dagestanian 109-10
 Daiiri (Batak) 79
 Daju (Nilo-Saharan) 454
 Dan (Niger-Congo) 450
 Danao 382
 Dandawa (Songhay) 572
 Dangme (Niger-Congo) 208
 Dani (Papuan) 491
 Danish **146-7**, 190, 225, 459-60, 599
 see also Denmark
 Darasa (Cushitic) 143
 Dardic 267, 306
 Dargwa (Caucasian) 53, 110, **148**, 337
 Dari (Persian) 492, 493, **494-6**
 Datoga (Nilo-Saharan) 454
 Davaweño (Austronesian) 50
 Dead Sea Scrolls 31
 Dendi (Songhay) 572
Denmark 471, 598
 see also Danish
 Dhangar (Kurukh) 341
 Dharuk (*New South Wales*) 43
 Dialecto Fronterizo (*Brazil-Uruguay*) 582
 Dibo (Nupe) 465
 Dida (Bété) 332
 Dido (Caucasian) 53, 110
 Digor (Ossete) 479
 Dingal (Rajasthani) 509
 Dinka (Nilotic) **149**, 454, 455
 Diola (Niger-Congo) **150-1**, 449, 450, 678n
 Ditammari (Niger-Congo) 450
 Divehi (Indo-Aryan) **152**, 266, 560
 Djerma (Songhay) 572, 573
Djibouti 3, 27, 143, 570-1
 Djonga (Bantu) 636
 Dōābī (Panjabi) 487
 Dobu (Austronesian) 50
 Dogon (*Mali*) 449, 450
 Dogri **153**, 266, 487
 Dolgan (Yakut) 681, 682
 Domari (Indo-Aryan) 515, 516
Dominica 203
 Dong **154**
 Dongolawi (Nubian) 81, 462, 463
 Doric 232
 Dravidian 96, **155-6**, 227, 300-1, 341, 393, 540, 610, 617
 Duala (Bantu) 71, **157**, 188
 Dungan (Chinese) 130
 Dusun (Austronesian) 50
 Dutch **158-60**, 222, 225
 Afrikaans 4-5
 Antilles 489
 Japanese 287
 Sinhala 559
 Suriname 177
 Ternate 620
 see also Frisian
 Dyirbal (*Australia*) 44
 Dyula (Mandekan) 395, 396-7, 398
 Dzhek (Caucasian) 110
 Dzongkha **161**, 627, 628
 Dzorgaish (Tibeto-Burman) 562

Easter Island 49
 East Futuna (Polynesian) 635
 Ebe (Nupe) 465
 Ebira (Niger-Congo) 450
 Eblaite (Akkadian) 11
Ecuador 21, 505-7, 581, 583
 Edo (Niger-Congo) **162-3**, 450, 693
 Efik (Niger-Congo) 157, **164**, 263, 450
Egypt 26, 252, 455
 Coptic 137
 Greek 232
 Nubian 462-3
 Egyptian, Ancient 6, **165**, 546
 Ekari (Papuan) 491
 Ekavian (Serbo-Croat) 550
 Elamite 12, 155, 587
El Salvador 439
 Elwana (Bantu) 594
 Embu (Bantu) 71, 320
 Ende (Austronesian) 50
 Enets (Samoyedic) 534, 535
 Enga (Papuan) 491
 Engadine (Romansch) 521
 English **166-74**, 225, 275
 Afrikaans 5
 Algonquin 17
 Amharic 23
 Arabic 26
 Assamese 41
 Australia 43
 Burmese 104
 Caribbean 203, 204
 Celtic 116
 Chamorro 120
 Fanakalo 701-2
 Fiji 191
 French 198-9, 201
 Frisian 205
 Fulani 207
 German 219-20
 Ghana 9, 243
 Hawaii 244

- India* 248-9, 399
Inuit 276
Ireland 280, 281
Kenya 319
Khasi 311
Latin 351
Luba 365
Luo 369
Malay 391
Malta 394
Micronesia 321, 406, 407
Mon 423
Norse 471
Oriya 476
Papua New Guinea 633
Philippines 600-1
Russian 525, 526
Scotland 209
Seychelles 551
Spanish 580
Sri Lanka 559
Swahili 592, 593
Tumbuka 640
Uganda 213
Vanuatu 93
Welsh 674, 675
Xhosa 679, 680
Yiddish 690
Zambia 84
Zulu 702
see also Britain
 English Creoles and Pidgins 93, **175-9**, 225, 286, 299, 331, 503, 633
Equatorial Guinea 71, 187
Eritrea 23, 24, 27, 28, 454
 Beja 81
 Cushitic 3, 143
 Ethiopic 184
 Italian 284
 Tigrinya 629
Erza see Mordvin (Uralic)
Esan (Edo) 162
Eskimo-Aleut 19, **180-1**, 276
Estonia 222, 471, 525, 527
Estonian **182-3**, 194, 195, 360, 661
Ethiopia 184, 454, 455, 456, 546
 Afar 3
 Amharic 23-4
 Cushitic 142, 143
 Gonga 228
 Luo 369-70
 Nuer 149
 Omotie 475
 Oromo 477-8
 Somali 570-1
 Tigrinya 629
 Turkana 645
 Wolaytta 677
 Ethiopic 23, **184**, 546, 629
Eton (Bantu) 188
Etruscan 350
Etsako see Yekhee (Edo)
Euskara see Basque
Even (Tungusic) 641
Evenki (Tungusic) 641
Ewe (Niger-Congo) 9, **185-6**, 450
Ewondo (Bantu) 71, **187-9**
Eyak (Na-Dené) 434, 435

Fá d'Ambó (Portuguese creole) 501
Falam (Kuki) 333, 334, 335
Falashi (Cushitic) 142
Fanakalo (Zulu) 84, 701-2
Fang (Bantu) 71, **187-9**
Fante (Akan) 9, 10
Faroese 146
Faroese **190**, 225, 460, 471, 599
Farsi see Persian
Fazandeiro (Portuguese creole) 284, 501
Fernando Po 178
Fiji 50, 91, 321
Fijian 49, 50, **191-2**
Filipino 601
Finland 471, 525, 530-2, 598
Finnish 183, **193-5**, 224, 323, 360, 661
Finno-Ugric 18
Flores 50, 51
Floup (Diola) 150, 151
Flytaal (Afrikaans) 4
Fon (Niger-Congo) **185-6**, 450
Fony/Fogny (Diola) 150, 151
Formosan 47, 49, **196-7**
Foulbéré (Fulani) 151
Fox (Algonquian) 17
Frafra (Gur) 449, 450
France
 Basque 76-7
 Breton 97-8
 Celtic 113, 114, 115
 Chad 542
 Dutch 158
 Kabyle 295
 Laos 348
 Mali 396
 Norse (Rouen) 598
 Occitan 108
 Portuguese 503
 Soninke 574
 Turkey 651
 Vietnamese 669
French **198-202**, 275
 Amharic 23
 Arabic 26
 Burundi 522
 Celtic 116, 198
 Central African Republic 67, 536
 Corsican 543
 English 167, 169, 198-9
 French Cree (Mitchif) 17, 20, 139, 202
 Fulani 207
 Italy 283, 284
 Kituba 322
 Latin 352
 Luba 365

- Luxembourg* 374
 Malagasy 387
 Maltese 394
 Mandekan 397
 Moore (Moré) 427
 Muong 433
 Occitan 468
 pidgins 202
 Romanian 519
Rwanda 522, 593
 Susu 590
 Tahitian 603
Vanuatu 93
 French Creoles **203-4**, 240, 551
 Frisian 159, 166, **205**, 225
 see also Dutch
 Friulian (Rhaeto-Romansch) 514, 520, 521
Fujian 127-8
 Fula (Fulani) 151
 Fulani 29, 151, **206-7**, 395, 396, 449, 450, 678
 Fulfulde (Fulani) 151
 Fuliru (Bantu) 71, 523
 Fur (Nilo-Saharan) 454
 Futhark (Runic) 472
- Ga (Niger-Congo) 9, **208**, 450, 451
Gabon 69, 71, 188
 Gabra (Oromo) 478
 Gaelic 114, 115, **209-10**, 280
 Gagauz (Turkic) 653
 Galela (Papuan) 620, 621
 Galician 108, **211**, 500, 513, 514, 581, 582
 Galla *see* Oromo (Cushitic)
 Gallurese (Sardinian) 285
 Galoli (Timor-Flores) 392
Gambia 206, 207, 331, 396, 397, 678
 Gà (Chinese) 127, 128, 129
 Ganda (Bantu) 71, **212-13**, 375, 457, 622
 Garhwali (Pahari) **214**, 267, 336
 Garifuna (Amerind) 21
 Garo **94-5**, 562
 Gaulish 114
 Gayo (Austronesian) 50
Gazelle Peninsula 51
 Gbagyi (Niger-Congo) 451
 Gbaya (Niger-Congo) 67, **215**, 450, 451, 537, 696
 Gbe **185-6**
 Gbedye (Nupe) 465
 Ge (Gbe) 186
 Gedeo (Cushitic) 142, 143
 Geechee (English Creole) 176
 Geg (Albanian) 14-15
 Gek (Caucasian) 110
 Geordie 171
Georgia 1, 32, 58, 232, 419, 479
 Georgian **216-17**, 419, 420
 Ge-Pano-Cariban (Amerind) 21
 German 169, **218-23**, 225
 Estonian 182
 Hungarian 256
 Lithuanian 361
 Low German 159
- Luxembourg* 374
 Romanian 519
 Sami 530
 Swedish 598
 Yiddish 688, 690
 see also *Germany*
 Germanic 4, 158, 166, 205, 218, **224-5**, 274, 275,
 471, 530, 688
Germany
 Celtic 113
 Frisian 205
 Papua New Guinea 633
 Sorbian 575
 Yiddish 691
 see also German
Ghana 206, 208, 450, 451, 452
 Akan 9
 Baule 80
 Dyula 397
 Gbe 185
 Gur 449
 Hausa 242, 243
 Moore 427
 Ghomala' (Niger-Congo) 451
 Ghulfan (Nubian) 463
 Gidatl (Avar) 63
 Gilaki (Iranian) **226**, 279
Gilbert Islands (Kiribati) 321
 Gilyak (Palaeosiberian) 481
 Ginukh (Caucasian) 53, 110
 Gisu (Bantu) 72, 376
 gLo-skad (Tibetan) 628
Goa 327, 403, 502
 Godoberi (Caucasian) 53, 110
 Gogo (Bantu) 71, 594
 Goidelic *see* Gaelic
 Gokana (Niger-Congo) 451
 Goldi (Tungusic) 641
 Göndi 156, **227**, 403
 Gongga (Omotic) **228**, 475, 477
 Gonja (Niger-Congo) 451
 Gorgani (Mazandarani) 226
 Gorkha/Gorkhali (Indo-Aryan) 444, 445
 Gorontalo (Austronesian) 50
 Gothic **229**, 530
 Gourmantché (Gur) 449, 451
 Great Nicobarese 45
 Grebo (Kru) 332, 449, 451
Greece
 Albanian 15
 Aromunian 39, 519
 Greek **230-4**, 275, 350-1, 494
 African languages 27, 88
 Albanian 13, 14
 English 169
 Macedonia 378-9
 Romanian 519
 Sanskrit 539
 Swedish 598
Greenland 146, 181, 276
 Gû (Gbe) 185, 186
 Gua (Niger-Congo) 451

- Guadeloupe* 203
 Guajiro (Amerind) 21
Guam 49, 120
 Guanche 88
 Guaraní 21, **235-6**
Guatemala 410, 411, 508
 Guaymí 21
 Guéré (Kru) 332, 451
Guiné 141, 207, 396, 450, 451, 501
Guinea 202, 206, 396, 590, 619
 Gujarati 56, **237-8**, 266, 556, 663
 Gujuri (Rajasthani) 238
 Gullah (English creole) 176
 Gumbo (Louisiana Creole) 203
 Gumuz (Nilo-Saharan) 453, 454
Gur (West Africa) 449
 Gurage (Semitic) 24
 Gurani (*Iran*) 340
 Gurenne (Gur) 449, 450
 Gurkhali (Nepali) 444
 Gurma (Gur) 449, 451
 Guro (Niger-Congo) 451
 Gurreh (Oromo) 478
 Gurung (Bodic) 562
 Gusii (Bantu) 71, **239**, 376
 Guugu Yimidhirr (*Australia*) 43
Guyana 91, 160, 171, 176
 Gwere (Bantu) 71, 213
 Gypsies *see* Romani
- Ha (Bantu) 72, **522-3**
 Hadhramaut (Arabic) 25
 Hadimu (Swahili) 595
 Hadiyya (Cushitic) 142, 143
 Hadrani (South Arabian) 29
 Hadu (Saho) 3
 Hadza (*Tanzania*) 316
 Haida (Na-Dené?) 434, 435
Hainan 118, 119
 Haitien 203, 204, **240**
 Haka (Kuki-Chin) 334
 Hakka *see* Kejia (Chinese)
 Halbdeutsch 222
 Halbi (*Madyha Pradesh*) 402-3
 Hallam (Kuki) 333
 Hamitic *see* Afroasiatic
 Hamito-Semitic *see* Afroasiatic
 Hamtai (Papuan) 491
 Han (Chinese) 669
 Hani *or* Akha 104, **241**, 344, 552, 562
 Harari (Semitic) 24
 Haroi (Aceh-Chamic) 119
 Hassaniya (Arabic) 28, 29
 Hattam (Papuan) 621
 Hausa 9, 29, 30, 117, **242-3**, 465, 572
Hawaii 171, 179, 533
 Hawaiian 49, **244**
 Hawuon (Austronesian) 51
 Haya (Bantu) 72, 457, 458
 Hebrew 11, 31, **245-6**, 546, 688, 690, 691
 Hehe (Bantu) 71
 Hema-Nord/Sud (Bantu) 458
- Hemba (Bantu) 365, 366
 Herero (Bantu) 71, 413
 Hiligaynon (Bisayan) 49, 50, 111, 112, **247**
 Hindi 90, 125, 227, **248-50**, 266, 336, 381, 383, 486, 509, 663, 664
 Hindko (Panjabi) 487
 Hindustani *see* Urdu
 Hinukh (Caucasian) 53, 110
 Hiri Motu (Austronesian) 50, **251**, 490, 633
 Hittite 12, **252-3**, 271, 274, 587
 Hmar (Kuki-Chin) 334
 Ho (Austroasiatic) 45, 46, 431
 Hokan (Amerind) 21
 Hokkien *see* Min (Cantonese)
 Holu (Bantu) 72
Hong Kong 130, 178
 Hopi (Uto-Aztecan) 438, 439
 Hororo (Bantu) 458
 Hre (Austroasiatic) 45, 46
 Huastec (Mayan) 410, 437
 Huichol (*Mexico*) 437
 Huihui (Aceh-Chamic) 119
 Huilliche (Araucanian) 507
 Huli (Papuan) 491
 Hunanese *see* Xiang (Chinese)
 Hunanoo (*Mindoro*) 111
 Hunde (Bantu) 71
 Hungarian 14, **254-6**, 566, 661
Hungary 525, 549, 567, 691
 Hunnish 649
 Hunzal (Caucasian) 110
 Huron 123
 Hurrian 12, 109, 252
 Hyperborean *see* Palaeosiberian
- Iban (Malayic) 49, 50, **257-8**, 389, 391, 392
 Ibanag (Austronesian) 50
 Ibani (Ijo) 263
 Ibibio (Niger-Congo) **164**, 450
 Ibo 242
Iceland 146
 Icelandic 225, **259**, 460, 471, 599
 Idakho (Bantu) 71, 375, 376
 Idoma (Niger-Congo) 451
 Ife (Niger-Congo) 451
 Ifugao (Austronesian) 50
 Igala (Niger-Congo) 451, 693
 Igbira (Niger-Congo) 450
 Igbo (Niger-Congo) **260-1**, 450, 451
 Igede (Niger-Congo) 451
 Igorots (*Philippines*) 264
 Ijekavian (Serbo-Croat) 550
 Ijo (Niger-Congo) 449, 451
 Ila (Bantu) 85, 634
 Illyrian 224, 275
 Ilocano 49, 50, **264**, 485, 601-2
India 265-6
- Ahom 8
 Ardhamagadhi 35
 Assamese 41-2
 Austroasiatic 431-2
 Bahnar 46

- Bengali 86
 Bhojpuri 90
 Bodo-Garo 94-5
 Chhattisgarhi 125
 Dogri 153
 Dravidian 155-6
 English 173
 English pidgins 178
 Garhwali 214
 Gondi 227
 Hindi 248-50
 Ho 46
Java 290
 Kashmiri 306-7
 Khasi 311-12
 Konkani 327
 Korku 46
 Kuki-Chin-Naga 333-5
 Kumauni 336
 Kurukh 341
 Maithili 383
 Malayalam 393
 Manipuri 399
 Marathi 402-3
 Mundari (Munda) 431
 Nepali 444
 Newari 445
 Oriya 476
 Panjabi 486-8
 Persian 495
 Romani 515
 Sanskrit 538-40
 Santali 541
 Sindhi 556-7
 Sora 46
 Syriac 32
 Tamil 610-12
 Telugu 617-18
 Tulu 639
 Urdu 663-4
 Indo-European 4, 13, 36, 64, 113, 224, 230, 265, **270-5**, 278, 350, 360, 539-40, 564, 632
Indonesia 49, 50, 51, 158, 389-92
 Achehnese 2
 Balinese 62-3
 Batak 78-9
 Bugis 99-100
 Iban 257-8
 Japanese 287
 Javanese 290-1
 Lampung 345
 Madurese 380
 Makasar 384
 Malay/Malaysian 389-92
 Minangkabau 418
 Papuan 490
 Rejang 510
 Sasak 545
 Sundanese 589
 Ternate 620-1
 Indonesian 290, **389-92**, 510
 Ingrian (Karelian) 194-5
 Ingush 121, 122
 Inuit 180, 181, **276-7**
 Ionic 232-3
Iran 27, 278-9, 648, 654
 Aramaic 32
 Armenian 36, 37
 Avestan 55-6
 Azeri 58-9
 Elamite 155, 156
 Gilaki 226
 Kurdish 339, 340
 Turkmen 655
 Iranian 65, 226, 265, 275, **278-9**, 339, 371, 479, 492, 494, 564, 569
 Iranun (Danao) 382
Iraq 27, 546
 Akkadian 6, 11-12
 Aramaic 32
 Kurdish 339, 340
 Mandaean (Gnostics) 32
 Sumerian 587-8
 Turkmen 655
 Iraqw (Cushitic) 142, 143
Ireland 113, 114, 172, **280-2**, 461, 471
 see also Irish
Iran 49, 99, 221, 391, 490, 620, 621
 Irish 115, 209, **280-2**
 see also *Ireland*
 Iron (Ossete) 479
 Iroquoian (Amerind) 21, 123
 Isan *see* Lao
 Iscatho (Sotho/Zulu urban jargon) 4, 702, 703
 Ishan *see* Esan (Edo)
 Isikhwetha (Xhosa) 679
 Isoko (Niger-Congo) 163, 451
Israel 245-6, 294-5, 546, 690
 Isukha (Bantu) 71, 375, 376
 Italian **283-5**, 513, 514
 Amharic 23
 Arabic 26
 Cocoliche 582
 Latin 352
 Maltese 394
 Sardinian 543-4
 Italic 350
Italy
 Albanian 15
 Celtic 113
 Franco-Provençal 108
 French 200
 Greek 232, 233
 Latin 350-2
 Rhaeto-Romansch 520
 Slovene 567
 Yiddish 691
 Itawit (Austronesian) 50
 Itelmen (Chukotko-Kamchatkan) 481
 Itsekiri (Niger-Congo) 451, 693
 Itzá (Mayan) 411
Ivory Coast 206, 450, 451, 452
 Akan 9, 10
 Baule 80

- Dyula (Mandekan) 396-7
 French pidgin 202
 Kru 332, 449
 Moore (Moré) 427
- Jaipuri (Rajasthani) 238
 Jakarta Malay 390, 392
Jamaica 176
 Jamaican Creole **286**
 Janjero (Omoti) 475
Japan
 Korean 329
 Papua New Guinea 633
 Japanese 18, 52, **287-9**
 Marshallse 406, 407
 Jarai (Aceh-Chamic) 50, 119
 Jarawa (Niger-Congo) 451
Java 50, 51, 99, 389, 418, 539
 Javanese 49, 50, **290-1**, 390
 Balinese 62
 Lampung 345
 Madurese 380
 Malagasy 387
 Sundanese 589
 Javindo (Dutch-Malay) 158
 Jewish Palestinian Aramaic 31
 Jewish Tat (Persian) 496
 Jews 31, 32, 33, 216, 222
 Crimean 615-6
 Hebrew 245-6
 Italian 284-5
 Judezmo 294
 see also Yiddish
 Jiarong (Bodic) 562
 Jinghpaw **292-3**, 552, 562
 Jita (Bantu) 71
 Jivaroan (Amerind) 21
 Jju (Niger-Congo) 451
 Jola *see* Diola (Niger-Congo)
Jordan 26, 546
 Judezmo 245, **294**, 489, 514, 582
 Jukun (Tiv) 451, 631
 Jumjum (Luo) 370
 Jutish (Danish) 146
- Kabardian (Circassian) 1, 136
 Kabiye (Niger-Congo) 451
 Kabras (Luyia) 376
 Kabré (Niger-Congo) 451
 Kabyle (Berber) 89, **295**
 Kacchī (Sindhi) 237, 556-7
Kachin 292, 562
 Kadai 52, 154, **296**, 355, 605
 Kadazan (Austronesian) 50
 Kado (Tibeto-Burman) 562
 Kadugli (Niger-Congo) 448
 Kagate (Tibetan) 628
 Kagulu (Bantu) 71, 594
 Kaitak (Caucasian) 110, 148
 Kaje (Niger-Congo) 451
 Kakwa (Nilotic) 454, 455, 645
 Kalabari (Ijo) **262**, **263**
 Kalami (Dardic) 267, 306, 307
 Kalanga (Bantu) 71
 Kalāṭī (Baluchi) 65
 Kalenjin (Nilo-Saharan) **297**, 375, 454, 646
Kalimantan 50, 257, 388, 390
 Kamano (Papuan) 491
 Kamaraz (Kashmiri) 267
 Kamba (Bantu) 71, **298**, 320
 Kambari (Niger-Congo) 451
 Kamchadal (Chukotko-Kamchatkan) 481
 Kamea (Papuan) 491
 Kami (Bantu) 71, 594
 Kamoro (Papuan) 491
 Kam-Sui (Kadai) 154, 296
 Kamtok (English pidgin) 157, 178, **299**
 Kanakanaey (Austronesian) 50
 Kana (Niger-Congo) 451
 Kandīālī (Dogri) 487
 Kandyān (Sinhala) 559
 Kanembu (Kanuri) 303
 Kāngri (Dogri) 487
 Kannada 155, 156, 169, 250, **300-1**, 402, 618, 639
 Kano (Hausa) 243
 Kanuri (Saharan) 29, **302-3**, 453, 454
 Kanyok (Bantu) 71, 365, 366
 Kaonde (Bantu) 71, 365, 366
 Kapampangan *see* Pampangan
 Kapucha (Caucasian) 109
 Karachai (Turkic) 337, 338, 648
 Karakalpak (Kazhak) 309, 648
 Karamojong (Nilo-Saharan) 455, 456, 645, 646
 Karanga (Shona) 554, 555
 Karapapak (Azeri) 58
 Karata (Caucasian) 53, 110
Karelia 525, 530
 Karelian 193, 194, 195
 Karen (Tibeto-Burman) 103, **304-5**, 561
 Karo Batak 79
 Kartvelian 109, 216, 217
 Kasa (Diola) 150, 151
Kashmir 279, 306, 628
 Kashmiri 267, **306**
 Kashubian (*Poland*) 498, 564
 Kâte (Papuan) 491
 Kawa (Wa) 672n
 Kawi (Javanese) 290
 Kayā (*Burma*) 304, 305
 Kazakh **308-10**, 343, 648
Kazakhstan 130, 223, 329, 526, 527
 Kechī (Baluchi) 65
 Kefa (Gonga) 228, 475
 Kei (Austronesian) 50
 Kejia (Chinese) 127, 128, 129
 Kekchī (Mayan) 410, 411
 Kela (Bantu) 71
 Kelao (Kadai) 356
 Kele (Bantu) 71
 Kembata (Cushitic) 142, 143
Kenya 70, 71, 454, 455, 456
 Cushitic 142, 143
 Gusii 239
 Kalenjin 297

- Kamba 298
 Kikuyu 319-20
 Luo 369-70
 Luyia 375-6
 Masai 409
 Meru 415
 Nubi 29
 Oromo 477-8
 Somali 570-1
 Swahili 592
 Teso 622
 Turkana 645, 646
 Kenzi 81
 Kerebe (Bantu) 71, 458
 Kerinci (Malayic) 391
 Ket (Palaeosiberian) 480, 481
 Kewa (Papuan) 491
 Kgatala (Tswana) 637
 Khakas (Turkic) 648, 649
 Khalka (Mongolian) 425, 426
 Kham (Tibetan) 627
 Khanag (Caucasian) 110
 Khandesi (*Gujarat*) 238
 Khanty (Ob-Ugric) 254, 256
 Kharia (Austroasiatic) 45, 46, 431
 Khari Boli (Hindi) 249, 250
 Khasi (Mon-Khmer) 41, 45, 46, 94, **311-12**
 Khas Kura *see* Nepali
 Khassonke (Mandekan) 397, 398
 Khayo (Luyia) 376
 Khinalug (Caucasian) 110
 Khmer (Mon-Khmer) 2, 45, 46, **313-14**
 Khmu (Mon-Khmer) 45, 46
 Khoe (Khwe) 315, 316, 679, 700
 Khoisan 4, **315-6**, 596, 679, 700
 Khorat (Thai) 348
 Khotanese (Iranian) 278, **317**
 Khowar 267, 306, **318**
 Khün (Tai) **346-7**
 Khunzakh (Avar) 53
 Khunzib (Caucasian) 53, 110
 Khvarshi (Caucasian) 53, 110
 Khwe (Khoisan) 315
 Kickapoo (Algonquian) 17
 Kiga (Bantu) 72, 457, 458
 Kikuyu 71, 298, **319-20**
 Kilba (Chadic) 243
 kiLuba (Bantu) 365
 Kimbundu (Bantu) 412, 413
 kiMwani (Swahili) 385
 Kinaray-a (Bisayan) 50, 112, 247
 Kipchak (Turkic) 337
 Kipsigis (Nilotic) 646
 Kirat (Newari) 446
 Kiribati 49, **321**, 407, 408
 Kirwa (Chagga) 415
 Kisa (Luyia) 376
 Kisii **239**
 Kissi (Niger-Congo) 451, 619
 Kituba **322**, 326
 Kitui (Kamba) 320
 Klao (Kru) 332, 449, 451
 Kleurling-Afrikaans 4
 Kodagu (Kannada) 301
 Koho (Austroasiatic) 46
 Kōlāmī (Dravidian) 227
 Kolana (Timor-Alor-Pantar) 392
 Kolnie-Deutsch 220
 Kolokuma (Ijo) 263
 Kom (Niger-Congo) 451
 Komering (Lampung) 50, 345
 Komi (Uralic) **323-4**, 661
 Komo (Bantu) 70, 71
 Komoro (Bantu) 594-5
 Kongo 71, 177, 322, **325-6**, 358
 Konjo (*Sulawesi*) 50, 99, 100
 Konjo/Konzo (Bantu) 71
 Konkani 266, **327**, 402, 403
 Konkomba (Niger-Congo) 451
 Kono (Niger-Congo) 451
 Konso (Cushitic) 142, 143
 Konyak (Kuki-Naga) 334
 Koranko (Mandekan) 397, 398
 Kordofanian (Niger-Congo) 447-8
 Korea 179, 287, 525
 Korean 18, **328-30**
 Koria (Bantu) 71
 Korku (Austroasiatic) 45, 46, 431
 Koroshi (Baluchi) 65
 Koryak (Chukotka-Kamchatkan) 481
 Kosali *see* Bhojpuri; Chhattisgarhi
Kosovo-Metohija 13, 14
 Kosraean (Micronesian) 407
 Kota (Dravidian) 301
 Koulán (Niger-Congo) 451
 Kpàrev (Tiv) 631
 Kpelle (Mande) 448, 451
 Kpone (Ga) 208
 Krahñ (Kru) 332, 451
 Krama (Javanese) 291, 589
 Krim (Niger-Congo) 619
 Krio 176, 178, 299, **331**
 Kriol (Australian pidgin) 179
 Krom-Hollands (Dutch-Malay) 158
 Krongo (Niger-Congo) 448
 Kru (Niger-Congo) **332**, 449
 Krumen (Kru) 332, 451
 Kryz (Caucasian) 110
 Kuanhua *see* Putonghua
 Kuanua (Austronesian) 51
 Kuba (Bantu) 365, 366
 Kuba (Lezghian) 354
 Kubachi (Caucasian) 110, 148
 Kui (Gondi) 227
 Kuki-Chin **333-5**, 399, 562
 Kukukuku (Papuan) 491
 Kulango (Niger-Congo) 451
 Kullo (Omoti) 228
 Kuman (Luo) 370, 454
 Kumauni (Pahari) 214, 267, **336**
 Kумык 59, 148, **337-8**, 648
 Kumzari (Iranian) 372
 Kunama (Nilo-Saharan) 453, 454
 Kunda (Nyanja) 71, 467

- !Kung 315, 316
 Kuo-yü *see* Pütöngghuà
 Kupa (Nupe) 465
 Kupang Malay 392
 Kuranko (Mandekan) 397, 398
 Kurdi (Kurdish) 340
 Kurdish 279, **339-40**, 371
 Kuria (Bantu) 71
 Kürin (Lezghian) 354
 Kurmanji (Kurdish) 340
 Kurukh (Dravidian) **341**
 Kusaal (Niger-Congo) 451
 Kuvi (Gondi) 227
Kuwait 27, 65
 Kuy (Austroasiatic) 45, 46
 Kwa (Niger-Congo) 9, 80, 185, 450
 Kwadi (Khoe) 316
 Kwahu (Akan) 10
 Kwama (Nilo-Saharan) 453, 454
 Kwanyama (Bantu) 71, 413
 Kwaya (Bantu) 71
 Kwese (Bantu) 72
 Kwéyol (French creole) 203
Kyrgysstan 32, 130, 526
 Kyrgyz 308, 309, **342-3**, 648
- Labrador* 276
 Lacandón (Mayan) 411
 Ladakhi (Tibetan) 626, 628
 Ladin (Rhaeto-Romance) 52, 514, 520
 Ladino (Judezmo) 294
 Lahnda (Panjabi) 486, 487
 Lahu 104-5, 241, **344**, 552, 562
 Lahuli (Tibetan) 628
 Lak (Caucasian) 110, 148
 Lakkia (Kadai) 356
 Lala (Bemba) 71, 85
 Lamaholot (Timor-Flores) 50, 392
 Lamani (*India*) 266
 Lamba (Bantu) 71, 85
 Lamnso' (Niger-Congo) 451
 Lampung (Austronesian) 49, 51, **345**
 Lamut (Tungusic) 641
 Land Dayak *see* Ngaju (Barito)
 Landoma (Temne) 619
 Landsmål (Norwegian) 459
 Langi (Bantu) 71
 Lango (Luo) 370, 454
 Lanna Thai **346-7**, 349, 605, 671
 Lao **348-9**, 416, 605
Laos 45, 46
 Hani 241
 Khün 346
 Miao 416, 417
 Lapp *see* Sami (Uralic)
 Larantuka Malay 392
 Larestani (Luri) 372
 Lārī (Sindhi) 556
 Lashi (Burmese-Lolo) 293
 Lāsī (Sindhi) 556
 Latgalian 64
 Lati (Kadai) 356
- Latin 275, **350-2**, 512, 513, 614
 African languages 27, 88
 Albanian 13, 14
 Arabic 28
 Balkans 39
 Celtic 97-8, 116, 198
 English 168-9
 French 198
 German 218
 Hungary 255
 Italian 283-4
 Konkani 327
 Lithuanian 360
 Occitan 468-9
 Portuguese 500
 Romanian 518-19
 Sanskrit 539
 Sardinian 543
 Spanish 579
 Swedish 598
 Welsh 674, 675
Latvia 222, 525, 527
 Latvian 64, 194, **353**, 530
 Lauan (Fijian) 191
 Lauje (Austronesian) 50
 Laurentian (Iroquoian) 123
 Lawa (Wa) 671, 672
 Lawangan (Austronesian) 50
 Laz (Kartvelian) 109, 217, 419, 420
Lebanon 26, 28, 546
 Ledo (Austronesian) 50
 Lega (Bantu) 71
 Lendu (Nilo-Saharan) 453, 454
 Lenje (Bantu) 70, 71, 85, 634
 Leonese (Spanish) 513, 580, 582
 Lepcha (Sino-Tibetan) 628
Lesotho 576, 578, 700-3
 Letzeburgisch 374
Leyte 51, 112, 673
 Lezghian (Caucasian) 59, 110, **354**
 Lhoba (North Assam) 562
 Li **355-6**
Liberia 206, 332, 414, 449, 450, 451, 452, 619
 Liberian English 178
Libya 26, 89, 456, 525, 607
 Limapuluh Kota (Malayic) 391
 Limba (Niger-Congo) 451, 619
 Limbu (Newari) 446, 562
 Lingala 71, 322, 326, **357-8**, 365, 537
 Lingua Franca 512-13
 Lingua Franqueada (Judezmo) 294
 Lingua Geral *see* Tupí
 Lio (Austronesian) 50
 Lisu 104, 105, 344, **359**, 552, 562
Lithuania 353, 497, 498, 525, 526
 Lithuanian 64, **360-1**, 530
 Livonian *see* Latvian
 Loanda Mbundu (Bantu) 412
 Lobi (Niger-Congo) 451
 Loda (Papuan) 621
 Logo (Nilo-Saharan) 453, 455
 Logooli (Bantu) 71, 375, 376

- Logudorese (Sardinian) 285, 543
 Loka of Mustang (Tibetan) 628
 Lokko (Mende) 414
 Lolo *see* Burmese
 Loloda (Papuan) 621
 Loma (Mande) 448, 451
 Lomavren (Indo-Aryan) 515
Lomblen Island 50
Lombok 50, 51, 545
 Lomwe (Bantu) 71, **362**, 386, 466
 Lotha (Naga) 334
 Lotuko-Maa (Nilotic) 455
 Louisiana Creole 203
 Lozi (Bantu) 71, **363**, 412, 578, 634
 Lü (Tai) **346-7**, 671
 Luanda Mbundu (Bantu) 72, 326, 413
 Luba (Bantu) 68, 71, 84, **365-6**, 367
 Lubumbashi Swahili 593
 Luchazi (Bantu) 71, 132, 133
 Ludian (Karelian) 194
 Lugbara (Nilo-Saharan) 453, 455
 Luimbi (Bantu) 132, 133
 Lul (Luo) 370
 luMasaaba *see* Gisu (Bantu)
 Lunda (Bantu) 71, **367-8**, 373, 412
 Lungchow (Southern Zhuang) 699
 Luo (Nilotic) **369-70**, 375, 455, 458
 Luorawetlan (Chukotko-Kamchatkan) 481
 Luri (Iranian) 279, 340, **371-2**
 Lusatian (Sorbian) 575
 Lushei (Kuki-Chin) 333, 334, 335
 Luvale (Bantu) 71, 133, **373**, 412
 Luwian (Anatolian) 252
 Luwo (Luo) 370
Luxembourg 201
 Luxemburgish 223, 225, **374**
 Luyi (Bantu) 363
 Luyia (Bantu) 69, 71, 369, **375-6**, 622
Luson 50, 51, 92, 111, 112, 264, 484, 485, 600, 673
 Lwo (Luo) 370
 Lycian (Anatolian) 252
 Lydian (Anatolian) 252
- Maba (Nilo-Saharan) 29, 302, 303, **377**, 453, 455
 Mabaan (Luo) 370
 Mabwe-Lungu (Bantu) 72
 Mača (Oromo) 478
Macao 502
Macedonia 39, 101, 516, 519, 564, 654
 Macedonian **378-9**, 473, 550, 564, 568
 Machakos (Kamba) 320
Madagascar 593
 see also Malagasy
Madeira 501
 Madi (Nilo-Saharan) 453, 455
 Madurese 49, 50, 290, 291, **380**
 Mafa (Chadic) 243
 Magadhi (Prakrit) 265, 266
 Magahi 90, 250, 341, **381**, 383
 Magar/Magari (Newari) 446, 562
 Magindanao (Austronesian) 50, **382**
 Māhārāṣṭri 265
- Mahi (Gbe) 186
 Mai Brat (Papuan) 621
 Maithili 86, 90, 91, 250, 266, **383**
 Makasai (Timor-Alor-Pantar) 392
 Makasar (Austronesian) 44, 49, 51, 99, 100, **384**
 Makonde (Bantu) 71, **385**, 386
 Makua (Bantu) 71, **386**, 466
 Malagasy 49, 51, **387-8**
 see also Madagascar
Malawi 71, 72, 73, 701, 702
 Lomwe 362
 Makua 386
 Nyanja 466-7
 Tumbuka 640
 Yao 683-4
 Malay 29, 30, 47, 49, 51, **389-92**, 490
 Achehnese 2
 Balinese 62
 Batak 78
 Cham 118
 Javanese 290
 Madurese 380
 Malagasy 387
 Rejang 510
 Sundanese 589
 Ternate 620
 see also Malaysia
 Malayalam 155, 156, 327, **393**, 610, 611
 Malayo-Polynesian 48, 49, 196
Malaysia 45, 46, 48, 390, 391, 539
 English 173
 Iban 257
 Japanese 287
 Malaccan Creole 502
 Malayalam 393
 Minangkabau 418
 Tamil 611
 see also Malay
Maldives 152, 560
Mali 89, 449, 450, 451, 452, 455
 Fulani 206
 Mandekan 395, 396, 397
 Songhay 572
 Soninke 574
 Tamasheq 607
 Malinke (Mandekan) 396
 Maliseet (Algonquian) 17
 Malpaharia (Bengali) 42
Malta 284
 Maltese 26, 27, 28, **394**
 Malto (Dravidian) 341
 Mālvi (Rajasthani) 238, 509
 Mālwar (Panjabi) 487
 Mam (Mayan) 410, 411
 Mamara (Senúfo) 452
 Mambai (Timor-Flores) 392
 Mamuju (Austronesian) 51, 100
Man, Isle of 113, 461, 471
 Manchu (Tungusic) 641-3
Manchuria 328, 329
 Mandacan, Modern 33
 Mandar (Austronesian) 51, 99, 100, 384

- Mandari (Nilo-Hamitic) 645
 Mandarin *see* Pütõnghuà
 Mande (Niger-Congo) 395, 447, 448, 451, 590
 Mandeali (Pahari) 267
 Mandekan 29, 80, 141, **395-8**, 448, 451, 574, 678
 Mandinka (Mandekan) 396
 Mandyak (Diola) 151, 451
 Manga Kanuri 303
 Mangala (Bantu) 357
 Mangarevan (Tahitian) 604
 Mangbetu (Nilo-Saharan) 453, 455
 Manggarai (Austronesian) 51
 Manica (Shona) 554
 Manipuri 334, **399**, 562
 Manja (Niger-Congo) 451
 Mankanya (Niger-Congo) 451
 Mano (Niger-Congo) 451
 Manobo (Austronesian) 51
 Mansi (Ob-Ugric) 254, 256
 Manx 113, 115, 210
 Mao (Naga) 334
 Maori 49, 51, 173, **400-1**
 Mapudungun (Araucano) 34
 Mapun (Austronesian) 50
 Marama (Luyia) 376
 Maranao (Danao) 51, 382
 Marathi 238, 265, 266, 300, 327, **402-3**
 Maraz (Kashmiri) 267
 Margi (Chadic) 243
Marianas 120, 407
 Mari (Uralic) **404-5**, 661
 Marma (Arakanese) 105
 Marshallese 49, **406-8**
Martinique 203
 Maru (Burmese-Lolo) 105, 293, 562
 Mārwarī (Rajasthani) 238, 509
 Masaba (Bantu) 72, 376
 Masai 319, 320, **409**, 455, 646
 Masbateño (Bisayan) 51, 112, 247
 Masenrempulu (Austronesian) 51
 Massalat (Maban) 453, 455
 Matengo (Bantu) 72
 Matutenho (Portuguese creole) 501
 Matuwari (creole) 177-8
 Mauricien 203, 204
 see also *Mauritius*
Mauritania 25, 26, 89, 206, 574, 614
Mauritius 91
 see also Mauricien
 Maya *see* Yucatec (Mayan)
 Mayan 21, **410-11**, 508, 694
 Mayo (Uto-Aztecan) 439
 Mazahua (Otomanguean) 422
 Mazandarani (Caspian) 226
 Mazatec (Otomanguean) 422
 Mazovian (Polish) 498
 Mbai (Nilo-Saharan) 455
 Mbala (Bantu) 72
 Mbanza (Niger-Congo) 451
 Mbene (Bantu) 70
 Mbola (Bantu) 72
 Mbum (Adamawa) 449
 Mbunda (Bantu) 72, 132, 133
 Mbundu (Bantu) 72, **412-13**
 Mbwela (Bantu) 72, 132, 133, 373
 Medlpa (Papuan) 491
 Medumba (Niger-Congo) 451
 Meglenite (Aromunian) 519
 Mehri (South Arabian) 29
 Meidob (Nubian) 462, 463
Melanesia 49, 179
 Melpa (Papuan) 491
 Mende (Mande) 395, **414**, 448, 451, 619
 Menggu (Mongolian) 425
 Menominee (Algonquian) 17
 Merina (Malagasy) 388
 Meru (Bantu) 72, 298, 320, **415**
 Meuneu (Lahu) 344
 Mewati (Rajasthani) 238
Mexico 20, 21, 410-11
 Mixtec 421-2
 Nahuatl 436-9
 Spanish 582, 583
 Yucatec 694-5
 Zapotec 697
 Miao (Miao-Yao) 52, 154, **416-17**, 561, 685
 Micmac (Algonquian) 17
Micronesia 49, 321
 Micronesian 406-8
 Mijikenda (Bantu) 72, 594, 595
 Mikir-Methei (Kuki-Naga) 334, 335, 562
Milne Bay 50
 Mimi (Maban) 453, 455
 Min (Chinese) 127-30, 391
 Minaean (South Arabian) 29
 Minangkabau (Austronesian) 49, 51, 78, 389, 391, 392, **418**
Mindanao 50, 51, 111, 112, 583, 584
Mindoro 111, 601
 Mingrelian 1, 109, 216, 217, **419-20**
 Miniferi (Saho) 3
 Miri (North Assam) 562
 Mirish (Baric) 562
 Miskito (Amerind) 21
 Miskito Coast Creole 176
 Misumalpan (Amerind) 21
 Mitchif (French Cree) 17, 20, 139, 202
 Mittelfränkisch 222
 Mixe-Zoque (Amerind) 21, 437, 694
 Mixtec (Amerind) 20, 21, **421-2**, 697
 Mizo (Kuki-Chin) 334
 Mlabri (Mon-Khmer) 45
 Mnong (Austroasiatic) 45, 46
 Moba (Niger-Congo) 451
 Mobilian (Amerind) 20, 21
 Moca/Mocha (Gonga) 228
 Mocaranga (Shona) 554
 Moghol (Mongolian) 425-6
 Mohawk (Iroquoian) 123
 Mokilese (Micronesia) 407
 Moksha *see* Mordvin (Uralic)
Moldova 101, 518, 519, 525, 653, 654
 Mon (Austroasiatic) 2, 45, 46, 103, 104, 106, **423**, 552

- Mongo (Bantu) 72
Mongolia 525, 648
 Mongolian 18, 75, 371, **424-6**
 Mon-Khmer (Austroasiatic) 45, 311, 313, 314, 423, 670
 Montagnais (Algonquian) 17
Montenegro 548-50
 Môôre (Niger-Congo) **427**, 449, 451
 Moose Cree 17
 Mopan (Mayan) 411
 Moraid (Papuan) 621
 Moravian (Slavonic) 499
 Mordvin (Uralic) **428-9**, 661
 Moré *see* Môôre
Morocco 26, 28, 89, 511, 608-9, 614
 Motu (Austronesian) 50, **251**, 633
Mozambique 70, 71, 72, 73, 525, 701, 702
 Lomwe 362
 Makonde 385
 Makua 386
 Nyanja 466
 Portuguese 501
 Sena 547
 Shangaan 636
 Shona 554-5
 Swahili 593
 Swazi 596
 Yao 683-4
 Mozarabic **430**, 500, 513, 579
 Mpuono (Bantu) 72
 Mūltānī (Panjabi) 487, 488
 Muna (Austronesian) 51
 Mundari (Munda) 41, 45, 46, 381, **431-2**, 541
 Munsee (Algonquian) 17
 Muong (Austroasiatic) 45, 46, **433**, 669
 Muskogean (Amerind) 21
 Mwanga (Bantu) 72
 Mwera (Bantu) 72, 386
 Mzab (Berber) 89
- Na-Dené 19, 181, **434-5**, 440
 Naga **334-5**, 399, 562
 Naga Pidgin (Nagamese) 42, 334
Nagorno-Karabakh 37-8, 58
 Nāgpurī (Konkani) 402
 Nahuatl (Amerind) 20, 21, 169, **436-9**, 508, 582, 694
 Nakh (Caucasian) 109, 121
 Nama (Khwe) 315, 316
Namibia 4, 71, 315, 316
 Nanai (Tungusic) 641
 Nancowry (Austroasiatic) 45, 46
 Nandi (Bantu) 72
 Nandi (Eastern Nilotic) *see* Kalenjin
Nauru 178, 321
 Nauruan (Micronesian) 407
 Navaho 19, 434, **440**
 Nawdm (Niger-Congo) 451
 Naxi 61, **441-2**, 562
 Ndandi (Bantu) 72
 Ndebele (Bantu) 72, **443**, 554, 701, 702
 Ndengereko (Bantu) 72
 Ndjuka (Bush Negro) 178
- Ndonga (Bantu) 72, 413
 Negerhollands 160
 Nembe (Ijo) 263
 Nenets (Samoyedic) 534-5
Nepal 42, 266, 341, 383, 445-6, 562, 626
 Nepali 161, 214, 267, 336, **444**
 Newari (Sino-Tibetan) **445**, 562
New Guinea *see* Irian
New Zealand 51, 173, 533
 Maori 400-1
 Ngaju (Barito) 257, 387, 388
 Ngalong (Tibetan) 161, 627
 Ngambai (Nilo-Saharan) 455
 Nganasan (Samoyedic) 534, 535
 Ngando (Bantu) 72
 Ngatikese (Micronesian) 407
 Ngbaka (Niger-Congo) 452
 Ngbandi (Ubangi) 449-50, 452, 536, 537, 696
 Ngindo (Bantu) 72
 Ngoko (Javanese) 291, 589
 Ngombe (Bantu) 72
 Ngonde (Bantu) 72, 701, 702
 Ngoni (Bantu) 72, 640
 Ngulu (Bantu) 71, 72, 594
 Nguni (Bantu) 315, 443, 576, 596, 679, 700, 701, 702
 Ngwana (Swahili) 593
 Ngyemboon (Niger-Congo) 452
 Nhang (Tai) 107
 Nias (Austronesian) 51
Nicaragua 176
 Nicobarese 45-6
 Niederfränkisch 159
 Niederpreussisch 159
 Niedersächsisch 159, 222
Niger 89, 454, 455
 Fulani 206, 207
 Hausa 242
 Songhay 572
 Tamasheq 607
Nigeria 68, 117, 450, 451, 452, 454, 455
 Arabic 28
 Edo 162-3
 Efik *and* Ibibio 164-5
 English pidgin 178
 Fulani 206
 Hausa 242
 Igbo 260-1
 Ijo *and* Kalabari 262-3
 Nupe 464-5
 Songhay 572
 Tiv 630-1
 Yoruba 692-3
 Nilo-Hamitic 455, 645
 Nilo-Saharan 149, 297, 302, 369, 377, 409, **453-6**, 462, 542, 572, 622, 645
 Nilotic (Nilo-Saharan) 455, 622, 645
 Nilyamba (Bantu) 72
 Nimadi (Rajasthani) 238
 Ningo (Ga) 208
 Niuean (Polynesian) 635
 Nivkh (Palaeosiberian)) 481

- Niya Prakrit 265
 Nkore-Kiga (Rutara) 72, 213, **457-8**, 523
 Nkoro (Ijo) 263
 Nobiin (Nubian) 462, 463
 Nogai (Turkic) 59, 310, 648
 Nootka 20
 Nordsächsisch 159
 Norn (Norse) 472
 Norse 113, 146, 190, 259, 460-1
 English 169
 Norwegian 459
 Old Norse 471-2
 Swedish 598
 Northern Gurage 24
 Northern Ndebele 702
 Northern Paiute (Uto-Aztecán) 438, 439
 Northern Samo (Niger-Congo) 452
 Northern Zhuang (Tai) 605, 699
Norway 146, 471, 530-2
 Norwegian 225, **459-61**, 599
Nova Scotia 209
 Nsenga (Bantu) 72, 701, 702
 Ntomba (Bantu) 72
 Ntumu (Bantu) 188
 Nu (Miao-Yao) 685
 Nubi (Arabic) 29
 Nubian 81, 455, **462-3**
 Nuer (Nilotic) 149, 455
 Nung (Southern Zhuang) 699
 Nupe (Niger-Congo) 452, **464-5**
 Nuristani (*Afghanistan*) 267
 Nyabungu (Bantu) 73
 Nyakyusa (Bantu) 72, 701, 702
 Nyala (Luyia) 376
 Nyali (Bantu) 70
 Nyamwezi (Bantu) 72, 585, 586
 Nyangatom (Turkana) 645
 Nyanja (Bantu) 72, 84, **466-7**, 547, 640, 701, 702
 Nyankole *see* Nkore-Kiga
 Nyaturu (Bantu) 72
 Nyemba (Bantu) 72
 Nyiha (Bantu) 72
 Nymylyan (Chukotko-Kamchatkan) 481
 Nynorsk (Norwegian) 459
 Nyore (Bantu) 71, 375, 376
 Nyoro (Bantu) 457, 458
 Nyungwe (Nyanja) 72, 467
 Nzema (Niger-Congo) 452

 Ob-Ugric 254-5
 Occitan 108, 199, 201, 219, 283, **468-9**, 513, 514
 Odawa (Cree) 17
 Oirat (Mongolian) 425
 Oiro (Turkic) 648
 Ojibwa (Algonquian) 17, **470**
 Okrika (Ijo) 263
 Old Franconian 158
 Old Norse **471-2**
 Old Prussian 64
 Old Slavonic **473-4**, 519
 Olmec (Mixe-Zoque) 410, 694
 Olonets (Karelian) 194

Oman 27, 65
 Ometo (Omotic) 228, 475, 677
 Omotic 6, 7, 228, **475**, 677
 Onandaga (Iroquoian) 123
 Oneida (Iroquoian) 123
 Ong-be (*Hainan*) 355, 356
 Onitsha (Igbo) 261
 Opëno (Luo) 370
 Orang Benua (Austroasiatic) 45, 46
 Orange River Afrikaans 4
 Oriya 41, 42, 87, 266, **476**
Orkney Islands 472
 Orma (Oromo) 478
 Orochen (Tungusic) 641
 Oromo (Cushitic) 29, 142, 143, 369, 370, 475, **477-8**, 571
 Oscan 350
 Osing (Javanese) 291
 Ossete 109, 279, **479**
 Ostniederdeutsch 222
 Ostpommersch 159
 Ostyak (Ob-Ugric) 256
 Otomanguean (Amerind) 21, 421, 697
 Otomi (Otomanguean) 422
 Ottawa (Cree) 17
 Otuho (Nilotic) 455, 646
 Owere (Igbo) 261

 Pachuco (Spanish American argot) 582-3
 Padang (Dinka) 149
 Paezan (Amerind) 21
 Pahari (Indo-Aryan) 214, 267, 336
 Pai *see* Bai
 Paite (Kuki-Chin) 333, 334
 Paiwan (Formosan) 197
Pakistan 248, 266
 Baluchi 65-6, 96
 Brahui 96, 155, 156, 341
 Gujarati 237
 Kashmiri 306
 Khowar 318
 Panjabi (Lahnda) 486-8
 Pashto 492-3
 Sindhi 556
 Siraiki 563
 Urdu 663
 Päkot (Nilotic) 646
 Pakpak (Batak) 79
 Pak Thai 348
 Palaeosiberian **480-1**, 681
 Palaic (Anatolian) 252
 Palau (Austronesian) 120
 Palaung (Mon-Khmer) 45, 46, 552, 671, 672
Palawan 50, 111, 112, 601
Palestine 32
 Pali 265, **482-3**
 Burmese 104
 Khmer 313
 Lü 346
 Sanskrit 539, 540
 Sinhala 558
 Thai 623, 624

- Pama-Nyungan (Australia) 43
 Pamir (Iranian) 279
 Pamona (Austronesian) 51
 Pampangan (Austronesian) 49, 51, **484**, 602
Panama 21
Panay 50
 Pangasinan (Austronesian) 51, 264, **485**, 601, 602
 Pangwa (Bantu) 72
 Panjabi 153, 250, 266, **486-8**, 515, 563, 663
 Panjgūrī (Baluchi) 65
Pantar Island 50
 Pa-o (*Burma*) 305
 Papel (Niger-Congo) 451
 Papia Kristang (Malaccan Creole) 502
 Papiamentu 160, **489**, 503, 582
 Papuan 46, **490-1**, 620
Papua New Guinea 50, 251, 490, 491, 633
Paraguay 21, 235-6, 581, 583
 Paraniotic 455
 Pare (Bantu) 593
 Parjī (Dravidian) 227
 Parthian 278, 494
 Parya (*Tajikistan*) 238
 Pashai (*Afghanistan*) 267, 268
 Pashto 279, **492-3**
 Pasisir (Malayic) 391
 Passamaquoddy (Algonquian) 17
 Patani (Ibo) 263
 Patani Malay 390
 Patwa (French creole) 203
 Pedi (Northern Sotho) 578
 Pehlevi (Persian) 494, 496
 Pehuenche (Araucanian) 507
 Pekingese *see* Putonghua
 Pennsilfaanisch 220
 Penutian (Amerind) 21
 Permyak (Komi) 324
 Persian 29, 30, 248, 278, 339, 426, 492, 493, **494-6**
 Baluchi 65
 Luri 371, 372
 Old Persian 12
 Sogdian 569
Peru 57, 505-7, 581, 583
 Pesisir (Austronesian) 51
 Petit Mauresque 202
 Petjo (Dutch) 158
Petra 31
 Phalok (Wa) 671
 Phende (Bantu) 72
Philippines 49, 51
 Bikol 92
 Cebuano 111-12
 Chamorro 120
 Danao 382
 Hiligaynon 247
 Ilocano 264
 Pampangan 484
 Pangasinan 486
 Spanish 583
 Tagalog 600-2
 Waray-waray 673
 Phoenician 546
 Pictish 114-15
 Pijin (*Solomon Islands*) 179
 Pikunche (Araucanian) 507
 Pilipino *see* Tagalog
 Pingal (Rajasthani) 509
 Pingelapese (Micronesian) 407
 Pinyin (Chinese Latin orthography) 128
 Pipil (Nahautl) 436
Pitcairn Island 179
 Plains Cree 17, 139
 Plains Indian Sign Language 20
 Pochismo (Spanish American argot) 582-3
 Pogolo (Bantu) 72
 Pohnpeian (Micronesian) 407, 408
 Pokomo (Bantu) 477, 595
 Pökoot (Nilo-Saharan) 455
 Polabian (Slavonic) 499
Poland 159, 222, 525
 Ukraine 659
 see also Polish
 Police Motu (*Papua*) 251
 Polish 82, 145, **497-9**, 564
 Lithuanian 361
 see also Poland
Polynesia 49, 400, 603
 Polynesian 321, 533, 635
 Pomeranian 498
 Pontic (Greek) 233
 Popoluca (Amerind) 21
 Porbat (Indo-Aryan) 444, 445
 Porohanon (Bisayan) 111
Portugal
 Caló 517
 Mozambique 547
 Portuguese 108, 162, 173, **500-4**, 514
 Afrikaans 5
 Crioulo 141
 Galician 211
 Kannada 300
 Kongo 325
 Luba 365
 Lunda 367
 Papiamentu (creole) 489
 pidgin 176
 Shona 554
 Sinhala 559
 Swahili 593
 Ternate 620
 Tupí 644
 Posha (Armenian) 37
 Potawatomi (Algonquian) 17
 Pōthwārī (Panjabi) 487, 488
 Poular (Fulani) 151
 Pōwādhīr (Panjabi) 487
 Prakrits 265, 317, 402, 482, 617
 see also Sanskrit
 Prampram (Ga) 208
 Prasun 268
 Provençal *see* Occitan
 Pubian (Austronesian) 51
Puerto Rico 160, 581, 582
 Punic (Semitic) 546

- Punjabi *see* Panjabi
 Purbī/Purbīyā (Bhojpuri) 90
 Purik (Tibetan) 628
 Pǔtōnghuà (Mandarin) 126-7, 128, 129
 Pwo (*Burma*) 304, 305
 Pyu (Sino-Tibetan) 561
- Qaidaq (Caucasian) 110, 148
 Qāshqāy (Turkic) 59, 654
 Qatabanian (South Arabian) 29
Qatar 27
 Qiang (Tibeto-Burman) 562
 Quara (Cushitic) 143
 Quechua (Amerind) 20, 21, 57, **505-7**, 582
 Quiché (Mayan) 20, 21, 411, 437, **508**
- Rade (Aceh-Chamic) 51, 119
 Rahanweyn Somali 571
 Rāi (Sino-Tibetan) 444
Raijua 51
 Rajasthani 238, 250, 266, 487, **509**
 Rajbangshi 42
 Rākshānī (Baluchi) 65
 Rapan (Tahitian) 604
 Rapanui (Austronesian) 49
 Rarotongan (Maori) 401
 Rāthīr (Panjabi) 487
 Rawang (Tibeto-Burman) 562
 Red Tai 349, 605
 Rejang (Austronesian) 49, 51, 345, **510**
 Rek (Dinka) 149
 Remi (Bantu) 72
 Réunionnais 203-4
 Rhaeto-Romance 513, 514
 Rheinfränkisch 222
Riau Islands 46
 Rifia (Berber) 89, **511**
 Riksmål (Norwegian) 459, 460
 Rindi (Baluchi) 65
 Rodiyas (Sinhala) 560
 Roglai (Aceh-Chamic) 119
 Romance 39, 108, 198, 211, 275, 283, 430, 468, 500,
 512-14, 518, 520, 543, 579
 Romani 266, 488, **515-17**, 519
 Albanian 14
Romania 37, 230, 474, 525
 Romanian 14, 39, 352, 514, 515, **518-19**, 564
 Romansch 514, **520-1**
 Romblomanon (Bisayan) 51, 112
 Rong (Sino-Tibetan) 628
 Ronga (Bantu) 636
 Roshani (Pamir) 279
 Rotinese (Timor-Flores) 51, 392
 Rotuman (Fijian) 191, 192
 Rubassa (Niger-Congo) 452
 Rufiji (Bantu) 72
 Ruguru (Bantu) 72, 594
 Rundi (Bantu) 72, 213, 458, **522-3**
 Rung (Tibeto-Burman) 562
 Rungu (Bantu) 72
 Runyarwanda (Bantu) 523
 Runyoro (Bantu) 457, 458
- Russenorsk (Norwegian) 460
Russia
 Abaza 1
 Abkhaz 1
 Avar 53
 Bashkir 75
 Chechen 121-2
 Chuvash 134-5
 Circassian 136
 Dargwa 148
 German 222-3
 Ingrian/Veps 194-5
 Kazakh 308-10
 Komi 323-4
 Korean 329
 Kumyk 337
 Lezghian 354
 Mari 404-5
 Mongolian 424-5
 Mordvin 428-9
 Norse 471, 598
 Pontic 233
 Tungusic 641
 Turkic 648
 Udmurt 656
Ukraine 659-60
 Yakut 681-2
 Yiddish 688-9, 690
 see also USSR, former
 Russian 75, 82, 83, 109, 110, 148, 169, 275, **524-9**,
 564
 Finnish 194
 Ossete 479
 Tatar 615
 Yiddish 690
 Rutara (Bantu) 457-8
 Ruthenian 82, 528, 564
 Rutul (Caucasian) 110
 Ruund *see* Lunda (Bantu)
 Ruvu (Bantu) 594
Rwanda 72, 158, 593
 Rwanda (Bantu) 72, **522-3**
 Ryukyuan (Japanese) 289
- Saamia (Bantu) 71, 376
 Sabaeen (South Arabian) 29
Sabah 50, 51
 Sabaki (Bantu) 594-5
 Sabaot (Nilo-Saharan) 455
 Sabir (Romance) 512
 Sadan (Austronesian) 51
 Sadani (Bhojpuri) 90
 Sadhu-Bhasha (Hindi) 249
 Saek (Tai) 107
 Safwa (Bantu) 72
 Sahadic (Coptic) 137
 Saharan *see* Nilo-Saharan
 Saho (Cushitic) 3, 142, 143
 Sahu (Papuan) 620, 621
St Lucia 203, 204
 Saipan Carolinian 120, 407
 Saka (Iranian) 278

- Sakalava (Malagasy) 388
 Sakuye (Oromo) 478
 Salar (Turkic) 648
Samar (Philippines) 51, 673
 Samaritan 31-2
 Samba (Bantu) 72
 Sambalpurī (Oriya) 41, 476
 Samburu (Nilo-Saharan) 455, 646
 Sami (Uralic) 183, 224, **530-2**, 661
 Samnorsk (Norwegian) 460
 Samoan 49, 51, **533**, 635
 Samoyedic **534-5**
 Sandawe (*Tanzania*) 316
 Sanga (Bantu) 72, 365
 Sangir (Austronesian) 51
 Sango (Niger-Congo) 67, 450, 452, **536-7**
San Marino 284
 Sanskrit 27, 62, 155, 265, 266, **538-40**
 Assamese 41
 Australia 44
 Balinese 62
 Bhojpuri 90
 Cham 118
 Gujarati 237
 Hindi 248
 Indo-European 270-1, 272
 Javanese 290
 Kannada 300
 Khmer 313
 Malagasy 387
 Malayalam 393
 Sundanese 589
 Tamil 612
 Telugu 617
 Tibetan 626
 see also Prakrits
 Santali (Austroasiatic) 45, 46, 383, 431, **541**
Sao Tomé and Príncipe 501
 Sara (Nilo-Saharan) 302, 453, 455, **542**
 Saramaccan (creole) 177-8, 503
Sarawak 257
 Sarawānī (Baluchi) 65
Sardinia 108
 Sardinian 285, 514, **543-4**
 Sarhaddī (Baluchi) 65
 Sariqoli (Pamir) 279
 Sarnami (Bhojpuri) 91
 Sasak (Austronesian) 49, 51, **545**
 Sassarese (Sardinian) 285
Saudi Arabia 27
 Sauraseni (Prakrit) 265
 Saurashtri (*Madurai, India*) 266
 Sawu (Austronesian) 51
 Scandinavian 146, 210, 471
Scotland 114, 115, 172, 209-10, 461
 Scouse 171
 Sea Dayak *see* Iban
 Sea Lapp 531-2
 Seba (Bantu) 71
 Sebei (Nilo-Saharan) 376
 Sedang (Austroasiatic) 45, 46
 Sehwi (Akan) 9
 Selayar (Bugis) 99, 100
 Selkup (Samoyedic) 256, 534, 535
 Sema (Naga) 334
 Semai (Austroasiatic) 45, 46
 Semitic 6-7, 11-12, 23, 25, 31, 117, 184, 245, 394, **546**, 629
 Semnani (*Iran*) 226
 Sena (Bantu) 72, **547**
 Senadi (Senufo) 449, 452
 Seneca (Iroquoian) 123
Senegal 141, 206, 207, 450, 451, 452
 Diola 150-1
 Mandekan 395, 396, 397
 Soninke 574
 Wolof 678
 Sengoi ((Austroasiatic)) 45, 46
 Senufo (*West Africa*) 450, 452
Serbia 13, 473, 654
 Serbian 13, 14, **548-50**, 564, 567, 568
 Serbo-Croat 548-50
 Serer (Fulani) 150, 151, 449, 452
 seSotho na Leboa (Northern Sotho) 577-8
 Seuta (Bantu) 594
 Seychellois 203, 204, **551**
Sgaw (Burma) 304, 305
 Shahdag (Caucasian) 110
 Shaho *see* Saho (Cushitic)
 Shāhpūrī (Panjabi) 487
 Shambala (Bantu) 73, 594, 595
 Shan (Tai) 103, 105, **552-3**, 605, 671
 Shapsugh (Circassian) 1
 Shawia (Berber) 89
 Shawnee (Algonquian) 17
 Sheldru (Irish) 281
 Sherbro (Niger-Congo) 450, 619
Shetland Islands 472
 Shhawri (South Arabian) 29
 Shi (Bantu) 73
 Shilha (Berber) 511, 608-9, 614
 Shilluk (Luo) 369, 370, 455
 Shina (Dardic) 267, 306
 Shinasha (Gonga) 228
 Shitire (Tiv) 631
 Shona 26, 73, 443, **554-5**
 Shopski (Bulgarian) 550
 Shor (Turkic) 648, 649
 Shoshone (Uto-Aztecan) 438, 439
 Shqip (Albanian) 13
 Shu (Bantu) 72
 Shuar (Amerind) 21
 Shubi (Bantu) 523
 Shughni (Pamir) 279
 Siamese *see* Thai
Siberia 181, 277, 480-1, 525, 534, 681, 690
Sibuyan Island 51
 Sicilian 285, 514
 Sidamo (Cushitic) 142, 143
Sierra Leone 176, 450, 451, 452
 Fulani 206
 Krio 331
 Mandekan 397
 Mende 414

- Susu 590
 Temne 619
 Sign languages 556
 Sijo (Korean) 329
 Sikka (Austronesian) 51
 Sikkimese (Tibetan) 626, 628
 Sikule (Austronesian) 51
 Sila (Nilo-Saharan) 454
 Simalur (Austronesian) 51
 Sindhi 65, 266, **556-7**, 563
 Sine'i (Kurdish) 340
Singapore 99, 127, 128
 English 173
 Japanese 287
 Malay 390
 Malayalam 393
 Portuguese 502
 Sinhala 559
 Tamil 611
 Sinhala 152, 266, 502, **558-60**, 611
 Sintu (Bantu) 68
 Siraiki (Panjabi) 487, **563**
 Sisaala (Niger-Congo) 452
 Siwa (Berber) 89
 Skepi (Dutch) 160
 Slave (Atheaskan) 434
 Slavenoserbian 473
 Slavonic 82, 101, 144, 224, 256, 274, 275, 378, 497,
 524, 548, **564**, 566, 567, 575, 659
 Old Slavonic 473-4, 519
 Slovak 145, 499, 564, **566**
Slovakia 516
 Slovene 549, 564, **567-8**
 Sô (Austroasiatic) 45, 46
 Soddo (Semitic) 24
 Soga (Bantu) 71, 213
 Sogdian (Iranian) 278, 279, **569**, 657
 Sokoto (Hausa) 243
Solomon Islands 179, 321
Solor Island 50
Somalia 25, 27, 477-8
 Somali (Cushitic) 29, 142, 143, 477-8, **570-1**
 Songe (Bantu) 73, 365, 366
 Songhay (Nilo-Saharan) 29, 396, 427, 447, 455, **572**
 Soninke (Mande) 29, 395, 397, 398, 448, 452, **574**
 Soqotri (South Arabian) 29
 Sora (Austroasiatic) 45, 46, 431-2
 Sorani (Kurdish) 340
 Sorbian (Slavonic) 145, 498, 499, 564, **575**
 Sorsogon (Masbate Sorsogon) 51, 247
 Sotho (Bantu) 73, 363, **576-8**, 596, 637, 638, 679
South Africa
 Afrikaans 4-5
 Bhojpuri 91
 English 173
 German 221
 Khoisan 315-6
 Ndebele 443
 Sintu (Bantu) 68, 73
 Sotho 576-8
 Swazi 596-7
 Tsonga 636
 Tswana 637
 Venda 667
 Xhosa 679-80
 Zulu 700-3
 South Arabian 29
 Southern Zhuang (Tai) 605, 699
Soviet Union, former see *USSR, former*
Spain
 Arabic 25, 27
 Basque 76-7
 Caló 516-17
 Catalan 108
 Celtic 113
 Hebrew 245
 Mozarabic 430
 Spanish 108, 275, 513, 514, 543, **579-84**, 614
 Arabic 26
 Araucanian 34
 Aymara 57
 Caribbean 204
 Cebuano 111
 Chamorro 120
 Gothic 229
 Judezmo 245, 294
 Latin 352
 Mayan 410
 Nahuatl 436-7
 Pampangan 484
 Pangasinan 485
 Papiamentu (creole) 489
 Paraguay 236
 Philippines 600-1
 pidgin 176
 Portuguese 501
 Quechua 505
 Sranan (*Suriname*) 160, 177
 Sre (Austroasiatic) 45, 46
Sri Lanka 539
 English 173
 Malay 391
 Portuguese Creole 501-2
 Sinhala 558-60
 Tamil 611
 Sri Lankan Creole Malay 390
 Stieng (Austroasiatic) 45, 46
 Stokavian (Serbo-Croat) 549
 Subanun (Austronesian) 51
 Subselvan (Romansch) 520
Sudan 24, 26, 28, 29, 448, 454, 455, 456
 Beja 81
 Cushitic 142, 143
 Dinka 149
 Fulani 206
 Luo 369-70
 Maba 377
 Nubian 462-3
 Turkana 645, 646
 Zande 696
 Suebic (*Iberia*) 500
 Sui (Kadai) 154
 Suk (Nilo-Saharan) 455
 Sukuma (Bantu) 73, **585-6**

- Sulawesi* 50, 51, 99-100, 384
Sulu (Austronesian) 51, 111, 112
Sumatra 2, 50, 51, 79, 345, 389, 391, 418, 510, 539
Sumba (Austronesian) 51
Sumbawa (Austronesian) 51, 545
Sumbwa (Bantu) 73, 585, 586
Sumerian 6, 11, **587-8**
Sundanese (Austronesian) 49, 51, 290, 291, **589**
Sungor (Nilo-Saharan) 455
Suriname 91, 158, 160, 176, 177-8
Surit (Aramaic) 32
Surselvan (Romansch) 520, 521
Susu (Mande) 29, 395, 448, 452, 574, **590**
Suvalkiečiai (Lithuanian) 64
Svan 109, 216, 217, 420
Swahili 29, 30, 68, 73, 212, 322, 585, **591-5**
 Kenya 319
 Luba 365
 Luo 369
 Luyia 376
 Malagasy 387
 Masai 409
 Mozambique 385
 Shona 554
 Yao 683
 Zulu 701
Swampy Cree 17, 139
Swatownese *see* Min (Chinese)
Swazi (Bantu) 73, **596-7**, 701, 702
Swaziland 73
Sweden 471, 530-2
Swedish 146, 219, 225, **598-9**
Swiss German 219
Switzerland 201, 284, 285, 520-1
Sylheti 42, 86
Syria 26, 28, 546
 Abkhaz 1
 Akkadian 11
 Aramaic 32
 Circassian 136
 Kurdish 340
 Yezidi religion 28
Syriac 27, 28, **31**, **32-3**, 36

Taabwa (Bantu) 73, 85
Ta-ang *see* Palaung
Tabarenho (Portuguese creole) 501
Tabaru (Papuan) 621
Tabasaran (Caucasian) 110
Tachoni (Luyia) 376
Tagalog 49, 51, 92, 112, 247, 484, 583, **600-2**
Tahitian 49, 51, **603**
Tai 52, 61, 107, 296, 346, 348, 552, **605-6**, 623, 698, 699
Taita (Bantu) 73
Taiwan 47, 50, 127, 128, 178, 196-7
Tai Yuan **346-7**, 349
Tajik 493, **494-6**, 666
 see also Tajikistan
Tajiki (Luri) 372
Tajiki (Pamir) 279
Tajikistan 96, 238, 279, 493, 494-6, 526
 see also Tajik

Takestani (Azari) 226
Talysh (Caspian) 226
Tama (Nilo-Saharan) 455
Tamang (Bodic) 562
Tamasheq (Berber) 88-9, **607**
Tamazight (Berber) 88, 89, **608-9**
Tamberma (Niger-Congo) 450
Tamil 156, 169, 393, 502, 559, 560, **610-12**
Tanah Datar (Malayic) 391
Tangale (Chadic) 243
Tanganyika, former 221, 409, 415, 592
Tangkhul (Konyak) 334, 335
Tangut (Burmese-Lolo?) 562, **613**
Tanzania 70, 71, 72, 73, 454, 455
 Cushitic 142, 143
 Luo 369-70
 Makonde 385
 Makua 386
 Masai 409
 Nkore 457-8
 Sandawe 316
 Sukuma 585-6
 Swahili 591-5
 Turkana 646
 Yao 683-4
 Zulu 701, 702
Tapanta (Caucasian) 1
Tarahumara (Uto-Aztec) 439
Tarascan (Amerind) 20, 21
Tarok (Niger-Congo) 452
Tashelhet (Berber) 89, **614**
Tasmania 43
Tasmanian 46
Tat (Persian) 109, 493, 496
Tatar 75, 308, 309, 324, **615**, 648
Tati (Iranian) 496
Tausug (Bisayan) 51, 111, 112
Tautala leaga/lelel (Samoa) 533
Tavgi (Samoyedic) 534, 535
Tavoyan (Arakanese) 105
Teda (Saharan) 453, 456
Teke (Kongo) 73, 326
Telugu 155, 156, 227, 301, 402, **617-18**
Tem (Niger-Congo) 452
Temiar (Austroasiatic) 45, 46
Temirgoi (Circassian) 1
Temne (Niger-Congo) 29, 286, 449, 452, **619**
Tengger (Javanese) 291
Teochow (Chinese) 127
Tepecano (*Mexico*) 437
Tequislatec (*Mexico*) 437
Ternate (Papuan) **620-1**
Teso-Turkana (Nilotic) 455, 456, **622**, 645-6
Tete (Nyanja) 467
Tetela (Bantu) 73
Tetun (Timor-Flores) 51, 392
Thado (Kuki) 333, 334, 335
Thagicu (Bantu) 320
Thai 314, 348, 423, 605, **623-4**
Thailand 539
 Hani 241
 Karen 304-5

- Lisu 359
 Malay 390
 Miao 417
 Mon 423
 Shan 552-3
 Sinhala 559
 So 46
 Tai Yuan 346-7
 Tharaka (Bantu) 73, 320, 415
 Tharelī (Sindhi) 556
 Thonga (Bantu) 636
 Tho (Tai) 605, **625**, 698, 699
 Thracian 275
 Thuri (Nilo-Saharan) 456
 Thuringian 222
 Tibetan 161, 425, 562, **626-8**
 Tibeto-Burman 292, 561, 562
 Tiddim (Kuki-Chin) 333, 334
 Tidore (Papuan) 620, 621
Tigre 81
 Tigrinya (Semitic) 23, 24, 29, 184, **629**
 Timbaro (Cushitic) 143
Timor 50, 51, 392
 Tindi (Caucasian) 53, 110
 Tiriki (Bantu) 71, 375, 376
 Tiv (Bantoid) 70, 450, 452, **630-1**
 Tjamoro *see* Chamorro
 Tlingit (Na-Dené) 434, 435
 Tlisi (Caucasian) 53, 109
 Toba Batak 78
 Tobaru (Papuan) 621
 Tobelo (Papuan) 620, 621
 Tocharian 275, 317, **632**
 Toda (Dravidian) 301
 Tofa (Turkic) 648, 649
Togo 449, 450, 451, 452
 Ewe and Fon (Gbe) 185
 Fulani 206
 German 221
 Hausa 242
 Mõöre 427
 Tok Pisin 179, 221, 490, **633**
 Tolai (Austronesian) 51
 Tolaki (Austronesian) 51
 Toma (Niger-Congo) 452
 Tonga (Bantu) 73, 85, **634**
 Tongan (Polynesian) 49, 51, 191, 321, **635**
 Tontemboan (Austronesian) 51
 Tooro (Bantu) 72
 Toraja (Austronesian) 51
 Torwali (Dardic) 267
 Tosk (Albanian) 14-15
 Totonacan (Amerind) 21
 Toura (Niger-Congo) 450
 Town Bemba 84, 701
 Transcarpathian Ruthenian 528
 Transvaal Sotho 577-8
Transylvania 518, 519
Trieste 567
Trinidad 90, 91, 176
 Tripuri (Bodo-Garo) **94-5**, 562
 Trique (Mixtec) 421, 422
 Truk (Micronesian) 407, 408
 Tsakhur (Caucasian) 110
 Tsakonian (Greek) 233, 234
 Tsangla (Tibetan) 161, 627
 Tsez (Caucasian) 53, 110
 Tsonga (Bantu) 73, 596, **636**, 701, 702
 Tsotsitaal (Afrikaans jargon) 4
 Tsotso (Luyia) 376
 Tsou (Formosan) 197
 Tswa (Bantu) 636
 Tswana (Bantu) 26, 73, 576, 578, **637**
 Tuamotuan (Tahitian) 604
 Tubuai-Rurutu (Tahitian) 604
 Tugen (Nilo-Saharan) 456
 Tugutil (Papuan) 621
 Tujia 105, 562
 Tukanoan (Amerind) 20, 21
 Tulama (Oromo) 478
 Tulu (Dravidian) 156, 301, **639**
 Tumbatu (Swahili) 595
 Tumbuka (Bantu) 73, 467, **640**, 701, 702
 Tung-an (Chinese) 130
 Tungusic 18, 641, **641-3**, 681
Tunisia 26, 28
 Tupí (Amerind) 20, 21, 169, 235, 236, **644**
 Turkana (Nilo-Saharan) 455, 456, **645-6**
Turkey 26, 27, 28, 648
 Abkhaz 1
 Aramaic 32
 Armenian 36, 37
 Bulgarian 101
 Circassian 136
 Greek 231
 Hittite 252
 Judezmo 294
 Kurdish 339, 340
 Laz 419
 Persian 495
 Serbo-Croat 549
 see also Turkish
 Turkic 18, 37, 58, 75, 134, 256, 308-9, 337, 342,
 424, 479, 615, **647-50**, 651, 655, 665, 681
 Turkish 29, 30, 59, 647, 648, **651-4**
 Ach'ar 216, 217
 Albanian 1, 13
 Romani 517
 Romanian 519
 see also Turkey
 Turkmen 59, 648, 651, 654, **655**
Turkmenistan 526
 Turku (Arabic pidgin) 28-9
 Turoyo (Aramaic) 32
 Tuscan 283, 284
 Tuscarora (Iroquoian) 123
 Tuva (Turkic) 648, 649
Tuvalu 321
 Twi (Akan) 10
 Tzeltal (Mayan) 410, 411, 694
 Tzotzil (Mayan) 410, 411, 694
 Tzutuhil (*Mexico*) 20
 Ubangi (Niger-Congo) 67, 215, 449-50, 536

- Ubykh (Circassian) 136
 Udi (Caucasian) 110
 Udmurt 323, 324, **656**, 661
Uganda 71, 72, 453, 454, 455, 456
 Ganda 212-13
 Gujarati 237
 Luo 369-70
 Luyia 375-6
 Nkore 457-8
 Nubi 29
 Swahili 591, 592, 593
 Teso 622
 Turkana 645
 Ugaritic (Semitic) 12, 546
 Ughbug (Caucasian) 110
 Uighur (Turkic) 647, 649, **657-8**, 666
Ukraine 473
 Bulgarian 101
 Polish 498
 Romanian 519
 Russian 525, 527
 Tosk 15
 Yiddish 689
 Ukrainian 82, 83, 361, 528, 564, **659-60**
 Umbundu (Bantu) 72, 412, 413
 Unami (Algonquian) 17
 Ungazik (Eskimo) 181
United Arab Emirates 27, 65, 559
United States
 Alaska 171, 181
 Algonquian 16-17
 American English 170-1
 Amerind 21, 22
 Aramaic 32
 Cherokee 123-4
 Crioulo 14
 Dutch 159
 French 202
 German 220
 Gullah 176
 Hawaii 171, 179
 Louisiana Creole 203
 Miao 416
 Navaho 440
 Norwegian 461
 Portuguese 503
 Samoan 533
 Spanish 582-3
 Tagalog 601
 Yiddish 689, 690
 Ural-Altaic 18
 Uralic 18, 182, 193, 254, 275, 323, 404, 428, 534,
 656, **661-2**
 Urartian 12, 109, 587
 Urdu 29, 30, 65, 237, 248, 266, 306, 311, 486, **663-4**
 Urhobo (Edoid) 163, 452
Uruguay 581, 582, 583
USSR, former 32, 33, 36, 525
 Azerbaijan 59, 60
 Dargwa 148
 Dungan 130
 Kalmyk 425, 426
 Karelian 193
 Kurdish 340
 Kyrgyz 342
 Lithuania 360, 525
 Moldova 519
 Persian 495, 496
 Uzbek 665
 see also Russia
 Ute (Uto-Aztecan) 439
 Utisha (Tiv) 631
 Uto-Aztecan (Amerind) 19, 21, 436
 Uzbek 495, 569, 647, 657, 658, **665-6**
Uzbekistan 329, 526, 655

 Vai (Mande) 448, 452
Vanuatu 93
 Varhadi (Konkani) 402
Vatican City 284
 Vedda (Sinhalese) 560
 Venda (Bantu) 73, 578, **667**
Venezuela 21, 503, 581
Venice 37
 Veps 193, 194
 Vicoli (Sindhi) 556
Vietnam 2, 50, 51, 525, 539
 Cham 118-19
 French 202
 Hani 241
 Miao 416, 417
 Muong 433
 Tai 348, 349
 Tamil 611
 Tho 625
 Vietnam Pidgin 179
 Zhuang 698-9
 Vietnamese 45, 46, **668-9**
 Vili (Kongo) 326
Virgin Islands 160
 Viryal (Chuvash) 135
Visayas Islands (Philippines) 50, 111-12
 Vlach (Romani) 515
 Vlachs 39, 518
 Voltaic (*West Africa*) 449

 Wa 45, 46, 552, **670-2**
 Waali (Niger-Congo) 450
 Wāgdī (*Gujarat*) 238
 Wahgi (Papuan) 491
 Waigali (Nuristani) 268
 Wakhi (Pamir) 279
Wallachia 518, 519
 Wallo (Oromo) 478
 Wanga (Luyia) 376
 Waray-Waray (Bisayan) 51, 112, **673**
 Wardai (Oromo) 478
 Wasa (Akan) 10
 Wellega (Oromo) 478
 Welsh 97, 113, 115, 172, 311-12, **674-6**
 Wendisch *see* Sorbian
 West Aramaic 31
 West Makian (Papuan) 621
 White Tai 349, 605

- Wiyot (Algonquian) 17
 Wobe (Niger-Congo) 451
 Wolaytta (Omotic) 228, 475, **677**
 Wolof (Niger-Congo) 150, 151, 176, 202, 449, 452, 574, **678**
 Woods Cree 17, 139
 Wu (Chinese) 127, 129, 130
 Wuming (Zhuang) 699

 Xhosa 73, 554, 596, **679-80**, 700, 701, 702
 Xiang (Chinese) 127, 130
 Xibo (Colloquial Manchu) 641

 Yagnobi (Pamir) 279, 569
 Yahudi (Arabic) 28
 Yaka (Bantu) 73
 Yakut (Turkic) 647, **681-2**
 Yalunka (Mandekan) 398, 590
 Yanomami (Amerind) 21
 Yao (Austro-Tai) 52, 355, 417, **685**
 Yao (Bantu) 73, 386, 466, **683-4**, 701, 702
 Yaouré (Niger-Congo) 451
 Yapese (Austronesian) 120
 Yas (Papuan) 491
 Yekhee (Edo) 162-3
 Yemba-Nwe (Niger-Congo) 452
Yemen 27
 Yemeni 25
 Yenisei Ostyak (Palaeosiberian) 481
 Yerwa Kanuri 303
 Yi 61, 104, 344, 441, 562, 606, **686-7**
 Yiddish 222, 225, 245, **688-91**
 see also Jews
 Yimbe (Niger-Congo) 451, 619
 Yorkshire 331
 Yoruba 29, 242, 331, 450, 452, **692-3**
 Yucatec (Mayan) 20, 21, 411, 437, **694-5**
 Yuè (Cantonese) 127, 129, 130, 685
Yugoslavia, former 13, 378, 525, 549, 564

 Yukagir (Palaeosiberian) 481
 Yupik (Eskimo-Aleut) 180, 181, 481
 Yurak (Samoyedic) 534-5
 Yurok (Algonquian) 17

 Zaghawa (Saharan) 453, 456
Zaire 158, 591
 Zalamo (Bantu) 73, 594
Zambia 70, 71, 72, 701, 702
 Bemba 84
 Chokwe 132, 133
 Lozi 363
 Lunda 367-8
 Luvale 373
 Nyanja 466
 Tonga 634
 Tumbuka 640
 Zamboangueno (Spanish creole) 583, 584
 Zande (Ubangi) 449, 452, 537, **696**
 Zanskari (Tibetan) 628
Zanzibar 156, 591, 593, 595, 611
 Zapotec (Amerind) 21, 422, 694, **697**
 Zarma (Songhay) 572
 Zaza (*Iran*) 340
 Zemaiciai (Lithuanian) 64
 Zenaga (Berber) 89
 Zezuru (Shona) 554, 555
 Zhuang (Tai) 107, 605, 625, **698-9**
 Zigula (Bantu) 73, 594
Zimbabwe 4, 71, 72, 73
 Lozi 363-4
 Ndebele 443
 Nyanja 466-7
 Shona 554-5
 Tonga 85
 Zinza (Bantu) 72, 457, 458, 586
 Zitako (Nupe) 465
 Zome (Kuki-Chin) 334
 Zoque (Amerind) 21

Acknowledgements

First of all, I am grateful to the authors whose explanations and examples have been quoted throughout, either because they help to demonstrate a special feature of the language under discussion, or because they relate to themes which emerge from the book as a whole – the multiple social uses of language and oral literature, and the complex ways in which languages have interacted with one another. I am also grateful to Gamma Productions, Vijay K. Patel and other makers of TrueType fonts used in the alphabet tables. The maps are based on outline maps produced by a shareware program, Clpmap, developed by W. W. Mayfield, 23219 Audrey Avenue, Torrance, California 90505.

The London Goodenough Trust for Overseas Graduates and the Institute of Linguists gave me the impulse to begin the *Dictionary of Languages*. Maureen, Elizabeth and Rachel excused me many hours of neglect as I worked on from Abkhaz to Zulu. It was a struggle to mould a shapeless text into a shapely reference book: somehow, Sarah Prest and others at Bloomsbury Publishing managed it.