DICTIONARY OF
LANGUAGES

ANDREW DALBY

THE DEFINITIVE
REFERENCE TO MORE THAN
400 LANGUAGES

DICTIONARY
OF LANGUAGES

DICTIONARY
OF LANGUAGES

The Definitive Reference to more than 400 Languages

Andrew Dalby

Preface

The language and language family headings in this book are in alphabetical order from Abkhaz to Zulu. Cross-references are given in SMALL CAPITALS. Maps, and sometimes boxes listing numerals or other examples, often bring together information on two or three related languages: the cross-references always serve as a guide.

It has usually been possible to give at least the numerals, 1 to 10, as an example of the way a language looks and sounds. Other information often displayed adjacent to the text includes foreign scripts and their equivalents in the familiar Latin alphabet. A surprising number of these scripts can now be found as TrueType fonts on the World Wide Web (see acknowledgements on p. 734).

This book is not designed as a bibliography or reading list. Often, however, information and examples in the language entries are drawn from sources to which an interested reader could go to find out more. Thus, wherever it may be useful, full references to sources have been given.

Putting sounds on paper

No ordinarily used writing system is adequate for recording all the sounds of any and all human languages. Alphabets as short as the Greek (24 letters) or the familiar Latin alphabet (26 letters) are not fully adequate even for most single languages. English, for example, by the usual count has about 40 'phonemes' or structurally distinct sounds.

Linguists therefore use special extended alphabets to record pronunciation precisely. The commonest is the International Phonetic Alphabet (IPA). Specialists in some language families have their own conventional alphabets and signs (see box at MORDVIN for an example).

Since the IPA has to be learnt, and this book is intended for non-specialists, the IPA has not been used here.

Languages that are usually written in the Latin alphabet are written here according to their usual spelling. For Chinese the official Pinyin transliteration is given, and for the languages of India I have kept close to the standard agreed at an Orientalist Congress a century ago. Languages usually written in other alphabets have been transliterated into Latin, giving the consonants the sounds they usually have in English, and the vowels the following sounds:

a   like *a* in English *father*
i   like *i* in English *machine*
u   like *oo* in English *boot*
e   midway between *ea* of *bear* and *i* of *machine*
o   midway between *oa* of *boar* and *oo* of *boot*

An additional consonant is familiar from nonstandard English:

'   this apostrophe is often used for a glottal stop, the consonant that replaces *t* as the third sound in the London colloquial pronunciation of *butter*.
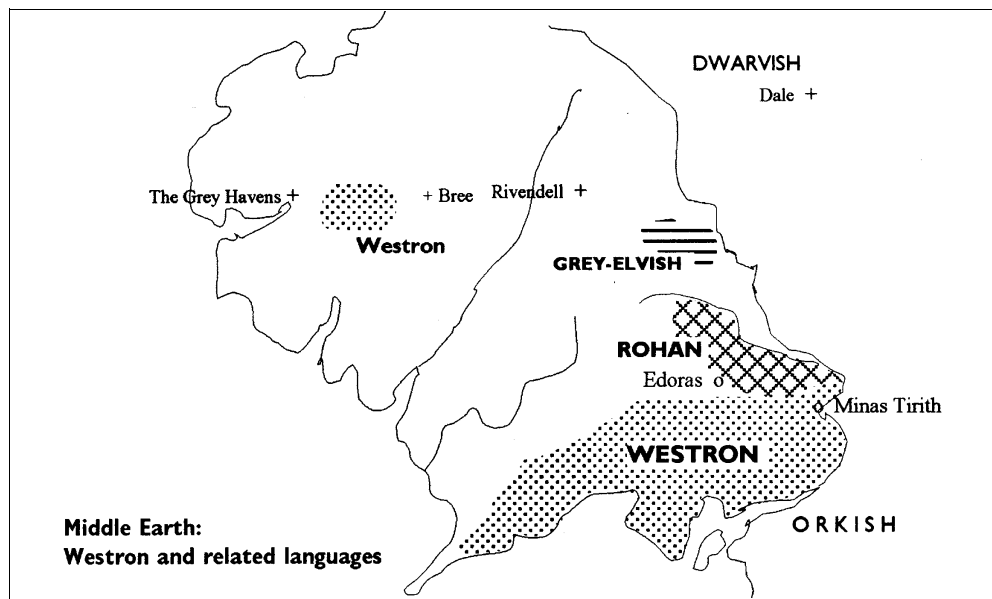
Three symbols have been borrowed from the IPA for sounds that are not easily distinguished otherwise:

ə   Often called by its Hebrew name *schwa*, this is the second vowel of English *father*
ɛ   The open *e* sound of English *bear*
ɔ   The open *o* sound of English *boar*

Three additional symbols, familiar in German and Turkish, have been used frequently in this book for sounds not found in English but common in many other languages:

ı   The vowel of Russian мы 'we'. To imitate it, say 'ugh' while gritting your teeth
ö   The vowel of French *coeur* 'heart' and German *hör* 'listen!' Make the 'uhh' sound of hesitation while rounding your lips
ü   The vowel of French *mur* 'wall' and German *für* 'for'. Say 'ee' while pursing your lips tightly

A dot below a consonant usually makes it a *retroflex consonant*, one that is formed with the tongue turned back towards the roof of the mouth – these are the sounds that help to typify an 'Indian accent' in speaking English, and they are indeed found in most Indian languages.

DWARVISH
Dale +

The Grey Havens + ⋯ + Bree Rivendell +

Westron

GREY-ELVISH ═

ROHAN
Edoras ○

◇ Minas Tirith

WESTRON

O R K I S H

Middle Earth:
Westron and related languages

A line above a vowel makes it long. An acute accent on a vowel means that stress falls on that syllable. In tonal languages, however, these and similar signs have sometimes been used to mark tones: ˉ for high level, ´ for rising, ` for falling, and an underline for low level tone.

In general, to make easier reading, words usually written with a forest of accents are accented only on the first occasion that they are used.

The statistics

Unless otherwise stated, the figures in this book give the number of 'native' or 'mother tongue' speakers for each language. Some, from English and French to Amharic and Tagalog, are spoken by many millions more as second languages. This is one reason why statistics in different reference books may seem to conflict (see also 'Facts, real facts and statistics', p. xiii).

The maps

Language boundaries are not like national boundaries: languages spread, and overlap, in a way that only very detailed statistical maps can show accurately. The two hundred maps in this book show simply and clearly where each language is spoken and, if possible, its nearest neighbours. Nearly all the maps are drawn to a standard scale – 320 miles to the inch. Just as the statistics allow a comparison of native speakers for each language, so the maps show what area of the earth each language covers.

Each map deals with a language or language group, and these are named in bold face. Shading indicates the main areas where these languages are spoken. Isolated places where the same languages can be heard are marked with a cross. Other neighbouring languages, not closely related to these, are named in lighter type. Major cities are marked by a circle. As an example, this map of J. R. R. Tolkien's Middle Earth shades the areas where Westron, Grey-Elvish and the language of Rohan are spoken.

There are a few smaller scale maps (for example, the map of 'Language families of the world' on pp. xii–xiii). All these are at the scale of 1,000 miles to the inch. Italic face is used in lettering these maps as an eye-catching reminder of the difference in scale.

INTRODUCTION

These are the major languages of the 21st century – their history, their geography and the way they interact. Astonishingly, no other book in English brings them together in this way.

The world has many more languages than these. From over five thousand that are spoken in 2004, a selected four hundred languages and language groups have entries here. Many more can be found in the index, but they are still only a minority of the total number of living languages.

Every language is a unique and uniquely important way to make sense of the world; but a choice had to be made. The languages selected here are those spoken by the great majority of the people of the world. These are the languages that the 21st century needs to know about *first*: national languages of independent countries, languages of important minorities that will make news, classical languages of the past. Most entries are for languages with more than a million speakers. Some, such as English, Spanish, Arabic, Russian and Chinese, are spoken by hundreds of millions.

All the languages that have *not* been given entries in this book have fewer than a million speakers. Some have only one or two speakers, and many that were until recently spoken by thriving communities are now extinct. This is an accelerating trend. It is easy to foresee a time, perhaps a hundred years ahead or less, when most of the languages left out of this book will not be spoken at all, and when many of those included will – so to speak – be struggling for speakers.

As a language falls out of use, one of those unique ways of making sense of the world is lost.

Why languages grow apart

All 'living languages' or 'mother tongues' – all the languages that children learn when they first learn to speak – are continually changing. The change happens in at least two ways: for language change comes from the very nature of childhood learning, and also from the demands that we make, throughout life, on the astonishingly flexible medium of communication that language is.

Look first at the way children learn to speak.

Language is a palette of sounds, a dictionary of words made up of those sounds, and a grammar of rules for combining the words meaningfully. Usually we are unaware of the making of sounds, the choosing of words and the applying of rules, yet this is how we speak and this is how we understand what others say. Every child that learns to speak practises sounds, builds up a dictionary, and works out a set of rules. Every child does all this largely unconsciously, with incomplete help and unreliable guidance from parents and friends and teachers who, themselves, are only half conscious of the rules. Every child does this *afresh*. The range of people from whom each child learns is different. And children are not clones of one another.

Thus everyone's sound patterns, everyone's dictionary and everyone's language rules are original and slightly different from everyone else's. This is how change and originality are built into the nature of human language learning. And this, incidentally, is why the 'grammar of a language', as opposed to the grammar of a single person's speech, is an abstract formulation – a highly useful one that we simply cannot do without when we want a standard language, or a foreign language, to be taught.

Since those living in a community interact most with others in the same community, everyday speech in any one community tends, over time, to diverge from that in others. We notice the differences: we talk of the 'accent' or of the 'dialect' of those whose speech uses identifiably different sound patterns, different words or different grammatical rules – though still so close to our own that we can understand it easily. Australian English and British English have grown apart, quite distinctly, in little over a hundred years.

Languages: how many days apart?

'The impressionistic statistics of glottochronology [see glossary] are nothing new. Speakers of AKAN are said to evaluate the closeness of languages by how long it takes a speaker of one to learn the other. Asante is a one-day language from Fante: EWE is further off from either.'

C. F. Hockett, *A course in modern linguistics*
(London: Macmillan, 1958) pp. 326–7

When most people lived all their lives in a single community, travelling little and seldom meeting outsiders, the language of a region – which might once have been 'the same' language – steadily differentiated into distinct local dialects. As the process continued, the dialects eventually became so different that speakers of one could not understand speakers of another. When this point is reached, then – by one often-used definition – we are not dealing with separate dialects any longer but with separate languages.

This is the single overriding reason for the great number of languages in the world today, and the effect is well demonstrated by the fact that a great many very different languages, each with small numbers of speakers, tend to be found in mountainous regions where communications are difficult, such as the Caucasus and the southern valleys of the Himalayas.

Why languages converge

But there are influences in the opposite direction too. If those who travel for study or work, and those who pay attention to press, radio and television, begin to make up a large proportion of the population, they will limit the tendency for community dialects to diverge. And those who want to make a good impression – as examination candidates or job applicants or employees or traders or politicians or preachers – have to limit the extent to which their own accent or dialect or choice of words will distract others from their message.

Thus the second process of language change takes effect – in which older children and adults continually adjust their speech to their hearers'

expectations, in order to get a message across. Speakers pick up new words, new phrases and new tones of voice from those around them. They imitate not only others who are speaking 'the same' language, but also those speaking a different dialect (perhaps a more prestigious one, the dialect of a capital city or a university), or indeed a completely different language (perhaps the language used in government or in an army or in business). We can, of course, learn several languages and keep them apart. But in practice we also need to mix them. The English that is standard in India has always differed in vocabulary from British English: the special vocabulary is naturally used, just as it was under British rule, in speaking of concepts that belong to the politics and way of life of the subcontinent.

How many dialects, how many dialect speakers make up a language? There is no answer to this question. The hundreds of millions of speakers of English speak it in very different ways, but all recognise what is in practical terms the same standard for writing and formal speech. Some African standard languages, such as TSONGA and Ronga, or EWE and Fon, differ from one another far less than do the local dialects of English: yet, given no overall literary or political unity, those who first devised their written forms had no basis on which to develop a standard language covering several dialects, while the speakers of any one dialect had no reason to attend to, or to respect, those who used an unfamiliar dialect.

Tracing language history

Every language displays, to the practised eye or ear, some of its own history: words clearly borrowed from other languages; voice inflections, and turns of phrase, that seem to be shared by two adjacent but otherwise very different languages. These are among the phenomena of language convergence.

The more pervasive, more regular phenomena of language change – those of divergence – can also be traced and reconstructed.

Historical linguistics uses three forms of evidence. Comparisons are pursued between two or more known languages that appear to share a

range of vocabulary and a set of grammatical rules. Written texts, preserved from an earlier period, are explored to reconstruct the sounds and patterns of the spoken language that lay behind them. The recorded history of peoples and their migrations is searched for possibilities as to when languages diverged and how they reached their known locations on the map. If all three forms of evidence are available, all three will be used and they will all act as controls on one another.

One result of this kind of interactive research is that it gives a clearer understanding of human history – of the way that people have 'constructed' their own 'identity' (as we might say now), linguistic, cultural, tribal, national, ethnic, racial. Where did they think that they belonged? What was their view of those who did not belong? What was the upshot – between cultural mergers and wars of extermination?

There are other results. The older written texts will be better understood, or understood for the first time. The recorded history will be reinterpreted, and something will be learnt of the silent majority whose lives and travels do not get into recorded history. A language history will be built up, tracing the two or more known languages back to the 'ancestral' language from which they diverged.

The next step seems to follow naturally. Even if the ancestral language was never recorded – no written texts – it now becomes conceivable to reconstruct it. This means building up a sound system, a grammatical system and a list of words and meanings which, after language change and divergence, would have resulted in the forms that actually exist in the two or more languages from which the investigation began.

In scholarly work, by convention, reconstructed words from unrecorded forms of language are marked with a *. The box at INDO-EUROPEAN gives some examples.

The reconstruction is a 'formula'. It may explain the known forms; it may do so in the neatest possible way; it may be open to confirmation or disproof when a third or fourth related language is brought into the comparison. Still, if there are no written records, nothing can prove that it represents a real language that was actually spoken at some time.

How to use languages

In a sense, there is no such thing as a 'flair for languages' – or, if there is, we all have it. If you have learned to speak your mother tongue, you have proved that you have the ability to learn languages.

But you spent a long time learning it. If you were lucky enough to grow up bilingual, you have spent a long time learning both languages. Not necessarily a long time in the classroom – not necessarily any time at all in formal language learning – but a long time, several solid years by any count, listening and speaking, sometimes being corrected, more often correcting yourself.

Mithradates: father of multilingualism

In 1784 Catherine the Great, Empress of Russia, initiated a research project to collect lists of about 225 common words from the languages of the world, and especially from the Russian Empire. The lists were eventually published in Adelung and Vater's compilation *Mithridates oder allgemeine Sprachkunde*.

Any classically educated nineteenth-century reader would have known why the book was called *Mithridates*. King Mithradates Eupator of Pontus (132–63 BC), who fought against the Romans and was defeated by Pompey, spoke twenty-two languages. He is the first historical figure famous for multilingual skills.

The way to retain the ability to learn languages is to go on using it. The younger that children are when they learn a second language, the more easily they will learn it, and the third language will come easier still. Unfortunately, youth isn't everything. They also have to *need* to learn the new language, and to *need* to practise it. In learning languages we are harnessing a skill that is inborn in human beings – but laziness, the least effort for the most reward, is also inborn in us, and if we can get by with one language, we will. In many countries in the world children are now growing up trilingual: they learn a local mother tongue, then a national language, then English. They are not three times as intelligent as children whose

mother tongue is English: but they *need* to use these languages, in successive stages of education, and they *need* to practise their linguistic skills when reading and watching television, when going about a city, and when dealing with businesses and government offices. Most English-speaking children, in Britain and North America and some other countries, can get by entirely with English: so most of them do, and relatively few of them ever learn a second language really well.

When learning a language, as a child or an adult, we need to know why we are learning it – and we need to practise it.

That the country needs linguists, and that a school curriculum demands a foreign language, are both good reasons for learning a foreign language – but they may not be quite good enough to overcome the laziness. Learning a language is hard work. That one needs to use the language in everyday life is the best reason, and the best opportunity, for learning it. Children, or college students, from a monolingual country have the best opportunity to learn a foreign language if they live abroad. And it is only a minority of British or North American children who do this.

So far we have talked of 'learning a language' – meaning, I suppose, learning to speak it, to understand it when spoken, to write it, to read it and to know something of the culture that underlies the written word. The learning process may continue until one speaks the language 'like a native'. It is a process that never ends: in a foreign language, as in one's own, there are always new words, and whole new sub-cultures, just about to be invented or patiently waiting to be discovered.

But apart from 'learning a language' whole, we also have the ability to learn a language selectively, and this ability is worth cultivating. It is used, most obviously, by those who learn a classical language. Many people, in many countries, learn Latin or classical Greek or Pali or Sanskrit. They learn to read the literature of those languages. Very few of them learn to speak the languages fluently: for most, that would be a useless skill.

Many develop this kind of knowledge of modern languages too. 'Language for special purposes' is now a recognised field of teaching – for business purposes, or for the ability to assimilate a technical text in one's own specialised field, or simply to understand essentials and make oneself understood when travelling.

But a good many linguists will agree that there is a threshold beyond which the learning of a language seems to develop a momentum of its own. Even after one has begun to learn a language for a very simple or specialised reason, the fascination of understanding more and more of a foreign way of life, its culture and its literature, takes hold.

The names of languages

Most people who speak English can happily call it 'English'. Though England is the name of a geographical region, it is not the name of a nation state, and speakers do not feel excluded or politicised by the term 'English'.

Language and political theory

The old Soviet Union was in some ways relatively enlightened in its handling of minority languages and their statistics. But the concept of *Language of the USSR* caused serious anomalies. German, spoken by millions in the USSR, was excluded from lists and from privileges because it was the national language of another sovereign state. The language of Moldavia had to be called 'Moldavian' to make it a separate language from Romanian. Tajik had to be similarly classed as a separate language from Persian-Dari. On the other hand Yiddish, Romani, Kurdish and Aramaic were allowed the all-important status of Language of the USSR because, even if the majority of their speakers lived in other countries, they had not the status of national languages in those countries.

The European Union has a similar ideological problem. Its Office of Minority Languages has to call Albanian *Arbëresh* when it is spoken in Italy and *Arvanitika* when it is spoken in Greece.

With many other languages it is not so simple. Until the 1940s the lingua franca of the south-east Asian archipelago was called 'Malay' by nearly everybody. For newly independent Indonesia the

term had been found unsuitable because of its connection with Malaya, still British-ruled. So the form of Malay that became the national language of Indonesia had to be called 'Indonesian' – and independent Malaysia, incorporating Malaya and three other British territories, had to follow suit and call its language 'Malaysian'. It is still 'Malay' in Singapore and Brunei.

Thus language names often carry a political charge. In this book the headings chosen will not please everybody. I have tried, however, to be uncontroversial. For national or minority languages of a single state I have usually chosen the current official name used in that state, or an obvious English equivalent. For languages that are more widely spoken I have preferred a neutral term if any exists, and I have always tried to explain, and index, the different names that are in use. Sometimes a language will have a different name in each of the different countries in which it is used.

Linguists (like other social scientists) love to invent words. Sometimes they have done so to solve this very problem: see MANDEKAN for an example. More often, linguists have invented names for language groups and families – and they have felt free to change the names whenever their view of a language relationship changes. So we have Semito-Hamitic, Hamito-Semitic, AFROASIATIC, Afrasian and Erythraic, all as alternative names for the same language family. The headings chosen here for language families are not intended to promote any particular view, but are in general the most widely used.

In some schools of linguistic research there is a custom of designing the names of proposed language groups to match a hierarchy – just like the different Latin terminations used by botanists and zoologists to distinguish sub-families, families and orders. Evidently such hierarchies are a useful tool for botanists and zoologists. For linguists they are more misleading, because language relationships do not work like that. I have not always bothered to mention such designer terms as *Hellenic* for GREEK and *Bodic* and *Bodish* for TIBETAN and its relatives (ugly names, these). I have used the terms 'family', 'branch' and 'group' without trying to pin them down too specifically, but I have generally called the most inclusive, generally recognised language

groupings 'families': I have not used the terms 'stock' and 'phylum' that some linguists prefer for designating very large groupings.

Facts, real facts and statistics

At the head of each entry an estimate of number of speakers appears. This is intended as a rough estimate of the number of people for whom this is the mother tongue or first language. The figures must be treated with suspicion.

Some of them come from national censuses. Are they accurate? That depends on what question was asked, how it was understood, and, besides, on a whole range of more emotive issues. In some countries there may be a cachet in claiming to adhere to a minority language which is actually falling out of use. In Ireland the IRISH language is a national symbol. Of the 1,000,000 who say they know it, how many can or do use it regularly? In many countries where nationalism is to the fore it may be safest to claim to speak the majority language even when one uses another mother tongue nearly all the time. This will swell the figures for languages like GREEK and TURKISH. In others again, minority areas may not be reached by any census.

Other figures – especially in countries where minority statistics are not officially published, or where certain minorities are not officially admitted to exist – come from non-official social and linguistic research. Usually these figures are extrapolated from sample surveys or from localised fieldwork. Sometimes they will turn out to go back to nothing more than hearsay. See AZERI (language of Azerbaijan, also spoken by a minority of unknown size in Iran) for an example of the resulting variation.

Language families of the world

If humans are genetically endowed with language, then, logically, all languages are related. This doesn't mean the relationship can be traced. This listing and the map on pp. xii–xiii show the families, and the single languages, mentioned in this book which have not yet been convincingly shown to be related to one another. When all languages have been shown to be

INDEPENDENT LANGUAGE FAMILIES OF THE WORLD

related, a revised version of this map and list will have only one entry.

Single languages which have not been proved to be related to any others are known as 'linguistic isolates'. Several are included here. But in the present state of our knowledge it is not sensible to try to draw a complete map of linguistic isolates. In addition to the few that are well known in linguistic literature (Ainu, Basque, Burushaski, Ket, etc.) there are many little-studied languages in Asia, Africa, New Guinea and the Americas for which no linguistic relationships have been discovered. Only a few of these will really be isolates. Most of them will eventually be shown to belong to one or other known family – if ever they are fully recorded and investigated before they cease to be spoken.

AFROASIATIC LANGUAGES

Ainu (see JAPANESE)

ALTAIC LANGUAGES, probably including TURKIC, MONGOLIAN and Tungusic languages (see MANCHU) and perhaps KOREAN and JAPANESE

AMERIND LANGUAGES, a family grouping that remains highly controversial, may perhaps include ALGONQUIAN LANGUAGES, ARAUCANIAN, AYMARA, Iroquoian languages (see CHEROKEE), MAYAN LANGUAGES, QUECHUA, Uto-Aztecan languages (see NAHUATL) and many others

Angan languages (see PAPUAN LANGUAGES)

AUSTRALIAN LANGUAGES

AUSTROASIATIC LANGUAGES

T

NA–DENE

A M E R I N D

A I

Questions and answers

Most of the time, our use of language is uncon-
scious: we say what we mean, and understand
what others mean, without concentrating on the
sounds or the individual words or their grammar
– and we learnt to do most of this unconsciously,
'instinctively'. This leaves a surprisingly wide
field for misunderstandings and misstatements
about the way language works.

Is language change wrong? You're entitled to your moral views. But change is built into language, into the way we learn it and the way we use it. It cannot be prevented.

Do some people speak ungrammatically? 'Grammar' is the sequence of rules through which human speech is produced. So the answer is no. We all have a built-in grammar, or several grammars, for the different languages and speech registers that we use. But when you are beginning a new language, your grammar may at first be so different from the one you are aiming at that no one can understand you . . .

Should parents correct children's speech? Yes: by example. Children have to learn to interact with others effectively.

Should parents teach children a different way of speaking from their own? Probably not, unless the parents can speak it fluently.

Should teachers correct children's speech? In every country children need to know at least one standard language if they are to succeed in everyday adult life. Schools that do not teach a standard language are failing in a crucial part of their job. Schools that punish children for using the language they learnt at home – whether Welsh, Black English, a Sign Language (just three examples from recent history) or any other – are also failing them. Human beings are naturally multilingual. No one needs to speak the same language in a job interview as when chatting with friends or family. A school's primary linguistic task is to *add* the standard language of their country to children's developing linguistic skills.

Should schools teach grammar? Yes. We *can* learn to speak without learning any grammar formally. But to learn to speak and write our standard language, or a foreign language, effectively, we need a basic understanding of grammar.

Are some languages unable to deal with modern civilisation? People use the languages they need. If a new skill, such as car maintenance or abstract mathematics or spice cookery, is demanded of the speakers of a language which has no technical terms for such things, they will learn and use the necessary words.

Are some languages more difficult than others? In the abstract, no. Whatever language they learn from those around them, whether it is English or Chinese or Nahuatl, by the age of about seven children have learnt it pretty well. Writing systems, which are conscious inventions, differ from natural language. Chinese script really is much more difficult to learn than alphabets like Latin or Arabic.

When we learn a new language in later life, it will be easier if it is close in structure and vocabulary to one we already know. There are some languages that even practised linguists have found unusually difficult – BASQUE and KHOISAN LANGUAGES among them.

Should we speak the way we write? Historically, speech comes first and is the most natural use of language. We need to speak to inform, interest and persuade others. Doing this effectively often means using a style that is different from our written language.

Is it difficult for children to grow up bilingual? In most communities it is necessary and natural. In some, notably among English speakers in Britain and the United States, bilingualism is very rare and quite difficult to achieve, not because children are less intelligent or less gifted linguistically, but because the environment they grow up in is not naturally bilingual.

Does our language affect the way we think? Yes. But anyone can learn another language.

The languages that die

The ever repeating news story of the early 21st century is the disappearance of languages. The phenomenon is often described as 'language death' – an easily understood metaphor. More emotive phrases, such as 'linguistic genocide', have been used too.

Sometimes real genocide is involved. In some countries, even now, those who find inconvenience in the continued existence of a small community speaking a strange language and following strange customs may be able to kill them all with impunity. But often the death of a language is a more gentle affair. This is how it happens. The older generation decides, family by family, that children will be brought up to speak a new language and not to speak the older community language. This decision is made without full awareness of the long-term

consequences, but it is made for serious reasons. In the past – and even now in many countries – children have been punished and ridiculed by teachers for speaking the wrong language at school. In some countries adults are breaking the law if they speak the wrong language in public. Almost everywhere in the world, the better you speak a national or international language, the more you will prosper. It is reasonable to make plans so that your children need not face ridicule and will be more prosperous than you have been.

The long-term consequence, of which most parents are not fully aware when they make their decision, is this: if all families in a linguistic community make the same choice, then the language will disappear for good when the older generation dies.

Those who are most closely involved – the last active speakers of the disappearing language – eventually realise, too late, that they are participating in the loss of a cultural resource. Linguists and anthropologists whose fate it is to work with disappearing languages are all too familiar with the sense of sadness felt by the last speakers, in the last years of their long lives, as they reflect that so much of their culture will vanish when they die. Some of these people work very hard, with or without scholarly amanuenses, to put their knowledge on record. Imekanu, one of the last speakers of Ainu (p. 288) and a skilled oral poet, set to work after her retirement from work as a Christian missionary and wrote down 20,000 pages of hitherto unrecorded Ainu epic poems.

Such sadness, for the present at any rate, is confined to grandparents and to those without children. Parents and children seem to have no time for it. That is why the decline in the total number of languages seems likely to continue at the same rate as in the 1980s and 1990s. It was estimated then that one language 'dies' every two weeks. If this pattern continues, by the year 2100 there will be only 2,500 languages in the world. Several of the languages with main entries in this book will already have gone, and speakers of many of the remainder will be aware that their language, too, is threatened. If the pattern continues for another century, only one language will be left. But will it come to that?

Curing 'language death'

Once a language disappears, it is very difficult – bordering on impossible – to bring it back.

For this, linguists must take some of the blame, because in spite of all their efforts they have never yet been able to describe any language fully. Some grammars fill several volumes; yet, however thorough the grammar, a native speaker can always beat it. There are English words even now omitted from the twenty packed volumes of the *Oxford English Dictionary*. And no phonetic description has quite done justice to the variations in sound quality, pitch and prosody that a community of speakers produce in their daily conversation. It may be true none the less that the work of hundreds of thousands of linguists has brought us relatively close to achieving a full description of some forms of English; by contrast, we are a long way from such a full description of most other languages. In particular, we have nothing like a full description – in many cases no description at all – of many of the languages that are in immediate danger of disappearing from use.

Even if we had full descriptions to refer to, learning a language wholly from books is not the easy way, and certainly not the way that most people can do it. The easy way – it takes a long time, but time passes – is to learn it from speakers. The easiest of all is to learn it as a child, from parents and other children. Unfortunately, if there are no longer any speakers, these ways of learning a language are no longer on offer.

These are the reasons why reviving a 'dead' language is very difficult. People do it, none the less. Hebrew was revived – but Hebrew was a very special case, because it was still widely studied at school and used in ritual. The settlers of the new state of Israel had usually learned some classical Hebrew and had no other shared language. They needed a lingua franca, and it might as well be Hebrew. It eventually became a mother tongue once more (as any lingua franca may do) and it is now the majority language of Israel. This is the only case so far in which language revival has gone all the way.

If the revivalists have more modest ambitions they may well succeed after a fashion. Since the

mid-19th century many previously unremembered languages have been revived to the extent that students can read the texts once more – Egyptian, Akkadian, Hittite, Sumerian and most recently classical Maya, to name only the most familiar examples. In classrooms and on scholars' desks, these languages have been brought back to life.

And then there are the languages, a rapidly growing number of them, that are revived to the point where they are actually used in a limited way by a whole community. The obvious case is Irish. Irish never died in the Gaeltacht, but in most of Ireland it had entirely ceased to be used at the beginning of the 20th century. Extensive and continuing efforts have ensured that practically everyone in Ireland treats Irish as a national symbol, listens to it and understands at least some of it, and reads Irish script easily. It's true that you can get by in Ireland without knowing any Irish; still, Irish is an integral part of Ireland's general linguistic culture. A century ago, few if any would have predicted this; Irish was widely assumed to be near death. Those who are reviving Cornish (which, unlike Irish, had completely ceased to be spoken) are hoping for similar success; so are the many Native American peoples who are now participating in movements to revive their lost or almost-lost indigenous languages.

It depends what you mean by revival. Cornish (p. 113) was not especially well recorded before it ceased to be spoken; no one really knows how close the sounds and grammar of the new schoolroom Cornish are to the sounds and grammar of the 16th century, when the language was still vigorous. Purists have observed that some of the revived Amerind languages now heard in schoolrooms and summer camps are in the nature of English-Amerind creoles, lacking the full grammatical and phonetic structure of the traditional languages; this is especially the case if, as with Cornish, the spoken tradition was broken and modern learners have no chance to hear the language spoken as it used to be. Whatever the differences, however, these revived languages, like Hebrew and Irish, have the potential to become community languages once more, symbols of a shared culture that was almost lost.

Reversing language decline

If we cannot do what the founders of modern Israel did (adopt for everyday use a language that was already widely learned at school) then we had better do what the makers of independent Ireland have done (reverse the decline of a language that was still spoken). Either of these strategies is much easier than reviving a language that had been lost, like Cornish.

This is what I mean by easy. If we are to reverse the decline in any minority language, we simply have to persuade national and local governments (a) not to treat public use of the language as a threat to national unity, and (b) to accept that the language requires legal and financial support to balance the tacit support that the current national language receives. Then we must persuade teachers and social workers (a) not to patronise, disadvantage or humiliate those who speak the language, and (b) to learn it themselves. We must persuade public services to (a) welcome use of the language in their contacts with the public, and (b) give it equal or higher status in their own communications. We must persuade journalists and politicians that young children are must better at learning languages than they themselves are. And then we will hav a chance of persuading parents (a) that given the need, young children will as readily learn two languages as one, (b) that youngsters who are bilingual or multilingual have more opportunities than others, and (c) that their children will be better and happier for having the opportunity to participate in their own traditional culture. We may complete all the other stages, but unless we complete this last stage – persuading the parents – we will not be able to reverse the decline of the language. Governments and teachers cannot make a language live.

So it isn't easy after all. The opinions that need changing are not held universally, and not in every country, but they are widely held and deeply ingrained. If we can change them, we will reverse the decline of minority languages. If we can do that, we may succed in transmitting to future generations the rich multi-linguistic culture mapped in the following pages.

Andrew Dalby

DICTIONARY
OF LANGUAGES

ABKHAZ AND ABAZA

300,000 SPEAKERS

Georgia, Turkey, Russia

A group of five dialects, belonging to the family of North West CAUCASIAN LANGUAGES, is spoken in the north-western extremity of Georgia, among mountains that slope steeply down to the Black Sea coast, and across the Caucasian watershed in the Cherkess republic of Russia. There may be 100,000 or more Abkhaz speakers in Turkey, and others in Syria, whose ancestors had fled to the Ottoman Empire to escape Christian rule when Russia conquered this part of the Caucasus in 1864. Since Georgia became independent from the Soviet Union, Abkhaz speakers (though a minority in the administrative district of Abkhazia) have been fighting the Georgians for their own independence, with military support from Russia.

The dialects of Georgia are conventionally classed as 'Abkhaz', and many speakers are bilingual in Georgian, which is now the national language. Speakers of the southern Abzhui dialect are frequently trilingual, able to converse in MINGRELIAN as well. The dialects in Cherkessia are 'Abaza': speakers here are bilingual in the related CIRCASSIAN.

Literary Abkhaz is based on the Abzhui dialect spoken at the capital, Sukhumi. It has fewer consonant phonemes than Bzyp' and is thus said to be slightly easier for non-natives to learn. The literary form of Abaza is based on the Tapanta dialect.

Abkhaz was occasionally written in Cyrillic script before the Russian Revolution: Latin, Georgian, a revised Cyrillic and (since the end of the Soviet Union) a revised Georgian script have been used in succession. To cope with its great number of consonants, 14 extra consonant letters were added to the Abkhaz variant of the Cyrillic alphabet.

For a table of numerals see CIRCASSIAN.

North West Caucasian languages on the map

Abkhaz and *Abaza* together make up four dialects: Bzyp', Abzhui or Abzhuwa, Ashkhar or Ashkharwa and Tapanta, the last being rather different from the rest. The dialects spoken in Georgia are grouped as Abkhaz and have about 90,000 speakers. The dialects in Cherkessia (Ashkharwa and Tapanta) are counted as Abaza and have about 60,000 speakers.

Adyge and *Kabardian* or Kabardo-Cherkess are the languages of Cherkessia and neighbouring districts, though most speakers are to be found in Turkey: together these languages are called CIRCASSIAN. There are four Adyge dialects, Bzhedugh, Shapsugh, Abzakh and Temirgoi – but Shapsugh and Abzakh have almost disappeared from Russia through emigration.

ACHEHNESE

2,000,000 SPEAKERS

Indonesia

Achehnese is spoken in northern Sumatra, where Banda Aceh was once the capital of an independent Muslim kingdom and a major port of call in the eastern Indian Ocean (see map at CHAM).

The closest relatives of Achehnese – far closer than the other languages of the big island of Sumatra – are the AUSTRONESIAN LANGUAGES of Vietnam, notably Cham, which also was once the 'national' language of an independent kingdom. But *why* are Achehnese and Cham closely related? When and by what route did an original single language split into these two that are spoken so far apart? That is as yet a mystery. Clues will come from the borrowings from Austroasiatic languages that occur in both Achehnese and Cham: further work may show which Austroasiatic language is responsible, whether Khmer, Mon or one of the Aslian languages of Malaya.

Achehnese also has borrowings from Sanskrit, some of them shared with Cham, Khmer and Mon. There are Malay loanwords, too, for Malay was used officially in Aceh even during its independence and (as Indonesian) remains the national language today. Dutch, language of the former colonial power, has influenced modern Achehnese heavily.

Formerly written in Arabic script, Achehnese now has a standard orthography in the Latin alphabet, agreed in 1979.

Numerals in Achehnese and Cham		
Achehnese		Cham
sa	1	tha
duwa	2	dwa
lhɛə	3	klɔ̆w
pɪət	4	pa
limɔng	5	limu
nam	6	năm
tujoh	7	tadjuh
lapan	8	talipăn
si-kurɪəng	9	thalipăn
si-ploh	10	tha pluh

AFAR

PERHAPS 500,000 SPEAKERS

Eritrea, Djibouti, Ethiopia

Afar, one of the Lowland East CUSHITIC LAN-GUAGES, is spoken in the desert region of southern Eritrea, in Djibouti (once the 'French Territory of the *Afars* and Issas') and in neigh-bouring parts of Ethiopia.

> The speakers call themselves '*A' far* and their language '*A'far af* 'mouth of the Afar'. In Arabic they are *Danakil* (singular *Dankali*), while in Ethiopia they are officially called *Adal*, the name of the old sultanate that once ruled this region.

Henry Salt's *Voyage to Abyssinia* (1814) con-tained the first published word list of Afar, which has been studied by many linguists since his time. Some textbooks and religious texts have appeared in Afar; both Latin and Ethiopic scripts have been used.

Speakers are Muslims and the region has long been under strong Arabic influence. Arabic is a national language of Djibouti and of Eritrea, and Arabic is the everyday language of the two major towns in Afar territory, Djibouti and Massawa. Yet the basic vocabulary of Afar contains rela-tively few Arabic loanwords.

The first ten numerals in Afar are: *enek, nam-may, sidoh, ferey, konoy, lehey, malhin, bahar, sagal, taban.*

Afar, Saho and BEJA on the map

Afar has three main dialects, *Southern* (including Bādu and Aussa subdialects), *Central* in Djibouti, *Northern* in Eritrea.

Saho or Shaho, the northern extension of the Afar dialect continuum, is a language of Muslim pastoralists on the seaward-facing mountains of northern Eritrea, with about 120,000 speakers. The three dialects of Saho are *Hadu, Miniferi* and *Assaorta*. The *Irob* people are Christians who are bilingual in Saho and Tigrinya.

BEJA, only distantly related to Afar and Saho, is the Northern Cushitic language of the nomadic 'Bedouin' of north-eastern Sudan and northern Eritrea.

Northeast Africa: Afar, Saho and Beja
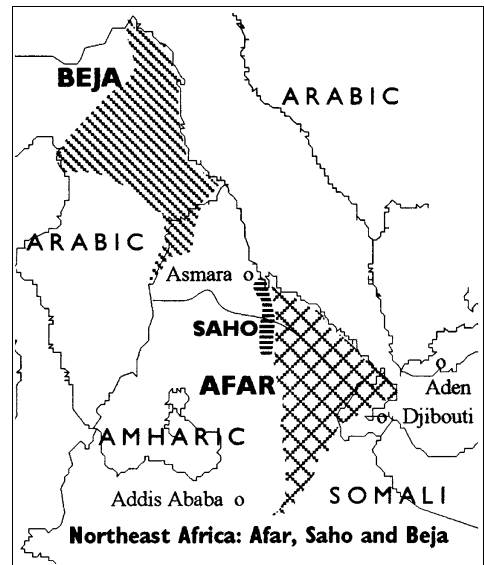
AFRIKAANS

6,000,000 SPEAKERS

South Africa

Afrikaans ('African') is a daughter language of Dutch: thus it belongs to the GERMANIC group of INDO-EUROPEAN LANGUAGES. It was one of the two national languages of white-ruled South Africa; it is still one of the eleven official languages of the country, spoken by the third largest linguistic community, after Zulu and Xhosa. There are Afrikaans-speaking minorities in Namibia and Zimbabwe.

> The white speakers of Afrikaans were called, in Afrikaans and South African English, *Boers*, literally 'farmers'. The word is the same in origin as German *Bauern* 'peasants'.

A local variety of Dutch had begun to develop not long after the foundation of Cape Town by the Dutch East India Company in 1652. Landmarks in the establishment of Afrikaans as a separate language are: the expansion of the Dutch-speaking settler population across South Africa in the 18th and 19th centuries; the development of a linguistically and racially mixed community, with intermarriage and the immigration of Malay and Indian language speakers; the annexation of the Cape by Britain in 1806, and the establishment of the self-governing Union of South Africa in 1910. Under the new democratic constitution of South Africa the formerly pre-eminent position of Afrikaans is likely to decay: English, less closely identified with racial exclusivity, is more generally acceptable in the role of lingua franca.

The local language was initially called *Cape Dutch* (or *Taal Dutch*, *Cape Coloured Dutch*, *Baby-Hollands*). It was slow to gain recognition as a real language, partly because it served at first mostly as an argot spoken within the sub-cultures of separate ethnic communities, and as a household jargon for communication with and among servants. The first printed book in Afrikaans was an Islamic religious text in Arabic script, in the *Cape Malay* or *Cape Afrikaans* variety of the language; it appeared in 1856.

Afrikaans has served as the basis for other, more ephemeral, mixed languages in South Africa. *Orange River Afrikaans* or *Kleurling-Afrikaans* was a creolised Afrikaans adopted by mixed populations called Griqua and Koranna, the majority of them former speakers of KHOISAN LANGUAGES, Nama and others. The Rehoboth Basters ('bastards'), a rural population descending from Dutch men and Khoi women, spoke a similar creolised variety.

Ironically, then, in view of its recent position as the ruling language of a racially exclusive community, Afrikaans has complex origins – which it demonstrates in a rich variety of loanwords from Portuguese, Malay, Bantu languages and Khoisan languages. It has also borrowed from English, and has in turn influenced the regional English of South Africa. The sound patterns of the two languages have naturally tended to converge.

> *Flytaal* or *Flaaitaal*, now better known as *Tsotsitaal*, is an Afrikaans-based jargon of black youths around Johannesburg and Pretoria: for its Zulu- and Sotho-based equivalent, *Iscamtho*, see SOTHO. It developed among criminal gangs west of the big city. In the 1940s and after, criminal gangs were the preferred role models for urban black South African youngsters: their language spread rapidly.

By contrast with Dutch, Afrikaans has no noun gender: *die man* 'the man', *die vrou* 'the woman'. A double negative, comparable to French *ne . . . pas*, is the usual rule: *hy staan **nie** op **nie*** 'he does not stand up'.

Afrikaans has contributed numerous loan-words to English, including the notorious *apartheid*, literally 'separateness'. *Kraal* 'enclosure' is in origin an Afrikaans loan from Portuguese *curral* 'farmyard', which is also the origin of American English *corral*. For a table of numerals see DUTCH.

AFROASIATIC LANGUAGES

Theodor Benfy demonstrated the relationship between Egyptian and the Semitic languages in 1844, and named the family *Semito-Hamitic* in 1869. These are key dates in the gradual recognition by scholars that the SEMITIC LANGUAGES so well known from Biblical and Orientalist study were part of a much wider family that also included BERBER LANGUAGES, CHADIC LANGUAGES, CUSHITIC LANGUAGES, OMOTIC LANGUAGES and Ancient EGYPTIAN.

Scholars who have looked at Afroasiatic linguistic relationships have no doubt of them. There once was a language something like the 'proto-Afroasiatic' that is now being reconstructed. But some of those who work on Semitic languages, on the Ancient Near East and on European prehistory have not yet thought through this century-old discovery. Even more strikingly than the Uralic and Indo-European families, the Afroasiatic language family cuts across usually perceived racial boundaries. It is an exciting challenge for archaeologists, anthropologists and linguists to trace its origin and the steps by which the individual groups diverged and spread.

Among present-day scholars of Afroasiatic languages, Christopher Ehret has been the most productive and successful: see his *Reconstructing Proto-Afroasiatic* (Berkeley: University of California Press, 1995). Ehret has postulated a series of divisions in which first the Omotic languages, then Cushitic, then Chadic, separated off as early dialects. This leaves a 'Boreafrasian' group, out of which emerged Berber, Egyptian and Semitic.

This is the language side of the story. In human terms, it goes with a hypothesis that proto-Afroasiatic was spoken, perhaps 18,000 years ago, in the Horn of Africa. Omotic, Cushitic and Chadic languages remain as the traces of a very early westward expansion: later, speakers of the earliest Boreafrasian dialects spread northwards across what is now the Sahara (but was a less arid environment then) and expanded both westwards and eastwards, eventually occupying the vast area that stretches from Morocco to Arabia.

As a result of this expansion, the Berber and ancient Egyptian cultures developed in North Africa itself, while across the Red Sea, the earliest Semitic dialects spread northwards from the Arabian peninsula to the Fertile Crescent, first emerging into history when speakers of the Semitic language AKKADIAN seized power in southern Iraq, where previously the quite unrelated SUMERIAN language had been pre-eminent.

It is because of the great length of time involved here that proto-Afroasiatic has been more difficult to work on than proto-Indo-European. There is another difficulty. All the Afroasiatic languages are built on word roots consisting of consonants, between which vowels are inserted to create various verb and noun forms. The Semitic languages have three-consonant roots. The others have mainly two-consonant roots, and it has not been clear how the Semitic forms could in practice have developed out of these. Ehret has presented persuasive evidence that single-consonant suffixes, with various fixed meanings, still found in some of the other Afroasiatic languages, became attached invariably to word roots in 'pre-proto-Semitic', thus resulting in the well-known three-consonant roots of modern Semitic languages.

Most Afroasiatic languages have two series of forms for nouns, 'independent' and 'construct': for example, Hebrew *dəbarīm* 'words' but *dibrē 'emet* 'words of truth'. Most languages of the

family have a masculine/feminine gender distinction, feminine forms being marked with a *-t*:

in Berber, *imishshu* 'male cat', *timishshut* 'female cat'; in Ethiopic, *bəʾəsi* 'man', *bəʾəsit* 'woman'.

The names of Afroasiatic

The name *Hamitic* or *Chamitic* was already used in very early scholarship to cover some of the groups now combined in 'Afroasiatic'. It comes from Ham, one of the Biblical sons of Noah. Benfy used *Semito-Hamitic* to make clear his view that the already recognised Semitic group was a co-ordinate half of the same family. Later Benfy's term was inverted, to make *Hamito-Semitic*, because scholars recognised that Semitic, though it belonged to the group, was just one of several divisions of it, while 'Hamitic' covered all the rest.

Afroasiatic is now a more popular term. It is less culture-specific: 'Hamito-Semitic' makes sense only to those who remember the Biblical story of Noah and his sons. But geographically, the term 'Afroasiatic' is a little too all-embracing, because there are many languages of Africa and Asia that do not belong to this family. *Erythraic* (after the Classical Greek name for the Red Sea), *Afrasian* and *Lisramic* (an invented compound of two Hamito-Semitic roots) have also been tried.

These terms are 'part of the dialect geography of linguistic terminology: roughly Semito-Hamitic for Eastern Europe; Hamito-Semitic for the rest of Europe; Erythraic in the focal area of the School of Oriental and African Studies, London; Afroasiatic in the U.S. generally; Lisramic and Afrasian still on the doorstep waiting to be adopted. Lisramic is the only one based on roots from the languages themselves (**lis* ''tongue, language''; Egyptian *rāməč* ''people'').'

After Carleton T. Hodge in *The non-Semitic languages of Ethiopia* ed. M. L. Bender
(East Lansing: African Studies Center, Michigan State University, 1976) p. 43

AFROASIATIC LANGUAGE GROUPS
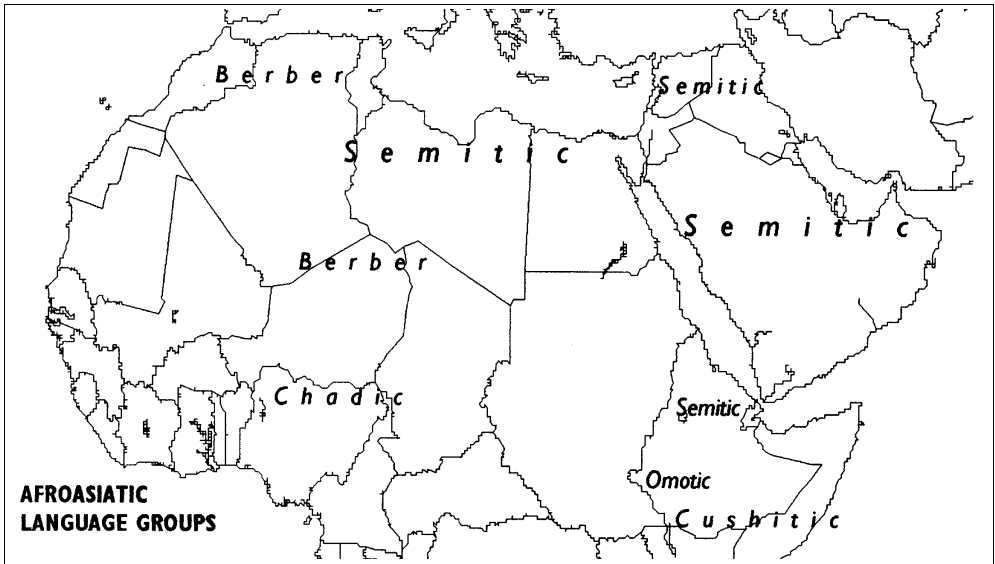
AHOM

EXTINCT LANGUAGE OF INDIA

Ahom is one of the TAI LANGUAGES (see map at SHAN). It was spoken by a conquering band who ruled the lower Brahmaputra valley, modern Assam, for six hundred years from their capital at Sibsagar.

The Ahoms' own name for themselves was *Tai*. In ASSAMESE they were called *Asam* (nowadays pronounced *Ohom*), a word related to *Shan* and to *Siamese*, both of them designations for Tai peoples. The state of *Assam* is named after them.

The Ahoms invaded Assam in the 13th century and established a kingdom. In the 18th century their power began to fade: the accession of King Lakshmi Singha in 1769 was followed by rebellions, Burmese invasions and civil wars. British rule, imposed after the treaty of Yandabu in 1826, allowed the Ahom kings to retain Upper Assam. But the local Indo-Aryan tongue, Assamese, had long since become their court language.

Ahom has had rather little linguistic influence on Assamese. But the Ahom kings encouraged a kind of literature almost unique in India. This was the prose chronicle, *buranji*, at first written in Ahom but destined in the 17th century to become one of the glories of Assamese literature.

The Ahom alphabet had 41 characters. Ahom was almost certainly a tonal language, like its Tai relatives, but the tones were not recorded in the script so they are now unknown.

The first ten numerals in Ahom are: *it, chang, chām, chi, hā, ruk, cit, pet, kāo, chip.*

The twelve months in Ahom script

After B. Barua, N. N. Deodhai Phukan, *Ahom lexicons* (Gauhati: Department of Historical and Antiquarian Studies in Assam, 1964)

AKAN

5,000,000 SPEAKERS

Ghana, Ivory Coast

Akan, often called Twi, is one of the Kwa group of NIGER-CONGO LANGUAGES. It was the ruling language of the Ashanti Empire and is now the most important African language of Ghana. Even among some who use other languages in daily life, Akan is the language of the priesthood and of liturgical texts in the Ashanti religion, still widely practised.

Since the 1950s *Akan* has been the preferred term for the language as a whole. *Twi* is an alternative, often used for the Asante dialect in particular. There are three standard forms of Akan, all mutually intelligible: Twi or *Asante* (of the region of Kumasi, centre of the *Ashanti* Empire), *Fante* and *Akuapem* (of southern Ghana).

Long before the rise of the Ashanti Empire, an Akan-speaking kingdom had its capital at Bono-Mansu, whose ruins are a hundred miles north of Kumasi. It was founded around 1300 and its gold mines made it a city of legendary wealth. In 1740, after defeat by Ashanti, it was destroyed. The Brong dialect, the most north-westerly of the Akan group, remains to show the former reach of the Bono kingdom.

The founder of the Ashanti Empire was Osei Tutu (ruled 1695–1731). By the early 19th century his successors controlled much of modern Ghana from their capital at Kumasi, and wherever their rule spread, so did their Asante language. Meanwhile Fante, the language of the coastal districts west of Accra, and Akuapem, of the hills north of Accra, had both achieved importance as inland trade from coastal ports increased – and both had been adopted as written languages for religious publications. A grammar of Fante had been published in Danish as early as 1764, a reminder that this was once the Danish Gold Coast; it was preceded by a glossary of Asante published in German in 1673.

When the British captured Kumasi in 1874, Akan in its three forms remained the single most useful African language on the Gold Coast and in its newly conquered hinterland, continuing to spread with roads, schools and missions. Religious and secular publications in Fante and Twi multiplied.

To replace the three standard orthographies for Asante, Fante and Akuapem – and so to make publishing in Akan a more rewarding activity – a new Akan orthography was promoted in the 1980s.

Although English, the official national language and the language of broadcasting, is still in many situations the most useful lingua franca, Akan has a growing role as the language of markets and transport across much of modern Ghana: in the north-east, where Hausa may be a second language for men, Akan is more likely to be the market language for women.

To the west of the Akan-speaking area, where Anyi and Sehwi (see BAULE) are the local languages, Akan is used in church, in trade and in most schools. To the south and east the situation is similar. In these districts, therefore, many people are bilingual in Akan. Akan has had a strong influence on GĀ (the local language of Accra) and EWE, both of which have borrowed Akan terms relating to food (including sea fishing) and statecraft.

Akan personal names have also been fashion-

able in these languages – names such as Kofi and Kwame, which relate to the days of the week (see table at BAULE). This is an important detail of Akan traditional belief, since each of the eight or nine paternal lineages, *ntoro*, in Akan society sets aside a specific day of the week for ritual washing: hence the common conversational gambit, 'Which *ntoro* do you wash?'

Akan has two tones – high and low. It also has a 'vowel harmony' rule based on the position of the root of the tongue: a word may contain either 'advanced' or 'unadvanced' vowels but not both. In the Fante dialect (as in the neighbouring Baule which, like Akan, belongs to the Tano or Volta-Comoe subgroup of Kwa languages), vowel harmony is more complex, depending also on the position of the lips: rounded and unrounded vowels may not be mixed in the same word.

The dialects of Akan

In the 1960 census, when Ghana had a total population of 6.7 million, first-language speakers of the eight main dialects of Akan were enumerated separately:

Asante or Twi	913,270
Fante	708,470
Brong or Abron	320,240
Akuapem or Akwapim	144,790
Akyem	203,820
Agona	49,080
Kwahu	131,970
Wasa	94,260

Brong (which some consider a separate language) has an additional 75,000 speakers in neighbouring districts of the Ivory Coast.

From *The languages of Ghana*
ed. M. A. Kropp Dakubu
(London: Kegan Paul, 1988)

West Africa: Akan
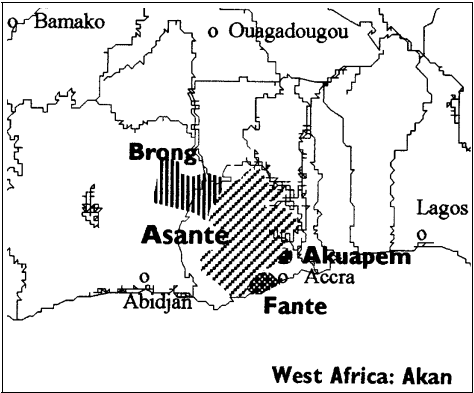
When not to trust

Kontromfí sè: Áfei ne ampã – 'Monkey says: Now I am going to tell the truth!'

Numerals in Akan

	Twi	Brong
1	biakō	ekō
2	abieng	enyõ
3	abiesã	esã
4	anang	enã
5	anum	enũ
6	asĩa	ensyĩ
7	asong	ensõ
8	awotwe	moqie
9	akrong	enkunõ
10	edu	edu

AKKADIAN

EXTINCT LANGUAGE OF IRAQ

One of the SEMITIC LANGUAGES, Akkadian was once a ruling language in Iraq and Syria. Some of the oldest literature in the world comes to us on clay tablets in Akkadian cuneiform script – including the *Epic of Gilgamesh*, and myths of the Creation and the Flood which have pervasive similarities with the HEBREW Biblical texts.

After it was supplanted by Aramaic and Greek (see below) Akkadian fell from use. For nearly two thousand years it was a forgotten language. Stone inscriptions and clay tablets in the strange wedge-shaped script were first encountered by modern travellers in the 17th century, and the painstaking work of decipherment began. It is because of these texts – literature, chronicles, business accounts, school exercises, even recipes – that knowledge of Akkadian exists today. Through deciphering Akkadian, the recorded history of the Middle East has been gradually extended thousands of years into the past.

Akkadian is named after Akkad, capital city of the empire ruled by Sargon (2350–2294 BC according to one chronology). The terms Assyrian and Babylonian are also found, usually as the names of the two dialects of Akkadian that became standard languages of the later empires of Assyria and Babylon.

The cuneiform writing system was not a new invention when it was first applied to Akkadian about 2350 BC. Until then it had been used for SUMERIAN, the ancient and unrelated language of the oldest kingdoms of southern Iraq. Speakers of Akkadian, advancing northwards from the Arabian peninsula, had occupied the region and their language had begun to replace Sumerian in everyday use around 2500.

In this process Akkadian borrowed heavily from Sumerian – but it did not borrow verbs. Sumerian verb forms were built up by adding affixes. Akkadian verbs, like those in other Semitic languages, are modified by means of internal changes, mainly of vowels. Sumerian verbs were not made to fit the pattern. However, when compared with other Semitic languages, Akkadian is seen to have a great number of Sumerian noun loanwords.

The Akkadian script also borrowed Sumerian word forms directly – using them as ready-made logograms for Akkadian words that had the same meaning but sounded quite different. In addition, whole long texts of Sumerian literature were recopied and preserved in Akkadian libraries, sometimes in bilingual versions meant to help in the understanding of the old classical language of Sumer.

By 2000, a distinction can be observed between the dialects of northern Iraq (Assyria and places linked with Assyria) and southern Iraq (the Sumerian cities, later Babylonia).

Eblaite is the language of the once-important city of Ebla, whose ruins were excavated at Tell Mardikh, forty miles south of Aleppo, in 1975. Far away from southern Iraq, this kingdom also had a writing system based on Sumerian cuneiform. Tens of thousands of tablets were recovered and are being deciphered. Eblaite was surprisingly similar to Akkadian, though not identical with it. It also seems to have many Sumerian loanwords.

Akkadian was at its most important in the second millennium BC. A great range of literature was written down in Old Babylonian, to be preserved in libraries or archives. Meanwhile, Akkadian was becoming the first truly international language of business and diplomacy in the Middle East. Akkadian texts of this period are to be found in Egypt and in Anatolia, far away from the regions where Akkadian was an everyday

language. Special forms of the language, such as the 'Amarna Akkadian', mixed with Canaanite, of the Tell el-Amarna tablets, were written by people who perhaps seldom spoke the language, using it simply as a recording medium. Akkadian was – for a while – at the centre of a multilingual culture, in which linguistic influences flowed in all directions. The spread of the cuneiform script mirrors these influences: after passing from Sumerian to Akkadian, it was afterwards used in different forms for Hurrian and Urartian (see CAUCASIAN LANGUAGES), Hittite, Ugaritic, Elamite (see DRAVIDIAN LANGUAGES), Old Persian and several other languages. All of these languages borrowed words from Akkadian.

New Assyrian, and late New Babylonian texts, show that by their time both dialects were strongly influenced by Aramaic, which must eventually have become the everyday speech of almost all of Iraq and was to be the administrative language of the Persian Empire. Assyrian, as far as can be seen from surviving tablets, ceased to be written altogether about 600, while Babylonian texts continue as late as the 1st century BC.

Akkadian numerals

Masculine		Feminine
ishtēn	1	ishtiat
shina	2	shitta
shalāsh	3	shalāshat
erba	4	erbet
ḥamish	5	ḥamshat
?	6	sheshshet
sebe	7	sebet
samāne	8	?
tishe	9	tishīt
esher	10	esheret

In Akkadian texts, numerals are usually written in figures rather than in words – so the pronunciation of some numerals is still unknown. As in many Semitic languages, a feminine numeral goes with a masculine noun and vice versa.

For large numbers, bases of 60 and 100 were used together: 60 *shūshum*, 100 *mētum*, 120 *shina shūshi*, 200 *shitta mētim*, 240 *erbet shūshi*, 300 *shalash me'āt*, 600 *nēr*, 1000 *lim* . . .

Periods of Akkadian

Northern Iraq	Dates BC	Southern Iraq
Old Akkadian	2500–2000	Old Akkadian
Old Assyrian	2000–1500	Old Babylonian
Middle Assyrian	1500–1000	Middle Babylonian
New Assyrian	1000–600	New Babylonian
–	600–1	Late New Babylonian

The Akkadian writing system

The characters in which Akkadian was written are called *cuneiform* 'made of wedge-shaped strokes'. They were usually formed with a pointed stylus on a wet clay surface which was then dried and sometimes baked. Huge libraries and archives of these clay tablets have been found in the excavation of big Assyrian and Babylonian cities. There are also monumental stone inscriptions: on these, engravers imitated the same style of writing.

There were over a thousand characters, some syllabic, some representing a whole word or concept. Many come originally from characters used for the quite unrelated Sumerian language. Scribes had to learn hundreds of characters, and might come to know the whole range, but businessmen were able to write letters and keep accounts with a knowledge of no more than 150 signs.

Some of the latest Akkadian inscriptions, erected by Persian emperors in the 5th century BC, were trilingual. These were the keys to decipherment – names, in particular, could be recognised in all three languages, and the signs so recognised made a beginning in the reading of other words. The Danish schoolteacher G. F. Grotefend took this first step in 1802. Decipherment was closer to completion by 1857, when the Royal Asiatic Society sponsored a competition in deciphering Akkadian. The result was that four similar – and essentially correct – translations of the same text were submitted.

ALBANIAN

5,000,000 SPEAKERS

Albania, Yugoslavia, Italy, Greece

Albanian forms a separate branch of the INDO-EUROPEAN LANGUAGES, quite distinct in its development from Greek and its other modern neighbours. It is the official language of Albania, also spoken by the majority in the Kosovo Metohija province of Serbia and by long-established minorities in Italy and Greece.

The local name of Albania is *Shqipëria* and of the language *Shqip*. An ancient term for a tribe from this region, classical Latin *Albani*, still survives in the name that the Albanian speakers of Italy give to themselves, *Arbëresh*; in the Greek name, *Arvanítis*; and in the Balkan names, such as Romanian *Arnăut*. As *Albania* it has become the English and international name for the country and its language.

Where, exactly, were the speakers of Albanian during the thousands of years that have passed since the proto-Indo-European dialects began to grow apart? Names and a few other words from languages of the ancient Balkans had been noted down by Latin and Greek authors. Some of these languages, such as Illyrian, are reminiscent of Albanian, but none has been identified as the exact precursor of the modern language.

Albanian itself was not recorded in writing until early modern times. Apart from single words, the first certain record of the language is a formula for baptism, written down in 1462 for occasions when no priest was available. One line of Albanian is given to the hero of a Latin play published in Venice in 1483, Thomas Medius's *Epirota* 'The Man from Albania'. In 1555 a prayer book in Albanian was printed. Composed by Gjon Buzuku, bishop of Shkodra, this is the oldest known book from Albania itself. The first Albanian publication from Sicily, also a religious text, appeared in 1592.

The greatest Albanian literature is poetry. There are oral epics, comparable to those in SERBIAN from neighbouring Bosnia. These and shorter folk poems were the inspiration for the poetry that was printed in Albania and Italy in the 17th, 18th and 19th centuries. A national renaissance (*Relindja*) in the late 19th century saw political independence, and the development of a unified literary language and of a new alphabet.

A Latin–Albanian dictionary, by Blanchus (Bardhi), appeared in 1635. J. H. Xylander, who published a grammar in 1835, showed that Albanian was an Indo-European language. He noticed regular patterns such as the correspondence of Albanian *gj-*, Latin *s-*, Ancient Greek *h-*. An example is the word for 'snake': Albanian *gjarpër*, Latin *serpens*, cf. Greek *hérpyllos* 'creeping thyme'.

The basic vocabulary of Albanian is Indo-European: *ne*, Latin *nos*, 'we'; *krimb*, Old Irish *cruim*, 'worm'; *ëndër*, Greek *óneiros*, 'dream'; *darkë*, Greek *dórpon*, 'supper'; *i parë*, Sanskrit *pūrvah*, 'first'. With Albania belonging first to the Roman Empire, then to Byzantium, and then for five hundred years to the Turkish (Ottoman) Empire, the Albanian language has borrowed many words from Latin, Greek and Turkish: *djallë*, Greek *diábolos*, 'devil'; *këndoj*, Latin *canto*, 'sing'. The majority of Albanian speakers are Muslims; numerous words and names are borrowed from Arabic, language of the Qur'ān.

A striking feature of Albanian is the use of periphrastic expressions to avoid a tabu word: thus for 'wolf' one says *mbyllizogojën* (from *mbylli Zot gojën* 'may God close his mouth!') and for 'fairy' *shtozovalle* (from *shtoju Zot vallet* 'may God increase their round-dances!').

The Balkans: language convergence? language substrate?

Albanian has striking similarities with other Balkan languages, BULGARIAN, SERBIAN, GREEK and ROMANIAN: Albanian *treg*, Serbian *trg*, Romanian *tîrg*, 'market'; Albanian *shtrungë*, Romanian *strungă*, Bulgarian *strǎga*, 'milking-place at entrance to a sheep-fold'; Albanian *i shtrëmbër*, Romanian *strîmb*, 'crooked', Greek *strabós* 'cross-eyed'; Albanian *mëz*, Romanian *mînz*, 'foal'. Often the same Latin word is used in these languages with the same special meaning: Latin *paludem* 'marsh', Rumanian *pǎdure* 'woodland', Albanian *pyll* 'woodland'. Some features are shared more widely, with TURKISH, HUNGARIAN or ROMANI.

Languages which share a culture, and are often used side by side, do grow together in the course of time. This is why English, since the Norman conquest, came to have many more similarities to French than its relative German has.

Like English and French, Balkan languages are only distantly related in the usual linguistic sense of the word. Yet the resemblances among them are so striking as to demand some special explanation. They are seen not only in individual words but in basic features such as the definite article – this has different forms in the different languages, but in all of them it forms a suffix to the noun: Romanian *zi* 'day', *ziua* 'the day'; Bulgarian *tsaritsa* 'queen', *tsaritsata* 'the queen'.

Many linguists consider that a linguistic 'substrate' is at work – an older language, once spoken all over the region, which, before it died, influenced all the surviving languages in the same ways. Some consider the main substrate language of the Balkans to be Latin, which in the last centuries of the Roman Empire, and until the Slavonic invasions, was certainly spoken widely all over the Balkans.

Others would argue that travel, migration, seasonal transhumance and bilingualism among Balkan peoples, lasting over a very long period (at least from the Roman Empire to the present day), brought all the languages closer to one another. Romanians (especially Aromunians), Albanians and Romani, in particular, were forever on the move.

Albanian alphabets

The only surviving copy of Buzuku's 1555 prayer book, with some missing pages, is in the Vatican Library. 'This is the first book in our language, and was very difficult to make.' Buzuku and some other pre-19th-century Geg writers used the Roman alphabet with extra characters for some unfamiliar sounds. Early texts in Tosk are sometimes in the Greek alphabet. But literature from both north and south was often written and printed in the same Arabic alphabet that was used for Turkish, the ruling language of pre-20th-century Albania.

The modern alphabet, adopted in 1908 at the Monastir (Bitolj) Congress after fierce debate, counts 36 letters. The Turkish government fiercely opposed its introduction. The dispute contributed to Albania's declaration of independence from Turkey on 28 December 1912.

The Albanian alphabet

a b c ç d dh e ë f g gj h i j k l ll m n nj o p q r rr s sh t th u v x xh y z zh

---

The map of Albanian

The two big dialect groups, Geg and Tosk, are quite different from each other, so different that some consider them separate languages. Geg has many more vowel sounds than Tosk: not only the seven that are recorded directly in the Albanian alphabet, but also five nasals, which can be written *â, ê, î, ô, ŷ*, and in Geg all twelve of these vowels may be either long or short. Geg (specifically the South Geg of Elbasan) was the literary standard of the early 20th century, but since 1945 Tosk has taken its place.

Geg is the language of northern Albania and the 1,750,000 Albanian speakers of Kosovo Metohija in Serbia. Tosk is spoken in southern

---

Counting down in Albanian

Dymbdhet muj na i ka vjeti,	**Twelve** months has the year,
Njimbdhet muj pela mazin,	**Eleven** months the mare a foal,
Dhet muej lopa viqin	**Ten** months the cow a calf,
Nand muej gruej djalin,	**Nine** months the woman a boy,
Tet cica na i ka dosa,	**Eight** teats has the sow,
Shtat sy na i ka naleti,	**Seven** eyes has the devil,
Gjasht muej bajn gjys vjeti,	**Six** months make half a year,
Pes gishta na i ka dora,	**Five** fingers has the hand,
Kater kamb' na i ka dhija,	**Four** legs has the goat,
Tri kamb' terezija,	**Three** legs has the scale,
Dy lula na i ka vasha,	**Two** flowers has the girl,
Vasha ka bylbylin-o,	The girl's taken by the nightingale,
Vasha po këndon malit-o,	The girl sings on the mountain
Sall per hater djalit-o!	Naked beside her lover!

This counting song is recorded in several versions from different villages. Some say 'The tenth month the army takes' and some say 'Two apples has the girl'.

After Max Lambertz, *Lehrgang des Albanischen* part 3 (Halle: Niemeyer, 1959) pp. 219–25; Eric P. Hamp, 'Albanian' in *Indo-European numerals* ed. Jadranka Gvozdanovic (Berlin: Mouton De Gruyter, 1992) pp. 835–927

Southern Europe: Albanian

Albania, in parts of central Greece, and in scattered districts in Italy. The Tosk dialects of Sicily and southern Italy are known as *Arbëresh*, those of Greece as *Arvanitika*. There are Tosk-speaking villages in Bulgaria and in Ukraine (near Melitopol), and a Geg-speaking community, Arbanasi, near Zadar on the Dalmatian coast.

Albanian communities in Greece and southern Italy have been there for many centuries, their numbers regularly refreshed by continuing migration.

Algonquian languages

LANGUAGE FAMILY OF NORTH AMERICA

From the 15th century onwards, explorers and settlers of the eastern seaboard of North America received their first impressions of New World natives from speakers of Algonquian languages. This group of AMERIND LAN-GUAGES has about fifteen living members, stretching from Labrador to Alberta and Wyoming; there were once many more, as far south as Carolina, but intensive European settlement in what is now the eastern United

Plains Cree

W o o d s C r e e

N a s k a p i

Swampy Cree

M o n

Blackfoot

Moose Cree

Atikamek

Cheyenne

O J I B W A

Ottawa o

[Delaware]

Washington o

States drove out and eradicated many Algonquian peoples.

Two trade languages have had an Algonquian basis. Mitchif or French Cree was spoken in North Dakota and in the Great Plains of Canada. The Atlantic Jargon of New Jersey was used between English and Dutch traders and New Jersey Indians. It was influenced by English grammar, but its vocabulary was largely Algonquian.

Algonquian languages have contributed a mass of place names to the map of North America. Massachusett, Narragansett, Connecticut, Cheyenne and Illinois are all names of extinct Algonquian peoples and languages. Delaware and Ottawa are names of languages that are still

spoken, though the few living speakers of Delaware live a thousand miles away from the state that is named after their ancestral tribe.

English has numerous Algonquian loanwords. *Squash*, with the meaning of 'pumpkin', comes from Narragansett, a now-extinct New England Algonquian language; so does *moose*. Others, such as *toboggan*, come from Abenaki, once an important language in Maine and Quebec, now with fewer than twenty speakers.

Algonquian languages on the map

The Algonquian or Algonkian group includes several languages with some thousands of remaining speakers: CREE, OJIBWA, Blackfoot (9,000 speakers), Cheyenne (2,000 speakers), Micmac (8,000 speakers) and Montagnais (7,000 speakers). Further Algonquian languages, still spoken but by very small numbers, are Abenaki, Algonquin, Arapaho, Atsina, Fox, Kickapoo, Maliseet and Passamaquoddy, Menominee, Munsee, Potawatomi, Shawnee and Unami (or Delaware).

Algonquian languages seem to be distantly related to the extinct *Wiyot* and the dying *Yurok* of north-western California, as was first suggested by Edward Sapir in 1913.

There are five main dialects of Cree. The Montagnais dialects in Quebec and Naskapi in Labrador (few remaining speakers) may, like Atikamek, be considered part of a dialect continuum with Cree.

Plains Cree is spoken in central Saskatchewan and Alberta. *Woods Cree* is spoken in northern Saskatchewan and north-western Manitoba. *Swampy Cree* is spoken in the remainder of Manitoba and most of Ontario. *Moose Cree* is spoken near Moose Factory, on the west coast of James Bay. *Atikamek* (3,000 speakers), sometimes regarded as a separate language, is spoken north of Trois-Rivières.

Ojibwa is now spoken in Indian reservations in Ontario, Manitoba, Saskatchewan, Michigan, Wisconsin, Minnesota, Montana and North Dakota. There are 50,000 speakers in Canada and 30,000 in the United States.

Ottawa and *Odawa* are alternative names for Eastern Ojibwa.

North America: Algonquian languages

t a g n a i s

Micmac
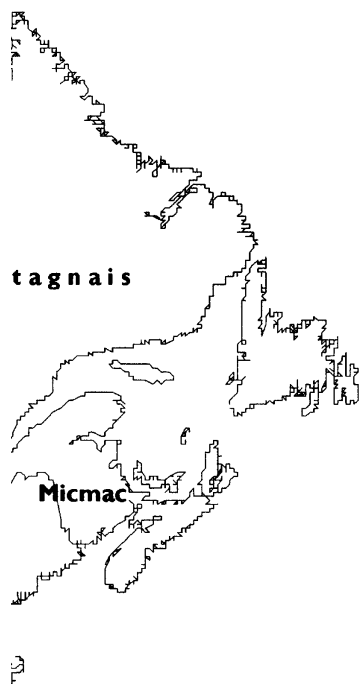
Altaic languages

A family of languages of northern and central Asia. Most research on them still centres on the individual groups that make up the family, TURKIC LANGUAGES, MONGOLIAN LANGUAGES and TUNGUSIC LANGUAGES.

The name *Altaic* is taken (paralleling *Uralic*) from the Altai mountain range of Asiatic Russia which seems to lie close to the point of origin from which Altaic languages must, many millennia ago, have begun to spread.

Language typology and the Ural-Altaic theory

Many linguists once believed that two well-known language families, Finno-Ugric and Turkic, were historically related. They were well known because both had representatives in Europe – and there was a temptation to link them because, although the individual word forms in these languages looked wholly different, the way that words were formed was almost identical. They were *agglutinative* languages, in which nouns and verbs were built of individual identifiable units: a verb might have, in addition to its basic 'root', a plural suffix and suffixes to identify the subject ('person'), the time past or present, the 'mood' factual or hypothetical, and perhaps others, all added in sequence, all separately audible (or at least visible when the word was written down).

This seemed wholly different from the other two families best known to European linguists, the Indo-European and Semitic languages. In typical *fusional* languages of both of these groups, words are indeed modified to reflect such features as plurality, person, time and mood, but the modifications appear to form an unanalysable whole – and sometimes appear inseparable from the root. A further contrast could be seen with *isolating* languages such as Chinese, in which all these ideas that need to be expressed take the form of separate words.

The postulated grouping was called Ural-Altaic – because both Finno-Ugric and Turkic languages had already recognised links with less-known groups. But scholars who worked on proving the wider relationship had little success. The actual words of the languages could scarcely ever be related to one another. And it became obvious that language *typology*, though relevant, was not in itself a firm guide to the *genetic* relationships between languages. Few now work on the Ural-Altaic link, which must lie very far back in prehistory.

A reaction followed. Linguists began to doubt even the lower-level relationships that had been assumed among the URALIC LANGUAGES (Finno-Ugric and Samoyedic) and among the Altaic languages.

Some now feel that the Turkic, Mongolian and Tungusic languages should be regarded as quite separate families. But this goes too far. Individual word histories do come together to show regular patterns of sound change, linking the three groups. Their grammars, too, are clearly related. The Altaic family is here to stay. Many linguists would include KOREAN in the family, but that link, also, must lie far back in prehistory. Some now add JAPANESE.

AMERIND LANGUAGES

The American continent was discovered, and began to be inhabited, perhaps twenty thousand years ago. It seems likely that early inhabitants spread gradually across the continent from the north-west, the first arrivals having reached it by crossing the Bering Strait.

Whether the oldest stratum of languages of the Americas consists of a single all-inclusive language family – the Amerind languages – is a question charged with controversy. If a single family, it is the most diverse of all those dealt with in this book, with about 300 languages in Canada and the United States; about 70 in Central America; about 600 in South America. Hundreds more, probably, have become extinct in the last two centuries. The total number of speakers of these languages is relatively small: 300,000 in the north; six million in Mexico and the other countries of Central America; eleven million south of the Isthmus of Panama.

The 15th-century European explorers thought at first that they had reached the East Indies, so they called the inhabitants of this new continent *Indians. Red Indian*, *American Indian*, *Amerindian* are attempts at clarification of the earlier misleading name. *America* is named after the discoverer of mainland Brazil, Amerigo Vespucci. His fellow-Italian Christophoro Colombo (in Spanish *Cristóbal Colón*) is commemorated in the name of *Colombia*, and in the term *pre-Columbian* for American peoples, cultures and languages as they existed before the European discoveries.

The study of American Indian languages began with Spanish missionaries in Central and South America in the 16th century. Some relationships have been known for a long time. Andrés Pérez Ribas, in 1645, suggested that the languages of Sinaloa were linked to Nahuatl. The group is now called Uto-Aztecan, and the relationship was worked on by the great American linguist Edward Sapir in the early 20th century.

In 20th-century work on the languages of 'Native America' over a hundred language families were identified, along with numerous 'isolates' – single languages apparently lacking any relatives. Some of the isolates were necessarily listed as such because the languages became extinct before they had been properly recorded. The Summer Institute of Linguistics (the Wycliffe Bible Translators) has recently been at the forefront of research, with grammars, dictionaries, Bible translations, and comparative studies.

In the abstract, most of these families and isolates might plausibly turn out to be related. Siberia and Alaska are fairly inhospitable places in themselves, and the number of separate times that people would independently determine on migration in this direction is unlikely to be large. So in arguing for just three migrations and presenting linguistic evidence for just three language families – ESKIMO-ALEUT LANGUAGES, now found on both sides of the Bering Strait; NA-DENÉ LANGUAGES, a family comprising a few languages of the north-west and Navaho; and Amerind languages, all the remainder – the linguist Joseph Greenberg was not doing anything surprising. But high-level classification in this area had always been controversial, and Greenberg's, the most ambitious of all, is no exception: see his *Language in the Americas* (Stanford: Stanford University Press, 1987).

He has not really demonstrated his thesis for all the small families and all the language isolates. For many languages now extinct, never fully recorded, it will probably be impossible to

demonstrate. But Greenberg has changed the agenda. As evidence assembles that the major language families are distantly linked, it becomes necessary in the case of the remainder to explain why they should not be linked with the rest.

Tukano multilingualism

Among the Tukano of the north-west Amazon it is incest if a man has a wife who speaks his language. Tukano villages have a language of men, languages of women, and a shared regional trade language.

After B. B. Kachru, 'Bilingualism' in *International encyclopedia of linguistics* (New York: Oxford University Press, 1992)

Many Amerind languages must always have had small numbers of speakers. Communication has, for centuries and perhaps millennia, demanded the use of second languages, lingua francas of various kinds, and pidgins. With Inca rule Quechua spread rapidly along the Andes in the last few centuries before the Spanish conquest. Tupí, now dying, was once the *lingua geral* of Portuguese-ruled Brazil. There was much bilingualism, notably in Nahuatl and Yucatec, in pre-Columbian Mexico. Amerind-European contact languages of North America have included Chinook Jargon of the north-west, Mobilian of the Mississippi valley, and Mitchif or French Cree. Plains Indian Sign Language, now kept alive by Boy Scouts, was once essential among Amerindian traders and travellers in the Midwest. Spanish, French and English came to fill a role as long distance languages among 'Native Americans' as well as among newly dominant peoples.

The north-western United States, a region of mountains and valleys, was the home of many small language communities. Soon after 1800 English sailors at the mouth of the Columbia River found that local people were speaking to travellers in a mixed trade language or 'jargon'. In the course of the 19th century it spread across much of Oregon, Washington State and British Columbia; in the 20th century it has died away as English has spread. This *Chinook Jargon* was clearly of Amerindian origin – its sound patterns were those of local languages including Chinook and Nootka – but it included French, English and Chehalis words as well as many of unknown origin. Whether it originated for local trade, or to deal with Europeans, is not known.

Mexico is a linguistic maze, clearly of long standing. There has been plant cultivation here from 5000 BC, sizeable villages from 2000 BC, ceremonial centres from about 1000 BC – always with a variety of cultures. NAHUATL was the language of government in the Valley of Mexico, and so it became the lingua franca in early Spanish times. But many neighbouring languages, in the Valley and bordering on it, are to all appearance not related to Nahuatl (Tarascan, QUICHÉ, Tzutuhil, MIXTEC) – and each major town tended to have its own well-marked dialect. They developed in parallel, without much travel and contact. So dialects diverged into languages, and there was rather little borrowing among them except for the names of obvious trade goods. There was no widespread literacy – even where it did exist, scripts were not fully linked to the sounds of words – and so little need for the emergence and maintenance of standard literary languages.

Amerind languages: what the dispute means

The squabble among scholars of American Indian languages – between the lumpers and the splitters – is not one of first principles. All natural human languages are related. It is about the use of scientific evidence. It matters because the evidence used and highlighted by linguists becomes an ingredient in the work of anthropologists and archaeologists – and thus, eventually, in people's understanding of their own cultural origins.

Consider two hypotheses. One is that, twenty thousand years ago or so, the early people who

crossed to Alaska already spoke two or more very different languages – not unlikely, considering the linguistic diversity of Siberia now and the symbioses that occur among linguistically diverse cultures there. If that were so, the time depth separating some 'Amerind languages' from others might be far more than twenty thousand years, and they might be less closely related to one another than they are to some languages of the Old World.

The second hypothesis is that all the people who crossed the Bering Strait in that early migration spoke the same language, and that the 'Amerind languages' have diversified from that one. We know nothing of the migration, but in the abstract this seems an equally likely hypothesis. If it were so, the time depth separating some Amerind languages from others would be about twenty thousand years: rather greater, probably, but not vastly greater, than that separating the branches of the Afroasiatic family. They would be more closely related to one another than they are to any of the languages of the Old World.

To decide between these hypotheses, linguists would have to be able to trace language relationships, by way of phonetic and lexical reconstruction, at least twenty thousand years back. Some (such as Eric P. Hamp) doubt that they will ever be able to do this, and argue that hypothetical family groupings implying such a timescale are useless because unprovable. Others (such as Joseph Greenberg) prefer to set out a hypothesis on the basis of initially weak evidence, and then get down to attempting a proof.

Greenberg has all the history of science on his side. This is how progress is made. But it has pitfalls, illustrated by the big world classifications of languages that are published from time to time – and illustrated by the headings AMERIND LANGUAGES, AUSTRO-TAI LANGUAGES and ALTAIC LANGUAGES in this book. These family groupings are still in the waiting room between inspired guesswork and established fact. Because classifiers like to classify, some languages firmly pinned down in reference books may turn out to belong somewhere quite different when the real work gets under way.

Major language families of the Americas

This book cannot deal fully with the immense variety of Amerind languages, most of which are spoken now by small and shrinking communities. The map gives a location to some families and linguistic isolates of historical importance. Cross-references (the small capitals in this list) point to larger-scale maps.

ALGONQUIAN, including Cree and Ojibwa

ARAUCANIAN

Arawakan, a family of vast range, extending from Belize to Paraguay, and including Black Carib or Garifuna (perhaps 100,000 speakers in Central American countries) and Guajiro (125,000 in Colombia and Venezuela)

Chibchan, including Guaymí and Cuna

Ge-Pano-Cariban, including Carib

Hokan

Iroquoian, including CHEROKEE

Jivaroan, including Shuar and Aguaruna

MAYAN LANGUAGES, including Cakchiquel, Quiché and Yucatec

Misumalpan, including Miskito (perhaps 100,000 speakers or more)

Mixe-Zoque languages, including Mixe, Zoque, and Popoluca

Muskogean, including Mobilian

Otomanguean, a large family including MIXTEC and ZAPOTEC

Paezan, spoken from Panama to Ecuador and including Páez

Penutian, including Chinook Jargon

QUECHUA and AYMARA, sometimes said to be related but regarded as isolates by others

Tarascan, a linguistic isolate of Michoacán, Mexico

Totonacan languages, including Totonac (250,000 speakers)

Tukanoan, including Tucano, an important lingua franca of the upper Amazon

Tupian, including Chiriguano, GUARANÍ and TUPÍ

Uto-Aztecan, including NAHUATL

Yanomami

Zuñi, a linguistic isolate of New Mexico

Penutian

Algonquian

Iroquoian

Uto-Aztecan

Zuñi

Muskogean

Hokan

Totonacan

Uto-Aztecan

Tarascan

Otomanguean

Mixe-Zoquean *Mayan*

Misumalpan

Chibchan

Cariban

Paezan

Arawakan

Tukanoan

Yanomami

Tupian

Jivaroan

Panoan

Ge

Quechua

Aymara

Tupian

Araucanian

SOME AMERIND LANGUAGE GROUPS

AMHARIC

14,000,000 SPEAKERS

Ethiopia

About two thousand years ago a gathering, *habashat*, of speakers of an early Arabic or South Arabian dialect crossed the Red Sea and founded a kingdom in modern Eritrea and Ethiopia. Their language was ETHIOPIC. Splintered by the very difficult communications in this mountainous country, the earliest Ethiopic has divided into several daughter languages, of which Amharic is the most important – it is the national language of modern Ethiopia.

> The country used to be called by outsiders *Abyssinia* (from *habashat*): Christian Tigrinya speakers are still called *habesh* in Eritrea (Muslim Tigrinya speakers reject the name 'habesh' and may be called *Jabarta*). *Amharic* is called by its own speakers *Amərinnya*, a name derived from the district of *Amhara*, apparently the historic centre of the language. *Tigrinya* (sometimes written *Tigriña*, or with the Italian spelling *Tigrigna*) is similarly in origin the language of *Tigre* province.

Thus Amharic is one of the SEMITIC LANGUAGES like Arabic and Hebrew. Originally a minor dialect of a region south of Axum, Amharic could claim to be the 'Language of the King' since the accession of the Solomonid dynasty in the 13th century, a crucial lifting of its status.

By the 17th century Amharic was a language of everyday communication, and particularly of the army, throughout the Christian empire of Ethiopia. Its spread was accelerated by the considerable use of and trade in slaves in traditional Ethiopia: the slaves were drawn from peoples of southern Ethiopia of various mother tongues, so slaves and owners necessarily used Amharic as a lingua franca. Thus it was the language in which the Jesuits made their short-lived attempt to convert the Ethiopians to Roman Catholicism. This was a calculated break with tradition, for classical Ethiopic was (and to some extent still is) the language of the established Ethiopian Christian church.

Amharic literature goes back to the royal praise poetry of the 14th century. Ethiopic, however, remained the language of literature and education until the 19th century. Even nowadays, educated speakers of Amharic use a form of speech which is much influenced by the classical language.

As the language of the modern central government, Amharic continues to gain ground as the everyday language of the capital, Addis Ababa, though the native language in the country round about is OROMO. Amharic also serves as a lingua franca in Ethiopia generally. The total of speakers who use the language regularly may be as high as 30,000,000.

Amharic still has a mainly Semitic vocabulary. As with Arabic, Amharic word structure is largely based on consonant-only roots, with inserted vowels marking number, tense and other grammatical features. Naturally there are loanwords from Cushitic languages, such as *wəshsha* 'dog', *səga* 'meat', and from modern European languages, *bolis* 'police' from French, *tayp* 'typewriter' from English, *fabriqa* 'factory' from Italian. Ethiopia was under Italian rule from 1935 to 1941.

Amharic is written in Ethiopic script, which, like Indian scripts, combines a consonant with a following vowel in a single complex symbol. The first box contains one-seventh of the full Amharic alphabet table, showing the consonant

letters combined with the vowel *ä*. For a table of numerals see TIGRINYA.

Based partly on works by M. L. Bender, including *Language in Ethiopia* ed. M. L. Bender and others (London: Oxford University Press, 1976)

The Amharic alphabet: the 'first order'

ሀ ለ ሐ መ ሠ ረ ሰ ሸ ቀ በ ተ ቸ ኀ ነ ኘ አ ከ ኸ ወ ዐ ዘ ዠ የ ደ ጀ ገ ጠ ጨ ጰ ጸ ፀ ፈ ፐ

h l h m s r s sh k' b t ch h n ny ä k h w ä z zh y d j g t' ch' p' c c f p

---

The Amharic alphabet: 'h' with the seven vowels

ሀ ሁ ሂ ሃ ሄ ህ ሆ

hä hu hi ha he hə ho

Northeast Africa: Amharic, Tigrinya and Tigre
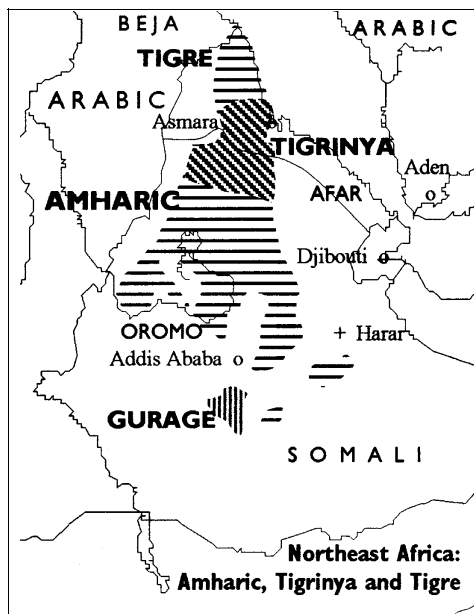
Amharic, TIGRINYA and Tigre on the map

The modern Semitic languages of Ethiopia are AMHARIC, Harari (13,000 speakers in the city of Harar), East Gurage (200,000 speakers), West Gurage (500,000 speakers), Soddo or Aymellel or Northern Gurage (100,000 speakers), Tigre and TIGRINYA.

There is relatively little dialect division in Amharic, though dialects of Shoa and Gojjam provinces are distinguished. The dialects of the old capital, Gondar, and the new capital, Addis Ababa, are both prestigious.

The heartland of *Tigrinya* speech is the highlands of southern Eritrea, and Tigre province across the Ethiopian border. Tigrinya is also dominant, serving as the national lingua franca, in the cities and towns of the Eritrean coast.

There are two dialects of *Tigre*, which may have about 100,000 speakers. The *Southern* or Highland dialect is spoken by the Mensa, Ad Timaryam and Ad Tekla tribes and most of the Red Marya. The *Northern* or Lowland dialect is spoken by the Habaab and Black Marya. This is the dialect in which the *Beni Amir* of Eritrea and Sudan are now bilingual (see BEJA).

ARABIC

PERHAPS 165,000,000 SPEAKERS

Algeria, Bahrain, Egypt, Eritrea, Iraq, Jordan, Kuwait, Lebanon, Libya, Mauritania, Morocco, Oman, Qatar, Saudi Arabia, Somalia, Sudan, Syria, Tunisia, Yemen

The importance of Arabic in the modern world goes back to its position as the language of the Qur'ān, the language of a conquering religion. As national language of nearly twenty countries, Arabic is now by far the most important of the SEMITIC LANGUAGES, which are a group within the AFROASIATIC family.

Arabic may possibly be the descendant of the language of the 'proto-Sinaitic' inscriptions of Mount Sinai, dated to about 1500 BC, but they are not fully deciphered. The first certain record of Arabic is in inscriptions from various parts of northern and central Arabia in the last few centuries BC. This is 'pre-classical' Arabic. Already several regional dialects existed.

Classical Arabic originates in the 7th century AD, when the masterpieces of pre-Qur'ānic poetry were composed (though written down only later) and when the Qur'ān itself was compiled. It immediately became the sacred text of the new religion of Islam – and, as Islam was spread by conquest, the Arabic language (in which the Qur'ān must be recited) spread with it.

The outreach of Arabic

So Arabic was from the beginning the literary and religious language of the rapidly spreading Islamic states of medieval North Africa and Asia. ARAMAIC of the Near East, COPTIC of Egypt, and BERBER LANGUAGES of North Africa retreated as Arabic advanced. Westwards, by the end of the 7th century, the Muslims held the whole north African coast. From here they advanced northwards across the 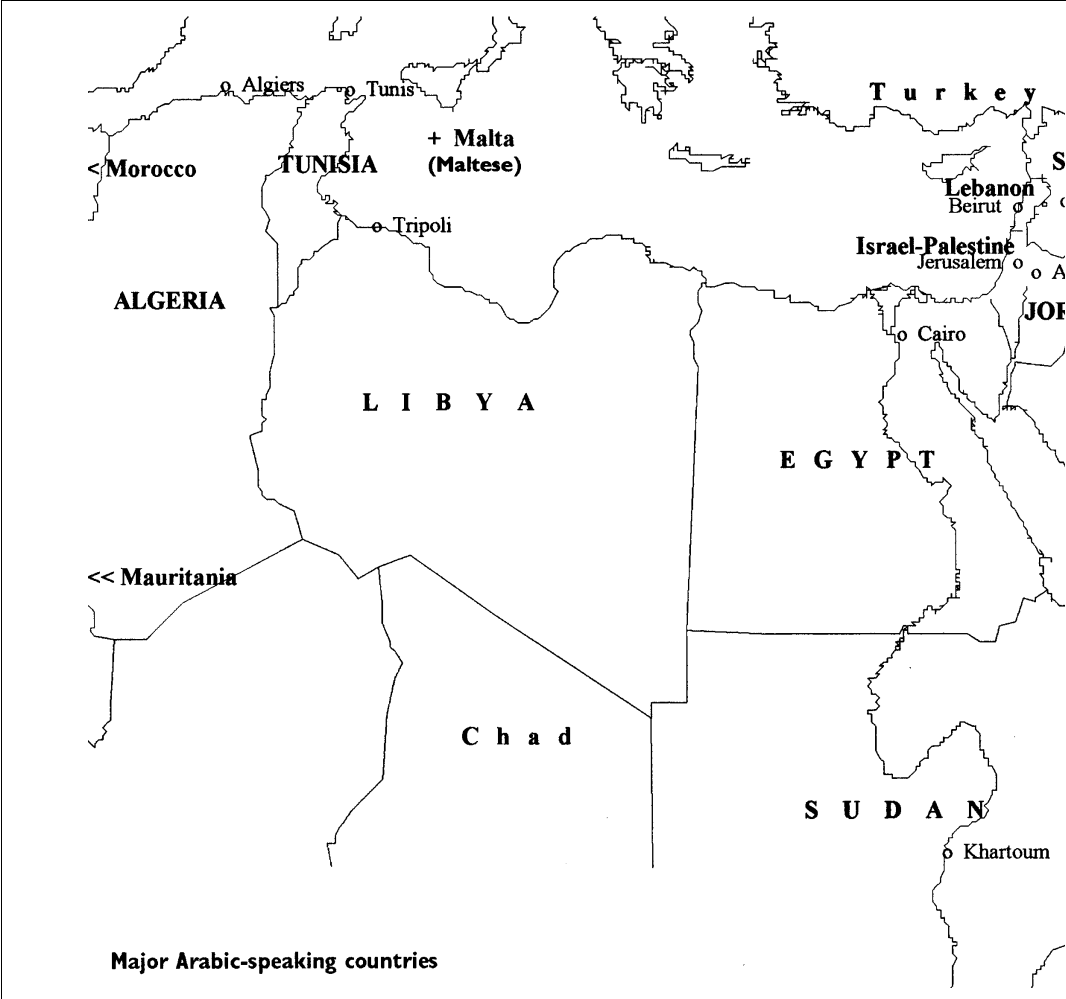Straits of Gibraltar (in 711) and on into Spain and southern France; southwards to the edge of the Sahara, and eventually across the desert to the plains of the western Sudan.

At this period Arabic must have spread very rapidly among people with different mother tongues, and there are signs – in the shape of Arabic loanwords in Spanish, for example – that pidgin and creole forms of the language existed. At least one Arabic text of the 11th century attempts to reproduce the pidgin Arabic of medieval Mauritania, with the remark: 'the blacks have mutilated our beautiful language and spoilt its eloquence with their twisted tongues' (quoted by S. G. Thomason and A. Elgibali).

In large parts of the conquered regions, Arabic eventually advanced from being the language of government, education and religion to become the language of everyday life. In most of these countries Arabic is now the national and majority language – but the Islamic countries were never entirely unified politically, and it is not surprising that over this great geographical range the language of the conquerors has split into dialects that are so different as to be mutually unintelligible.

In the Arabian Peninsula itself the dialects of Arabic have a history that goes back beyond Islamic times. They remain vigorous – even beyond the confines of the peninsula, for in northern Somalia the Yemeni and Hadhramaut dialects of Arabic are used as second languages, especially by traders.

Elsewhere, probably, a common language grew up among widely recruited armies, among

Algiers o

o Tunis

T u r k e y

**+ Malta
(Maltese)**

< Morocco

TUNISIA

Lebanon
Beirut o

S

< Morocco

o Tripoli

Israel-Palestine
Jerusalem o   o A

ALGERIA

JOR

o Cairo

L I B Y A

E G Y P T

<< Mauritania

C h a d

S U D A N

o Khartoum

Major Arabic-speaking countries

Loanwords from Arabic

Italian *dogana*, French and international *douane* 'customs house'	*diwān* 'office'
Italian *fondaco* 'warehouse', Spanish *alóndiga* 'cornmarket'	*funduq* 'inn'
French *alchimie*, English *alchemy*; French *chimie*, English *chemistry*	*kīmīā* 'alchemy' (originally from Greek *khēmeíā*)
French *chiffre* 'figure', English *cypher*; Spanish *cero* 'zero', English *zero*	*sifr* 'zero'
French *sucre*, English *sugar*, Spanish *azucar*	*sukkar* 'sugar' (originally from Sanskrit *śarkarā*)
Shona *ndarama* 'money', Tswana *talama* 'button'	*darāhim* 'money' (plural of *dirham* 'coin', borrowed from ancient Greek *drakhma*)

of Malta (MALTESE) has noticeable Italian features. In Egypt, COPTIC survived alongside Arabic down to modern times.

Arabic and the transmission of cultural loanwords

'In many ways the Arabs did not create the culture they brought to Africa. They . . . handed on what they had previously acquired from other peoples, mainly the peoples of antiquity. That is why we find today in African languages words of such diverse origin as the words for pen, money, army and shirt from LATIN; for philosophy, paper, diamond and list from GREEK; for lead, temple, poor man and sulphur from Babylonian [AKKADIAN]; for offering, angel, praise and prayer from Syriac [ARAMAIC]; for soap, sugar, banana and musk from SANSKRIT. Most of the words they brought are genuinely Arabic, however, such as *zabīb* ''grape'', *sufūr* ''copper'', *katan* ''cotton'', *dawāt* ''ink'', *ibrīq* ''kettle'' and *dhahab* ''gold''.'

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

---

governments and the governed, and it is from this medieval Arabic common language or *koiné* that the modern dialects (map 1) descend.

Their differences come partly from the different later cultural history of their homelands. Thus in Morocco, Algeria and Tunisia Latin and Berber were native, and Berber languages are still spoken; French became a ruling language in the 19th century and is still widely known. The Arabic of Spain (long extinct) was also influenced by Berber because the armies that conquered Spain were recruited in north Africa. The Arabic of modern north Africa contains many local terms, such as plant names, which are of Berber origin. The Arabic dialect

The further reaches

Eastwards, too, Islam spread rapidly – to Iran and central Asia, to large areas of northern India, to the Malay Archipelago, and many coastal kingdoms and provinces from Mozambique all the way to central Vietnam. Although Arabic did not become the everyday language of the majority in these areas, it became and remains a crucial influence. Wherever Islam is the religion, Arabic must be the language of religious education and traditional culture (map 2). Islamic fundamentalists, for example in Algeria and Turkey, tend to promote the sacred language and to depreciate local languages. And 'just as Latin and Greek supply the European language community with scientific terms and an educated vocabulary, Arabic performs that function within the Islamic

world; thus lexical items like *jāmi* ' "mosque", *madrasa* "school", *qā'ida* "rule" exist in nearly all languages of Muslim peoples' (W. Fischer).

In parts of the Islamic world, religion remained the main field in which Arabic was used. But this can amount to a big slice of daily life. Arabic is the common language of those who undertake the pilgrimage to Mecca, and is used for communication in all the towns and cities on the usual pilgrimage routes. Arabic came to be essential in some everyday fields, such as trade. On the edges of the Arabic-speaking world, strongly marked dialects, shading into pidgins and creoles, have developed. Alongside the dialectal Arabic of Mauritania, Chad, Sudan and lowland Eritrea, there are Arabic pidgins and creoles in Nigeria, Chad, Sudan and upland Eritrea.

Well beyond the Arabic language zone some isolated Arabic-speaking communities were established: some survive. There are still a few Arabic speakers in Afghanistan, claiming to be descendants of the Arabian nomads who once formed the armies of Islam.

Although so closely associated with Islam, Arabic is also the language of some Christian communities, notably the Maronites of Lebanon and Cyprus. Arabic in Syriac script, called *Karshuni*, is used in some religious texts of the Syrian Orthodox Church (whose everyday language is ARAMAIC). The first Arabic printed book was in Christian-Arabic Karshuni, though medieval block prints from Egypt take the history of printed Arabic further back – well before the introduction of printing into Europe.

Distinctive forms of Arabic, sometimes called *Yahudi*, are spoken by Jews of Morocco, Tunisia, south-eastern Turkey and other countries.

Although persecuted for centuries, the Yezidi religion is maintained in some Kurdish villages in Iraq, Iran, Syria and south-western Turkey by up to 150,000 believers. Originating from an early heretical Islamic sect, and with some Jewish, Christian and pagan elements, its religious rites are carried out in Arabic. Each believer has a special bond with a brother or sister in the next world.

Classical and modern

While regional dialects have continued to differentiate across the Arabic-speaking world, classical Arabic has – for 1,300 years – remained alive and vigorous. It is the official language of Arabic-speaking governments, the written language of literature, and the spoken language of inter-regional communication and trade. Without the Arabic tradition of education, centred on the Islamic religion and the Qur'ān, this position could not possibly have been maintained.

In the last two centuries, with the spread of modern education and communications, the position of classical Arabic has strengthened even further. It is the natural medium, in all Arabic-speaking countries, for the press, broadcasting and films.

Naturally 'modern standard Arabic' has its own regional accents and shows some dialect variation, reflected more or less noticeably in individuals' speech. But it is now so different from the various colloquial spoken forms of Arabic that these are often studied quite separately from the classical written language.

In the centuries after Muhammad, Arabic had become the vehicle of one of the great literary cultures of the world. Beside original writings, translations from Greek and Syriac contributed to the flourishing of Arabic science, medicine and philosophy. Many of these texts were afterwards translated into Latin for European scholars, helping to create an early renaissance – preceding the real Renaissance – in Western scholarship. After some centuries of stagnation, Arabic literature is flourishing once more in the 20th century.

Arabic in the world

The dialects of Arabic group as follows: Central Asian, Iraqi, Arabian, Syro-Lebano-Palestinian, Egyptian, North African (Maghrebi), MALTESE, West African (Hassaniya), Chadic (Shua), Sudanese. The Arabian dialects show the greatest differentiation.

*Turku* is a name (no longer used locally) for an

Arabic pidgin spoken east of Lake Chad and in the Bodélé depression.

Sudanese Creole Arabic (or Juba Arabic or Mongallese or Bimbashi Arabic) is an important lingua franca of the southern Sudan, a first language for some people, also used by many whose first languages are Nilo-Saharan.

Nubi is an Arabic creole that grew out of this, now spoken in Kenya (notably in the Kibera suburb of Nairobi) and Uganda (notably at Bombo near Kampala) by the descendants of slave troops who were recruited from various Sudanese peoples into the Egyptian army during the Egyptian conquest of the Sudan in the 19th century. Most of them transferred to the British Army in East Africa in the 1890s.

Well beyond the countries in which it has become the language of everyday life, Islamic conquests and religious proselytism have spread the knowledge of the language as a means of communication. Here Arabic remains necessary as the language of the Qur'ān and of education, but other languages are used outside these contexts: TEMNE, SUSU, SONINKE, MANDEKAN, BERBER LANGUAGES, SONGHAY, YORUBA, HAUSA, FULANI, KANURI, MABA, TIGRINYA, OROMO, SOMALI, SWAHILI, AZERI, CHECHEN, PERSIAN, TURKISH, URDU, BENGALI, MALAY, CHAM. All these have been strongly influenced by Arabic in their role as languages of Islamic states and peoples.

A second map looks at the South Arabian

The South Arabian dialects
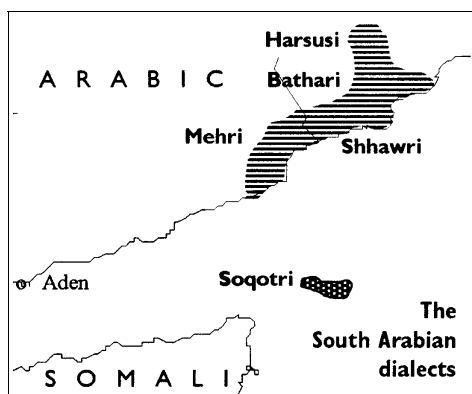
languages. Epigraphic South Arabian is a language known from inscriptions of the early city states of south-west Arabia, possibly beginning as early as the 8th century BC and continuing to the 6th century AD. There are five dialects, Sabaean, Minaean, Qatabanian, Ḥaḍramī and Awsanian. The scripts used in early South Arabian inscriptions are closely related to early Arabic scripts and to the writing system used for Ethiopic. Some modern dialects of southern Arabia, not closely related to Arabic, seem likely to represent a modern form of the same language (though the ancient inscriptions have not been deciphered fully enough to be quite certain). The best known of the modern South Arabian languages are Mehrī, Shḥawrī and Soqotrī (of the island of Socotra).

Arabic numerals

	Standard Arabic: masculine	Standard Arabic: feminine	Hassaniya dialect	Nubi Arabic creole
1	aḥad	iḥdā	wahad	wai
2	ithnāni	ithnatāni	azenein	tinin
3	thalātha	thalāth	tlata	talata
4	arba'a	arba'	arba	arba
5	khamsa	khams	khamsa	khamsa
6	sitta	sitt	sitta	sita
7	sab'a	sab'	saba	saba
8	thamāniya	thamānin	esmania	tamanya
9	tis'a	tis'	tissa	tisa
10	'ashara	'ashr	ashra	ashara

Arabic script

Arabic script developed in the 6th century AD. It originates from an early Aramaic alphabet, which in turn was derived from Phoenician.

The Arabic alphabet

ن م ل ى ق ف غ ع ظ ط ض ص ش س ز ر ذ د خ ح ج ث ت ب ا
ي و ه
a b t ṯ j ḥ kh d ḏ r z s š ṣ ḍ ṭ ẓ ' gh f q k l m n h w y

In the box the full form of each individual letter is shown, arranged from left to right to match the transliterations in the second row. But Arabic is written and read from right to left, and much-simplified forms of some letters are used in normal writing and printing when they are joined into words.

Thus *al-Qur'ān* 'the Qur'ān' is written الفرآن.

Arabic is the fastest and most cursive of all scripts: but short vowels are in general not written and the reader must supply them from a knowledge of the language and its structure. Vowels can, if necessary, be fully marked in Arabic writing by the use of signs above and below the line, but these are generally used only in texts of the Qur'ān itself, where correct pronunciation is essential for religious reasons.

Arabic script has been carried to many parts of the world with Islamic culture and religion. In slightly varying forms it is used for a number of other languages quite unrelated to Arabic, notably Persian and Urdu. It is no longer the usual script for Turkish, Malay, Swahili and Hausa, as it was until early in the 20th century.

Aramaic and Syriac

200,000 SPEAKERS

United States, Georgia, Iraq, Turkey, Iran, Syria

One of the SEMITIC LANGUAGES, Aramaic was an international language of the Near East from the 7th century BC and was the ruling language of the great Persian Empire from the 6th to the 4th centuries. Syriac, a later form of Aramaic, was the principal language of the borderland between the Roman and Parthian empires. Both Aramaic and Syriac have been important languages of religion. Tenuously, Aramaic still survives as a living language today.

In the Bible, Aram was one of the sons of Shem. The Aramaeans, whose ancestor he was supposed to be, first appear in Biblical and other Near Eastern texts as a nomadic people of the desert east of Palestine and south-east of Syria, threatening, and eventually conquering, Damascus and other cities in Syria and northern Iraq.

Old Aramaic inscriptions, the first direct evidence of the language, are found in all these regions and are dated to the 10th century BC onwards.

Classical (or Imperial) Aramaic was the form of the language that was used in the Assyrian, Babylonian and Persian empires and in the period that followed. It was under Persian rule that Aramaic spread most widely. In the last few centuries BC, inscriptions in Aramaic were erected all the way from the Greek-speaking Hellespont to the Prakrit-speaking plains of the lower Indus valley. Aramaic was a world language.

Yet, with the destruction of the Persian Empire by Alexander the Great and his successors, Aramaic was now no longer the national language of a major state. It remained in very widespread use in the Greek and other kingdoms that supplanted Persia, but it now gradually split into dialects, usually grouped into 'West Aramaic' and 'East Aramaic'.

In the course of the Jewish migrations recorded in the Old Testament, Aramaic had become the everyday language of the Jews. It was the language of some of the Dead Sea Scrolls, and it was certainly the daily language spoken by Jesus Christ. Although nearly all of the Old Testament itself was composed in Hebrew, there are Aramaic passages in the Biblical books of Ezra (iv.8 to vi.18, vii.12 to 26) and Daniel (ii.4 to vii.28). Later Jewish religious texts, notably the vast collection known as the *Jerusalem Talmud*, are in Aramaic throughout. In due course, with continued Jewish migrations, Aramaic too fell out of everyday use: but a Jewish religious education still embraces both Hebrew and Aramaic. This *Jewish Palestinian Aramaic* is normally written in Hebrew script.

Jewish Palestinian Aramaic is the best-known representative of 'West Aramaic'. To the same dialect group belong the languages, as recorded in many inscriptions, of two important states on the border of the Roman Empire – Palmyra (once called Tadmor) and Nabataea, whose capital was at Petra. The spoken language of Petra was apparently Arabic: its inscriptions are in a dialect of Aramaic that is clearly mixed with early Arabic forms. Samaritan, language of a separate Jewish community of Roman and early medieval Palestine, was a fourth dialect of 'West Aramaic'.

The surviving Samaritans of Nablus in Israel now speak Arabic, but their religious texts and ceremonies are in Samaritan Aramaic, which they continued to use as a literary language till the 19th century.

Centring in the originally independent kingdom of Edessa (now Urfa), to the north-east of Palestine and Roman Syria, *Syriac* became the vehicle of a flourishing Christian culture. The New Testament and many other Greek religious texts were translated into Syriac. An original literature of hymns, sermons and theological and historical works grew up in this form of Aramaic, with its distinctive alphabet (see box). Syriac was the medium through which early Christian writings reached Armenia and Arabia and were re-translated into those languages. Syriac literature is dated to the 3rd to 13th centuries AD.

East Aramaic includes Syriac, and also Mandaean and *Babylonian Aramaic*. The latter was the language of the Jews of the Parthian and later Persian Empire: it is the language, therefore, of the *Babylonian Talmud* and of many magical texts of the 4th to 6th centuries. *Mandaean* or Mandaic is the distinctive written language of the Gnostic Christians of Iraq, whose literature dates from the 3rd to 8th centuries.

Much further east, the Nestorian Christians took their Syriac language across central Asia, where Syriac inscriptions near Bishkek and Tokmak in Kyrgyzstan date back to the 14th century. For nearly two thousand years there was a Syriac-speaking Christian community, the 'St Thomas Christians', in south India. Their Bibles and other religious manuscripts in Syriac were rediscovered with astonishment by Western scholars in the 19th century, though by that time few if any of the congregations were able to understand the language.

Modern dialects of Aramaic are still spoken by small minorities in Iraq, Turkey, Iran and Syria (see map). Numbers are difficult to determine. The majority of speakers, it is said, are in émigré communities in Armenia and Georgia (where they migrated after the Russo-Persian War of 1827) and in the United States. It is a mark of the cohesion of these émigré communities that Aramaic was recognised as a 'national language of the USSR', with a new Cyrillic orthography. The major modern dialect, *Sūriṭ* or 'Assyrian', has 15 vowels, in three sets of five: each word shows vowel harmony, with all its vowels belonging to one set.

Periods of recorded Aramaic

Palestine, Jordan and western Syria		Eastern Syria and Iraq
Old Aramaic	1000–700 BC	Old Aramaic
Classical Aramaic	700–100 BC	Classical Aramaic
Palestinian, Samaritan, Palmyrene, Nabataean	100 BC–AD 700	Babylonian Aramaic, Mandaean
	AD 200–1300	Syriac
Modern West Aramaic	AD 1800 to date	Modern East Aramaic ('Modern Assyrian') émigrés

---

Modern Aramaic on the map

An East Aramaic dialect, *Ṭūroyo* (sometimes called 'modern Assyrian' or 'Neo-Syriac') is spoken by Christian communities of the Syrian Orthodox Church whose traditional homes are on the Ṭūr 'Abdīn plateau in Turkey. Their religious texts are in Syriac and Arabic, in the western Syriac script. Most of the 30,000 remaining Turkish speakers of Turoyo now live in Istanbul.

A second East Aramaic dialect, *Sūriṭ*, is spoken by 'Chaldaean' Christians in Iraq and in a few villages in eastern Turkey. Some Nestorian or 'Assyrian' Christian communities remain in Iran (the plateau of Urmiā), Iraq (near Mosul) and Turkey: they also speak Surit and they have a written form of their language for daily use. Both

Nestorians and Chaldaeans use classical Syriac, in eastern Syriac script, for worship. The majority of the Christian communities no longer live in these country districts. Most of those who have not emigrated now cluster in the big cities, Tehran and Baghdad. There were also Jewish communities speaking Surit, but they have migrated en masse to Israel, and may already have ceased to use their language.

Numerals in Aramaic

Biblical Aramaic		Sūriṭ
ḥadh	1	kha
tərēn	2	tre
təlāthā	3	ṭlå
'arbə'ā	4	årpå
ḥamshā	5	khamsha
shittā	6	ishtæ
sibh'ā	7	shåvå
təmanyā	8	tmænyæ
tish'ā	9	îchå
ashrā	10	îsrå

David Cohen, *Les langues chamito-sémitiques*
(Les langues dans le monde ancien et moderne
ed. Jean Perrot, pt 3) (Paris: CNRS, 1988)

The third East Aramaic dialect is *Modern Mandaean*, spoken near Akhwaz and Khorramshahr in Iran. In three villages near Damascus –

Ma'lūla, Jubb'adīn and Baḥ'a – a West Aramaic dialect is still spoken. It is sometimes called Ma'lūla.

Aramaic writing systems

Aramaic in writing

𐡀 𐡁 𐡂 𐡃 𐡄 𐡅 𐡆 𐡇 𐡈 𐡉 𐡊 𐡋 𐡌 𐡍 𐡎 𐡏 𐡐 𐡑 𐡒 𐡓 𐡔 𐡕
' b g d h w z ḥ ṭ y k l m n s ' p ṣ q r š t

The oldest known Aramaic alphabets (an example is shown in the top line of the first box) were developed from an early Phoenician script. These Aramaic alphabets had many local variants, such as the Nabataean shown in the third line of the box. They were the parents from which were developed the Syriac, Arabic and Hebrew scripts of medieval and modern times.

Still sometimes seen in print is the Syriac alphabet, which is essentially that of the Christian Syriac literature of Roman times and after. Syriac script, like the Arabic from which it derives, is able to indicate vowels by using diacritical marks. The second alphabet box shows the consonant signs of the East Syriac (above) and West Syriac (below) alphabets as now printed.

The Syriac alphabet

ܐ ܒ ܓ ܕ ܗ ܘ ܙ ܚ ܛ ܝ ܟ ܠ ܡ ܢ ܣ ܥ ܦ ܨ ܩ ܪ ܫ ܬ
' b g d h w z ḥ ṭ y k l m n s ' p ṣ q r š t
ܐ ܒ ܓ ܕ ܗ ܘ ܙ ܚ ܛ ܝ ܟ ܠ ܡ ܢ ܣ ܥ ܦ ܨ ܩ ܪ ܫ ܬ

Fonts from *Gamma Unitype* (Gamma Productions)

ARAUCANIAN

PERHAPS 300,000 SPEAKERS

Chile, Argentina

One of the AMERIND LANGUAGES, Araucanian is still spoken by one of the larger linguistic communities of the southern half of South America (see map at QUECHUA). It has no close relatives.

> Its speakers call themselves Mapuche, 'people of the earth [*mapu*]', and their name for their language is *Mapudungun*. In Spanish it is *Araucano*: this word comes from the name of the town *Arauco* 'limestone', as does the botanical name of the monkey puzzle tree, *Araucaria imbricata*.

Araucanian was once spoken over most of central Chile, from Copiapo, five hundred kilometres north of Santiago, southwards to the island of Chiloe, and all the way across Argentina to the bay of Comodoro Rivadavia. The warlike Mapuche did not submit easily to Spanish-speaking suzerains, and suffered repeated military defeats and massacres. Nowadays Araucanian speakers have almost disappeared from Argentina, while even in Chile Araucanian-speaking country now begins several hundred kilometres south of Santiago, around Temuco, the town that used to be called *Mapuchu cara*, 'city of the Mapuche'. Few speakers are now monolingual in Araucanian; most, if not bilingual, are at least accustomed to using Spanish as well as their mother tongue.

Araucanian has contributed to Spanish – and thus to English – the word *gaucho* (Araucanian *cauchu* 'nomad, adventurer'), which became the name for the mixed Araucanian and Spanish nomadic population of the pampas. Many Chilean place names are of Araucanian origin, such as *Caramavida* from *carü mahuida* 'green mountain'.

The first ten numerals in Araucanian are: *kiñe, epu, küla, meli, kechu, kayu, relke, pura, ailla, mari*.

Based on M. Malherbe, *Les langues de l'humanité* (Paris: Laffont, 1995) pp. 1196–200, and other sources

ARDHAMAGADHI

CLASSICAL LANGUAGE OF INDIA

One of the INDO-ARYAN LANGUAGES, Ardhamagadhi is the language of the canonical Jain scriptures. It is known in two forms. One, found in the earliest sutras, is roughly contemporary with Pali. The other, dating from around the 6th century AD, may be based on the spoken language of what is now eastern Uttar Pradesh; an almost identical dialect was also used for Buddhist dramas.

The Jains (Sanskrit *Jaina*, 'follower of a *jina* or victor') hold to a strongly ascetic religion that arose in northern India around the same time as Buddhism. It has never been a mass movement. Its believers are to be found mostly in Gujarat and Rajasthan. Most Jains have the surname *Shah*.

Ardhamagadhī is 'half-Magadhī', a name that links the language to the Prakrit dialect of north-eastern India, which was also claimed to be the closest relative of Pali, the language of Buddhism.

There is no single Jain holy book: the canonical scriptures as known today are incomplete and the two sects (*Dīgambara*, 'sky-clothed', whose ascetics go naked; *Svētāmbara*, 'white-clothed') disagree over their authenticity. Jains (including émigré communities, such as those in East Africa) still use Ardhamagadhi in religious ritual, though most have not learnt the language and do not understand what is said.

The first ten numerals in Ardhamagadhi are: *ege, do, tao, cattaro, paṃca, cha, satta, aṭṭha, nava, dasa*.

ARMENIAN

2,000,000 SPEAKERS

Armenia, Iran

For more than 1,500 years Armenia was ruled by Persia, or by kings who owed allegiance to Persia. As a result, Armenian is so heavily influenced by Persian that scholars once counted it as an Iranian language. It is now certain that Armenian forms a separate branch of the INDO-EUROPEAN LANGUAGES. It is spoken in the Republic of Armenia, in north-western Iran, and by several large and well-established Armenian communities in other parts of the world.

The Armenians call themselves *Hai*. The name of *Armenia* is found in Persian and Greek texts from the 6th century BC onwards.

Most Indo-European scholars think that early Armenian speakers reached eastern Asia Minor, perhaps as late as the first millennium BC, after migrating across Asia Minor from the west. No clear archaeological evidence has been found to support this, but Armenian seems to have similarities with Greek and with the extinct Indo-European dialects of the early Balkans. The language is spoken approximately where the Urartians, whose language was Caucasian and not Indo-European, reigned in the 9th to 6th centuries BC.

The history of Armenia becomes clearer after the first Persian conquest, around 500 BC. There was, from this time onwards, Greek cultural influence: Greek plays were performed, and even written, at the Armenian court. The overlordship of Armenia was to be a long-running dispute between the Roman and Sassanian (Persian) empires. Armenia was in classical times a far larger state than it is now, and the language was spoken across eastern Turkey as far south as Edessa, capital of the state of Osroene, which was bilingual in Syriac (see ARAMAIC) and Armenian.

Christianity in Armenia may be dated to the conversion of King Trdat by St Gregory the Illuminator in AD 314. Armenian literature began in the 5th century with literal translations of Christian texts: for the Armenian Orthodox church, though eventually independent, was at first heavily dependent on Greek teaching. Words coined by the translators, anxious to find a precise match for Greek and Syriac religious terminology, still survive in modern Armenian. Persian political dominance, meanwhile, gave rise to numerous loanwords. Words of Iranian and Persian origin actually form a majority of the entries in Armenian dictionaries.

Classical Armenian, the *grabar*, essentially the language of the translations and the earliest Armenian literature, is still taught. It is now a very different language from the modern spoken dialects. These, in the 19th century, formed the basis of two essentially new literary languages or dialects, Eastern Armenian of the present Republic of Armenia (which became independent from the Soviet Union in 1991), and Western Armenian of Turkey and the diaspora.

Rich in consonants like the CAUCASIAN LANGUAGES that surround it, full of foreign loanwords from its dramatic history, Armenian still has a clearly Indo-European structure. Nouns have seven cases, singular and plural. As in ancient Greek, classical Armenian verbs have an *e-* prefix to mark the past tense, though this is now only added to monosyllabic forms: *berē* 'he bears', *eber* 'he bore'.

South-west
Asia:
Armenian

Armenians across the world

Both in their own homeland and in their migrations, Armenians have long lived in a multilingual environment. Persian, Russian and Turkish have been necessary not only politically but also to communicate with near neighbours. There is a fine body of oral epic poetry in Armenia and eastern Asia Minor. It is not surprising that oral poets in recent centuries often had bilingual skills, in Armenian and Turkish, or even trilingual, able to entertain Persian-speaking audiences as well.

The *Hemsinli* are Armenians of Asia Minor who converted to Islam in the 15th and 16th centuries: they remain settled in valleys of Artvin province in Turkey, where many work in the transport business. *Posha* or *Boşa*, a variety of Armenian with many Indo-Aryan words, is spoken by a Gypsy-like nomadic group, the Lomavren, of Azerbaijan, Armenia and the Lake Van region of Turkey.

A considerable population of Armenians settled in the Crimea, under Genoese rule, in the 14th and 15th centuries. They eventually moved on to Poland and Transylvania in the 17th century, by which time they spoke not Armenian but a Kipchak Turkic dialect (which they wrote in Armenian script) and they had converted to Roman Catholicism.

Other Armenian colonies are also long-established. The first Armenian daily newspaper was published in Constantinople in 1832. The Armenian Mekhitarist monastery on the Venetian island of San Lazaro has printed Armenian literature and religious texts since its foundation by Mekhitar in 1717 – but Armenians had settled in Venice ever since the 13th century. Gherla (German *Armenierstadt*) and Dumbrăveni (German *Elisabethstadt*) in Romania were once Armenian towns. There have been large Armenian communities in Smyrna, Vienna, Lwów, Moscow, Cairo and Aleppo. Beirut, Paris, Montreal and Los Angeles are important Armenian centres today.

The Armenian dialects

The lands where Armenian was spoken have been politically divided for many centuries. In medieval times the south-western half, 'Cilician Armenia' or 'Little Armenia', makes frequent appearances in the history of the Crusades. 'Greater Armenia' to the north-east was approximately where the Republic of Armenia is now. As a result, already in the 10th century there were two dialects of spoken Armenian. One was the ancestor of the Eastern Armenian still spoken in the Republic of Armenia and in Iran. The independence of Armenia (the medieval 'Greater Armenia'), asserted at the Russian Revolution, was crushed by Turkish and then Russian troops. Independence was again declared in 1991.

The other dialect, Western Armenian, was the language of Cilician Armenia and was still spoken in that region until the genocide of 1913, when most of the Armenian Christians living within the borders of what is now Turkey (the majority Turkish speakers, the minority retaining their ancestral Armenian) were killed. Western Armenian is still spoken by the 75,000 Armenians of Istanbul and is the dominant dialect in the worldwide Armenian diaspora.

Nagorno-Karabakh (Mountain Karabakh, half of the old Khanate of Karabakh) is an Armenian-

speaking enclave within the borders of Azerbaijan. The latest Armenian attempt to annex Nagorno-Karabakh, in 1988, has led to war and the expulsion of hundreds of thousands of Armenians from Azerbaijan. Azeris likewise fled Armenia. Armenian traders still form large minorities in other cities of the Caucasus.

Armenian in writing

The Armenian script was invented by Mesrop Mashtots' in 405 so that religious texts could be translated into the local language from Syriac and Greek. According to legend, he invented quite different alphabets for Georgian and for a third language of the Caucasus then called Albanian.

The first *original* text in Armenian, appropriately enough, was a biography of Mashtots by his pupil Koriwn. The usual Armenian text font is small and slanting. For clarity a larger, rounded font, more appropriate to newspaper headlines, is used in the box. The transliteration given, with its range of diacritical marks, is the one preferred by the majority of linguists.

Numerals in Armenian

մէկ	1	mēk
երկու	2	erkow
երեք	3	erek'
ձորս	4	čors
հինգ	5	hing
վեց	6	vec'
եօթը	7	eōt'ə
ութը	8	owt'ə
ինը	9	inə
տասը	10	tasə

The Armenian alphabet

Ա Բ Գ Դ Ե Զ Է Ը Թ Ժ Ի Լ Խ Ծ Կ Հ Ձ Ղ Ճ Մ Յ Ն Շ Ո Չ Պ Ջ Ռ Ս Վ Տ Ց Ի Փ Ք Օ Ֆ

ա բ գ դ ե զ է ը թ ժ ի լ խ ծ կ հ ձ ղ ճ մ յ ն շ ո չ պ ջ ռ ս վ տ ց ի փ ք օ ֆ

a b g d e z ē ə t' ž i l x c k h j ł č m y n š o č' p ǰ ř s v t r c' w p' k' ō f

Aromunian

PERHAPS 200,000 SPEAKERS

Greece, Albania, Macedonia

Aromunian is one of the ROMANCE LANGUAGES, very close linguistically to ROMANIAN (see map there).

It is the most direct reminder of the fact that the Balkans belonged to the Roman Empire from the 1st century BC onwards. Latin was once spoken across much of south-eastern Europe. It is to be seen in the public announcements and private inscriptions still surviving in museums. It can be found in loanwords in Albanian, modern Greek and the south Slavonic languages.

The Latin speech of the Balkans was last recorded in the words of peasant soldiers in the Byzantine army in a battle of AD 587. And then it went underground for eleven hundred years.

The arrival of Slavonic speakers in large numbers from the 5th century onwards simply drove the Romance speech of the Balkans out of public view. Thus, while the medieval descendants of Latin became national languages in Spain, France and Italy, it shrank back to the more inaccessible mountain areas of north-western Greece, central Albania and southern Macedonia.

It is here – particularly in the Pindos mountains – that the *Aromunians* or *Vlachs* are now to be found. The word *Vlach* is the same historically as *Wallachian*, *Walloon* and *Welsh* – it had come to mean simply 'speakers of a strange language' – but, curiously, its origin appears to be the name of the Celtic tribe of eastern Gaul, *Volcae*, who were once on the linguistic frontier between Celtic and Germanic speech.

The Vlachs of northern Greece have been a highly prosperous community, prominent in local and long distance trade. The Gypsies of Turkey refer to Gypsies who are no longer nomadic as *Velakhos*.

There has been very little literature in Aromunian. A landmark is the Aromunian Haji Daniil's *Lexicon tetraglosson* 'four-language dictionary' (Voskopojë, 1764) whose aim was to teach Greek to the Albanian, Aromunian and Macedonian speakers of the region. In the 20th century Romanian scholars have been at the forefront of research on Aromunian, a language which gets no official recognition in its native territory.

Turn back! Turn back!

ἐπιχωρίῳ γλώττῃ εἰς τοὐπίσω τραπέσθαι ἄλλος ἄλλῳ προσέταττεν, **τόρνα**, *τόρνα* μετὰ μεγίστου ταράχου φθεγγόμενοι, οἷα νυκτομαχίας τινὸς ἐνδημούσης ἀδοκήτως αὐτοῖς.

In their peasant speech each man told the next to turn back, shouting *Torna, torna!* in great confusion, as if panicked by an unexpected night attack.

A unique record of the Romance speech of the Balkans, later to emerge as Aromunian, from the 7th-century Greek chronicler Theophylactus Simocatta.

Floară gălbinioară,	Yellow saw-wort,
dimând-a tutuloru,	ask them all,
dimând-a feateloru,	ask the girls,
se yină se-ñi me alumbă	to come to me
duminecă dimneaţa	on Sunday morning
şi luni de cătră seară,	and Monday in the evening,
cu roauă se-ñi-mi adună,	to gather me with the dew
ş-pre avră se-ñi-mi poartă	and bear me with the breeze
pre iapă nefitată,	upon an ass that has not foaled,
pre feată nemărtată,	a girl that has not married,
pre gione neinsuratu	a boy that has not taken a wife,
ş-pre cale necalcată.	and a road that is untrodden.

Aromunian folk song collected by C. Récatas. Every word is of Latin origin

After Th. Capidan, *Les Macédo-roumains* (Bucharest, 1937)

Assamese

12,000,000 SPEAKERS

India

Assamese is the easternmost of the INDO-ARYAN LANGUAGES of India. It is spoken in the Indian state of Assam, in the middle Brahmaputra valley, where that wide river flows south-westwards eventually to reach the Bay of Bengal. Until modern times this was the eastern outpost of the whole Indo-European language family.

Assamese, called by its own speakers *Asamiya*, is named after the Indian state (Assamese *Asam*, pronounced *Ohom*) which in turn gets its name from the Tai people who dominated the region from the 13th to the 19th centuries (see AHOM).

Assamese developed its special character because, unlike Bengali and Hindi, it was not simply a language of the Indian plains. Mountains shadow the lower Brahmaputra valley on both sides. Their inhabitants speak numerous quite unrelated languages, Tai, Austroasiatic and Sino-Tibetan. Assamese became the lingua franca of all these peoples, and in the process lost many of the features that almost all other Indian languages share – notably the retroflex consonants made by curling the tongue backwards towards the palate. Assamese, by contrast, shares with English the unusual feature of alveolar consonants (English *t, d*): these are formed when the tongue touches the alveolar ridge above the upper teeth.

Assamese vocabulary, basically inherited from SANSKRIT, includes borrowings from Khasi (*bhur* 'raft'), from Munda languages (*kadu* 'gourd') and from Ahom (*jeka* 'moist'). Bodo and related Sino-Tibetan languages are the source of many borrowed words (*celek* 'lick'; *gaba-mar* 'embrace';

thalamuri-mar 'slap') and apparently of the diminutive suffix *-ca* (e.g. *kala* 'black'; *kalca* 'blackish'). Bengali and English have both influenced Assamese strongly. For a table of numerals see BENGALI.

Assamese is first recognisable as a separate language from Bengali in poetry of about 1400 onwards. Sankara Deva, greatest of the Vaishnavite devotional poets, flourished around 1500. Assamese prose writing begins with translations of the Sanskrit *Bhagavadgītā* and *Bhāgavatapurāṇa* a century later. All this literature came from the small kingdoms of western Assam. Meanwhile at the Ahom court, to the east, Assamese became the court language in the 17th century, and prose chronicles, earlier composed in Ahom, began to be written in Assamese – of a noticeably modern form. Literature faded in the last half-century of Ahom rule, a chaotic period with rebellions, Burmese and British interventions, and massacres.

The British, who finally took control of Assam in 1826, tried to impose Bengali as the language of courts and education, but gave up in 1873 and made Assam a separate province of their Indian empire. *A few remarks on the Assamese language and on vernacular education in Assam*, by 'A Native' (Anandaram Dhekiyal Phukan), published at Sibsagar in 1855, influenced the change of policy. So did the publications of the American Baptists at the Sibsagar Mission Press, which encouraged the use of modern, colloquial language in Assamese literature.

Oriya, Bengali, Assamese: the map

Oriya is the language of Orissa State. *Sambalpuri* is the dialect of the Sambalpur lowlands, while

Bhatri, a quite distinct dialect, is spoken in Bastar District of Madhya Pradesh.

Malpaharia, an aberrant western dialect of Bengali, is spoken in the Rajmahal hills, in the Santal parganas of Bihar, where Malto was spoken until recently.

Bengali is divided into several dialect groups. Calcutta, whose language helps to form the standard in Indian West Bengal, belongs to the Central group; Dacca, which increasingly sets the standard for Bangladesh, speaks an Eastern dialect.

Sylheti, often counted as a separate language (see BENGALI), has as many as 5,000,000 speakers in Sylhet District, a hundred miles north-east of Dacca.

Rajbangshi and a group of related dialects extend Bengali northwards from Rangpur in Bangladesh towards the Darjeeling Terai and south-eastern Nepal.

The *South-eastern Bengali* dialect of Chittagong and Noakhali is so different from the standard that it has been considered a separate language.

Chakma, spoken in the Chittagong Hills, has its own script, resembling that of some southeast Asian languages. It claims 68,000 speakers in Tripura and Mizoram, India, and more in Bangladesh.
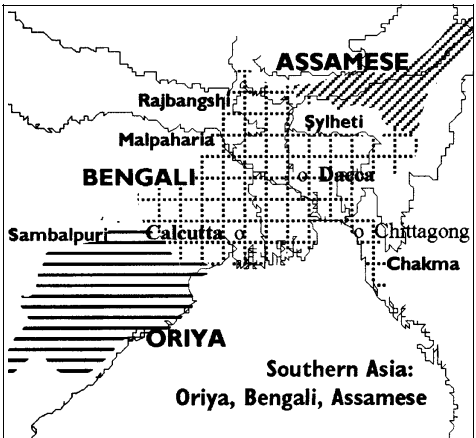
The influence of the Mission Press at Sibsagar (continuing that of the Ahom court!) ensured that the dialect of eastern Assam would become the modern standard for *Assamese*. This eastern dialect, fairly uniform because the region has been politically unified for several centuries, is spoken on both banks of the Brahmaputra from Sadiya down to the modern capital Gauhati.

The western dialects of Kamrup and Goalpara Districts (historically separate small kingdoms) are very different from eastern Assamese and from one another. Kamrupi stresses initial syllables of words, like Bengali.

Naga Pidgin or *Nagamese* is a variety of Assamese – perhaps a kind of creole, but linguists differ over definitions here – that has become the lingua franca of the polyglot Indian state of Nagaland. It has at least 500,000 speakers. Among some Naga peoples it has been used for 150 years or more in trading with one another and with the valley of Assam. Although now used informally in many schools, Naga Pidgin has so many variant forms that it is difficult to produce an acceptable set of textbooks in it.

Assamese script

অ আ ই ঈ উ ঊ এ ঐ ও ঔ কখগঘঙ চছজঝঞ টঠডঢণ তথদধন পফবভম য৵লৱ শষসহ

a ā i ī u ū e ai o au k kh g gh ṅ c ch j jh ñ ṭ ṭh ḍ ḍh ṇ t th d dh n p ph b bh m y r l v ś ṣ s h

---

Southern Asia:
Oriya, Bengali, Assamese

Assamese in writing

The Assamese alphabet is almost the same as the Bengali, though the sound system is quite different. Assamese spelling is not phonetic, but it does help to show the origins and derivations of words. In the early 19th century the influential Baptist Mission Press introduced a modernised spelling closer to actual pronunciation, but then gradually reverted to the traditional Sanskritised spelling familiar to educated Assamese readers.

AUSTRALIAN LANGUAGES

When Europeans first began to explore and settle in Australia, about three hundred languages were spoken by the hunter-gatherer peoples (possibly 300,000 people in total) who inhabited the continent. Now there are fewer than 150 languages, and few indeed will survive beyond the next generation. The language that rules Australia is ENGLISH.

Australia is the 'Southern Land', already named *Terra Australis* in Latin on 16th-century maps, long before any European had seen it.

Most scholars claim that all Australian languages belong to a single family, and that within it a single branch, Pama-Nyungan, accounts for the indigenous languages of nine-tenths of Australia. Far greater linguistic diversity is found in the Arnhem Land region, the northern third of Northern Territory and the north-eastern segment of Western Australia.

The languages of Tasmania were perhaps not related to those of Australia. The two land masses were separated, 12,000 years ago, by a rise in sea level that flooded the Bass Strait. There were as many as twelve languages spoken on the island when Europeans began to settle there. Many speakers were exterminated, and the languages ceased to be spoken around 1900 before any serious linguistic records had been made, though a few amateurish wordlists and one or two recordings of songs do exist.

Research on Australian languages has been slow to develop, though some wordlists were made in the late 18th and 19th centuries. The first book about an Australian language was the Rev. L. E. Threlkeld's introduction to Awaba, a now extinct language of the Lake Macquarie region, published in 1827.

Australian loanwords in English

Kangaroo: Guugu Yimidhirr, the language of the Endeavour Bay district, is the source of the first Australian loanword to reach English. In that language *gangurru* is the term for a large black kangaroo, the male *Macropus robustus*. The word was noted down by Sir Joseph Banks, a member of Captain Cook's exploring party in 1770.

Boomerang: The *bumariñ*, as it is called in the Dharuk language of New South Wales, was first described in English in 1825: 'a short crested weapon which the natives of Port Jackson propel into a rotary motion, which gives a precalculated bias to its forcible fall'. By 1846 it was famous enough to find its way into a simile in an American poem:

Like the strange missile which the Australian throws,

Your verbal boomerang slaps you on the nose.

Based on *Australian words and their origins* ed. Joan Hughes (Melbourne: Oxford University Press, 1989) and other sources

---

In the last few decades much has been done to record dying languages, often by eliciting almost-forgotten vocabulary from the last one or two surviving speakers. A great deal of oral literature and music has been collected. Most Australian peoples tell legends in prose. Songs are reflective rather than narrative, but some genres of songs can be linked with the legend cycles.

In most or all Australian languages there is a separate speech register, with different vocabulary and sometimes a different sound pattern, used when speaking in the presence of a taboo

relative such as a man's mother-in-law (see box). Special registers or secret languages may also exist for use among young men undergoing initiation.

The first external influence on Australian languages came from MAKASAR traders, who began to visit Arnhem Land in the 17th century. A few resulting loanwords have been identified in languages of the north-west, such as *rrupiya* 'money' which came via Makasar from Sanskrit *rūpya* 'wrought silver' (compare Indian English *rupee*).

The first English-speaking settlements, notably the prison camp at Botany Bay, followed soon afterwards. Interaction with settlers of British and European origin, who now occupy most of Australia and rule it all, has been fatal for its indigenous peoples and cultures. The way of life of many tribes has been destroyed by disease, land seizure, forced migration, money and religion.

Australian languages are alike, historically, in having no numeral system: only 'one', 'two', 'several' and 'many' could be specified verbally in most languages. Days could also be counted in sign language by pointing at different areas of the palm of the hand. Many languages have now borrowed English numerals, which are necessary in dealing with money and other once-alien concepts.

'Mother-in-law language' in Dyirbal

The most extreme example of taboo language is that of Dyirbal-speaking tribes: here every single lexical word, except a few kin terms, has a different form in the *Jalnguy* 'avoidance' and *Guwal* 'everyday' styles. 'Blue-tongue lizard' is *banggara* in Guwal but *jijan* in Jalnguy, while 'ring-tail possum' is *midin* and *jiburray* respectively.

There are not, however, as many lexical forms in Jalnguy as there are in Guwal. Typically there is a one-to-many relationship: *jijan* is in fact the Jalnguy term for any lizard or guana, while *jiburray* covers all possums, squirrels and gliders. That is:

Guwal	Jalnguy
banggara 'blue-tongue lizard'	
biyu 'frilled lizard'	
buynyjul 'red-bellied lizard'	*jijan*
gaguju 'water skink'	
bajirri 'water goanna'	
midin 'ring-tail possum'	
jula 'striped blue possum'	
mungany 'Herbert River ringtail possum'	*jiburray*
yiwarrmany 'stinking honeysucker possum'	
burril 'flying squirrel'	

These correspondences reveal how speakers of Dyirbal, subconsciously, classify the natural world.

After R. M. W. Dixon, *The languages of Australia*
(Cambridge: Cambridge University Press, 1980) p. 61

AUSTROASIATIC LANGUAGES

To trace the history of the Austroasiatic languages is a challenge that no linguist has yet faced up to fully. The membership of the family has, in part, been in doubt until very recently. Only two of the languages in the family, Vietnamese and Khmer, have national status, yet it extends across much of southern Asia. Its speakers are, some of them, hunter-gatherer peoples; some, mountain farmers highly unwelcoming to outsiders; some, bearers of old and long-recorded lowland civilisations. They seem to have little in common.

It has long been realised that KHMER and MON were related, and the well-established group of Austroasiatic languages typified by these two is universally known as *Mon-Khmer*. Both are recorded in early inscriptions as languages of medieval Buddhist kingdoms heavily influenced by the culture of classical India. Mon rule was supplanted by Burmese and Thai expansion, and the language is now in decline; Khmer survives. Close to Khmer are some minority and hill languages of south-east Asia, including Sre, Mnong, Stieng, Bahnar, Hrê, Sedang, Kuy, Bru, Sô and others.

The Mon-Khmer family was soon seen to include a northern group whose largest members are WA, Palaung, Khmu and – far off in north-eastern India – KHASI. Many languages of very small communities belong to this group, including Mlabri, spoken by 300 hunter-gatherers called 'Spirits of the yellow leaves'.

The *Aslian* languages of inaccessible districts in the Malay Peninsula include Sengoi (Semai), Temiar, Orang Benua and others with even smaller numbers.

More recently it has been shown that the *Munda* or Kharia group of languages, spoken entirely in east central India, is distantly related to Mon-Khmer, and the name Austroasiatic came into use to reflect this broader level of classification. The Munda group includes Korku, Bhumij, SANTALI, MUNDARI, Ho, Kharia, Sora and others. The division between the Mon-Khmer group and the Munda languages must have taken place many thousands of years ago.

Much more controversial has been the inclusion of VIETNAMESE and MUONG (with some tiny minority languages of Vietnam and Laos) in the Austroasiatic family. This is because Vietnamese – the only member of this *Viet-Muong* group on which much work has been done – has for two thousand years been under the influence of Chinese. Whatever its shape at the beginning of this period, Vietnamese is now a tonal language with a sound pattern rather resembling that of Chinese. Moreover, it was traditionally written in Chinese script and its grammar and style had adopted many Chinese features. Some thought it a Sino-Tibetan language or tried to trace links with the Tai group. The resemblances between Vietnamese and its Austroasiatic neighbours were hard to see; yet they have now been demonstrated to the satisfaction of nearly all specialists.

Nowhere do the speakers of major Austroasiatic languages appear to be recent migrants. This is a language family that must once have filled more of the map of south-east Asia and eastern India, now reduced to scattered islands by encroaching Indo-Aryan, Sino-Tibetan, Tai and Austronesian languages including Bengali, Assamese, Burmese, Thai, Lao and Cham.

Languages of the Andaman and Nicobar Islands

On the Nicobar Islands of the Bay of Bengal a Nicobarese group of Mon-Khmer languages is recognised – Car, Nancowry, Great Nicobarese

and others. In these languages taboo leads to word avoidance on a considerable scale, and so to a rapid turnover of vocabulary. Thus, although they look very different from one another and from other Mon-Khmer languages, the relationships are historically closer than this would suggest.

On the Andaman islands two entirely separate language families exist – not clearly related to each other, or to the Austroasiatic languages, or to any other family, although Joseph Greenberg has postulated a relationship between these and the Papuan and Tasmanian languages. The first wordlist of an Andamanese language was made in 1795. The survival of the half dozen remaining Andamanese languages is now threatened, since there are fewer than 500 speakers in total.
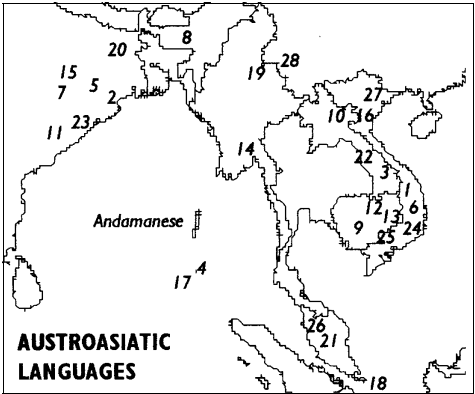
AUSTROASIATIC LANGUAGES

Austroasiatic languages on the ground			
Bahnar	1	85,000	Vietnam
Bhumij	2	200,000	India
Bru	3	120,000	Vietnam, Laos
Car	4	15,000	Nicobar Islands
Ho	5	750,000	India
Hrê and Sedang	6	150,000	Vietnam
Kharia	7	150,000	India
KHASI	8	500,000	India
KHMER	9	8,000,000	Cambodia
Khmu	10	350,000	Laos and other countries
Korku	11	300,000	India
Kuy	12	650,000	Cambodia
Mnong	13	200,000	Vietnam
MON	14	200,000	Burma, Thailand
MUNDARI	15	850,000	India
MUONG	16	800,000	Vietnam
Nancowry	17	5,000	Nicobar Islands
Orang Benua	18	10,000	Riau Islands
Palaung languages	19	500,000	Burma, China
SANTALI	20	4,000,000	India, Bangladesh
Sengoi or Semai	21	15,000	Malaysia
Sô	22	130,000	Laos, Thailand
Sora	23	270,000	India
Sre or Koho	24	100,000	Vietnam
Stieng	25	70,000	Vietnam, Cambodia
Temiar	26	10,000	Malaysia
VIETNAMESE	27	55,000,000	Vietnam
WA	28	1,000,000 or more	Burma, China

AUSTRONESIAN LANGUAGES

This is a family of over a thousand languages with a total of perhaps 270,000,000 speakers who live from Madagascar eastwards to Easter Island, and from Taiwan southwards to New Zealand.

Most of these thousand languages are spoken by very small groups in regions that are still almost untouched by world politics and communications. But one of them is Malay (Indonesian and Malaysian), which has long been crucial to the spread of trade and culture in the south-east Asian archipelago and is now a major international language.

> *Austronesian*, a modern Graeco-Latin formation meaning 'of the southern islands', is a good name – accurate, yet not tied too tightly to any existing geographical or political label. It was devised by Wilhelm Schmidt in 1899. The first detailed reconstruction of *Uraustronesisch* 'proto-Austronesian' was worked out by the German scholar Otto Dempwolff between 1920 and 1938, but the family relationship had been noted as long ago as the 17th century. *Malayo-Polynesian* is sometimes used as a synonym for Austronesian, sometimes as a name for a grouping that excludes the languages of Taiwan.
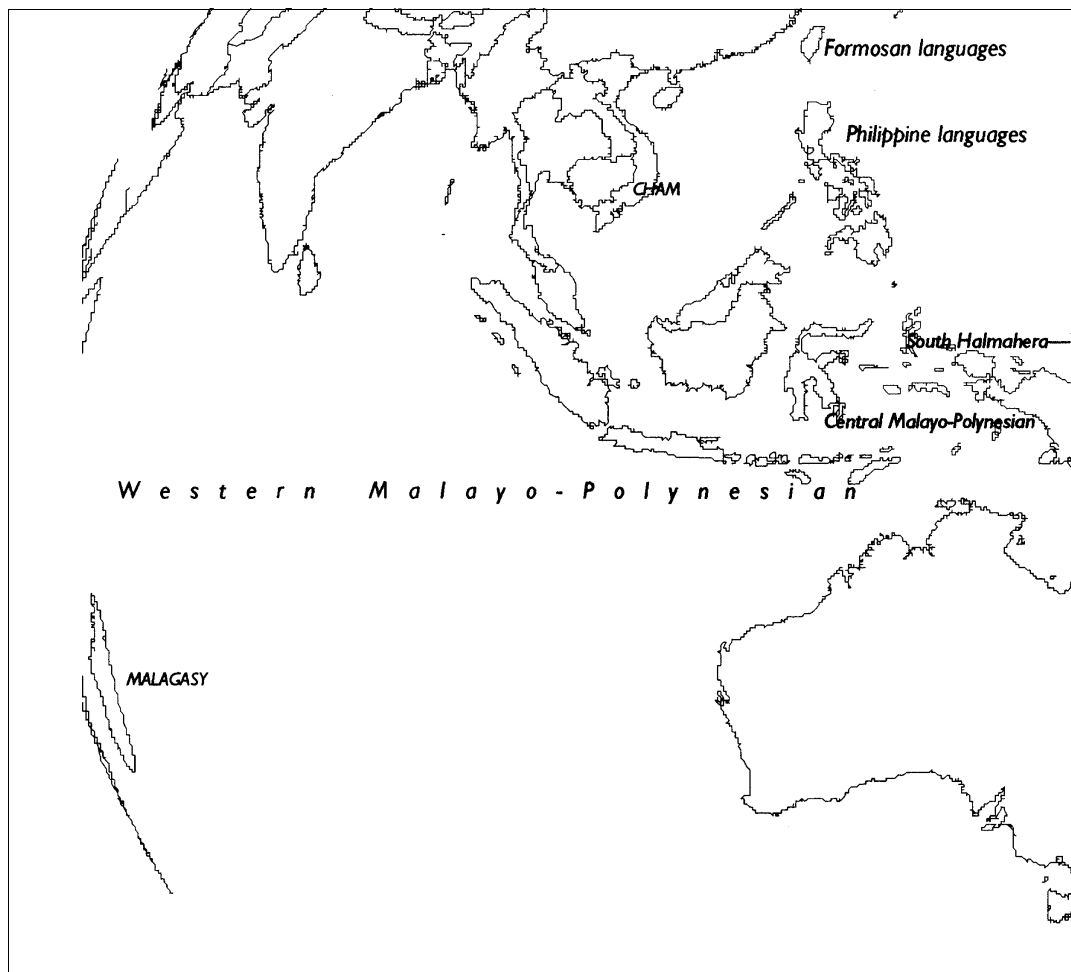
When all these related languages are set side by side, there is the most striking diversity in the relatively small region of highland Taiwan. This diversity represents very long-established dialect divisions and separate development. It suggests that from Taiwan itself or somewhere nearby, in a series of seaborne migrations, the people who spoke early Austronesian languages migrated to spread their culture and language across the Indian and Pacific Oceans. Looking back to an even earlier stage of language prehistory, it is becoming accepted that Austronesian languages are distantly related to several language groups of southern China and south-east Asia in a larger family of AUSTRO-TAI LANGUAGES.

The proto-Austronesian language, then, is thought to have been spoken on Taiwan and perhaps on the adjacent coast of south-eastern China, and it may have begun to differentiate into four dialects about 6,000 years ago. Three of these long-forgotten dialects were the ancestors of the three language groups now spoken by the 'aboriginals' of Taiwan (see FORMOSAN LANGUAGES), which are now gradually giving way to elite Chinese. The fourth dialect is what would now be called proto-Malayo-Polynesian. It is the ancestor of all the remaining modern languages of the family – and its speakers were evidently mariners. For the distinctive feature of Malayo-Polynesian, as contrasted with all other major language groupings of the world, is that to reach their present locations, early forms of every one of these languages must have been carried across the sea.

The Austronesian family is a glorious laboratory for comparative linguistics. Groupings and subgroupings should be relatively easy to work out, because (1) few of the languages have been influenced, until very recently, by languages of other families, and (2) mutual influences ought to have been limited by the seas and mountains that separate many of them from one another. In practice, Austronesian subgroupings have been controversial.

What do the subgroupings mean in historical terms? In most cases – as with the Formosan languages – on one side of each successive dividing line will be the languages that 'stayed behind', and on the other side will be the single proto-language spoken by those who 'moved on', the ones who colonised a whole new island or island group.

Formosan languages

Philippine languages

CHAM

South Halmahera

Central Malayo-Polynesian

W e s t e r n   M a l a y o - P o l y n e s i a n

MALAGASY

Taboos and secret languages

In many Austronesian languages, ethnologists have reported, taboo and forbidden topics force speakers to use circumlocutions or secret languages. The word *taboo* is a loanword in English, apparently borrowed from Tongan at the time of Cook's expedition.

As J. G. Frazer noted in *The Golden Bough* (London: Macmillan, 1911–15), Achehnese fishermen used a special language when fishing; Malay tin-miners and Sumatran gold-miners had a special language when at work, and collectors of aromatics, such as camphor and eagle-wood (see box at CHAM), whatever their religion, avoided offending forest spirits when gathering the valuable product.

In many parts of the Austronesian-speaking territory, from Madagascar to Tahiti and Maori New Zealand, names of parents-in-law, of children-in-law, of chieftains, and of the dead, are tabooed. They must be avoided in speech – and, often, words that sound like these names must be avoided as well.

Dusun and Kadazan dialects are spoken by the largest indigenous ethnic group of the east Malaysian province of Sabah. Priestesses, *bobohizan*, communicate with the dead using the so-called 'ancient Kadazan' language, which has a wholly different vocabulary.

AUSTRONESIAN LANGUAGE GROUPS

CHAMORRO

HAWAIIAN

O c e a n i c

West New Guinea

O c e a n i c

SAMOAN

FIJIAN

TONGAN

MAORI

Austronesian languages: the major groupings

The map identifies the subdivisions of Austronesian that are now accepted by most linguists. It also shows by name some of the languages that have separate entries in this book.

The FORMOSAN LANGUAGES remain to show where the series of migrations began.

From here, by way of the Batan Islands, Austronesian speakers spread across the whole of the Philippines (modern languages include TAGALOG, PAMPANGAN, ILOCANO, PANGASINAN, BIKOL, CEBUANO and HILIGAYNON) and to Guam (CHAMORRO), then gradually expanded westwards to the western half of Indonesia (BUGIS, MAKASAR, SASAK, SUNDANESE, JAVANESE, MADURESE, IBAN, MALAY, MINANGKABAU, LAMPUNG, REJANG, BATAK, ACHEHNESE) with further migrations to the coast of Indochina (CHAM) and to Madagascar (MALAGASY). This first expansion formed what is now known as the Western Malayo-Polynesian (once called *Indonesian*) language group.

A new migration went from the Philippines southwards, and this is the origin both of the Central Malayo-Polynesian languages (which are spoken in the Moluccas and Lesser Sunda Islands) and the much larger Eastern Malayo-Polynesian group.

One Eastern subgroup expanded across South Halmahera and the western extremity of New Guinea (Irian). The other apparently moved along the northern coast of New Guinea – and that was where the next great migration began, in a gradual expansion across the islands of Melanesia, Micronesia and Polynesia that gave rise to the Oceanic group of languages. These include FIJIAN, KIRIBATI, MARSHALLESE, SAMOAN, TAHITIAN, TONGAN and many other languages with very small numbers of speakers. The most distant migrations were those that gave rise to HAWAIIAN, MAORI and the language of Easter Island, *Rapanui*. The remarkable Polynesian migrations took Austronesian languages across a vast area of ocean, 5,000 miles from north to south and 6,500 miles from west to east.

Austronesian languages: the individual languages

This is a list of all the languages of the vast Austronesian group that have more than 100,000 speakers.

Language name	Number of speakers	Location
Abung	500,000	southern Sumatra
ACHEHNESE	2,400,000	northern Sumatra
Aklanon	350,000	northern Panay
Amis (see FORMOSAN LANGUAGES)	130,000	Taiwan
Atoni or Timor	650,000	western Timor
Bajau and Mapun	115,000	Sulawesi, Palawan, Sabah
BALINESE	3,000,000	Bali, Lombok, eastern Java
Banggai	100,000	central Sulawesi
BATAK LANGUAGES	3,500,000	Sumatra
Biak	40,000 and as lingua franca	off Bird's Head
BIKOL	3,000,000	Luzon
Bima	500,000	eastern Sumbawa
Bingkokak	150,000	south-eastern Sulawesi
Blaan	200,000	Mindanao
Bolaang Mongondow	900,000	north-eastern Sulawesi
BUGIS	3,600,000	southern Sulawesi
Capiznon	450,000	north-eastern Panay
CEBUANO	12,000,000	Visayas, Mindanao
CHAM	235,000	Cambodia, Vietnam
Davaweño	125,000	Mindanao
Dobu	8,000 and as lingua franca	Milne Bay
Dusun and Kadazan	280,000	Sabah
Ende and Lio	220,000	central Flores
FIJIAN	340,000	Fiji
Gayo	180,000	northern Sumatra
Gorontalo	900,000	northern Sulawesi
HILIGAYNON	4,600,000	Visayas
Motu and HIRI MOTU	15,000 and as lingua franca	Papua New Guinea
IBAN and Sea Dayak languages	1,200,000	Borneo coasts
Ibanag	300,000	Luzon
Ifugao dialects	110,000	Luzon
ILOCANO	5,300,000	Luzon
Itawit	100,000	Cagayan Province
Jarai	200,000	Vietnam
JAVANESE	75,000,000	Java
Kankanaey	180,000	Luzon
Kei	85,000 and as lingua franca	Kei and Kur Islands
Kinaray-a	300,000	south-western Panay
Komering	700,000	southern Sumatra
Konjo	200,000	southern Sulawesi
Lamaholot	310,000	Solor, Lomblen, Pantar and Alor islands
Lauje	125,000	central Sulawesi
Lawangan	120,000	southern Kalimantan
Ledo	130,000	Sulawesi
MADURESE	9,000,000	Madura
Magindanaon	915,000	Mindanao

MAKASAR	1,600,000	southern Sulawesi
MALAGASY	10,000,000	Madagascar
MALAY and related dialects	30,000,000 and as lingua franca	(see map at MALAY)
Mandar and Mamuju	345,000	western Sulawesi
Manggarai	400,000	western Flores
Manobo dialects	320,000	Mindanao
MAORI	100,000	New Zealand
Maranao	600,000	Mindanao
Masbateño	330,000 and as lingua franca	Masbate Province
Masenrempulu dialects	210,000	southern Sulawesi
MINANGKABAU	6,500,000	Sumatra
Muna	200,000	Muna and Buton islands
Ngaju	250,000	Borneo
Nias, Sikule and Simalur	600,000	off Sumatra
Pamona	105,000	southern Sulawesi
PAMPANGAN	1,850,000	Luzon
PANGASINAN	1,635,000	Luzon
Pesisir	400,000	southern Sumatra
Pubian	400,000	southern Sumatra
Rade	120,000	Vietnam
REJANG AND LAMPUNG	2,000,000	Sumatra
Romblomanon	200,000	Romblon and Sibuyan islands
Roti	130,000	Roti, Timor
Sama and Bajau	250,000	Philippines, Sabah; language of the 'Sea Gypsies'
SAMOAN	325,000	Samoa
Sangir	205,000	Sangir
SASAK	2,100,000	Lombok
Sawu or Hawuon	100,000	Sawu, Raijua, Sumba, Flores, Timor
Sikka	175,000	Flores
Subanun dialects	155,000	Mindanao
Sumba	200,000	Sumba
Sumbawa	300,000	western Sumbawa
SUNDANESE	27,000,000	western Java
TAGALOG	10,500,000 and as lingua franca	Philippines
TAHITIAN	100,000	Tahiti
Tausug or Sulu	480,000 and as lingua franca	Philippines, Borneo
Tetun	300,000	Timor
Tolai or Kuanua	60,000 and as lingua franca	Gazelle Peninsula
Tolaki	125,000	south-eastern Sulawesi
TONGAN	130,000	Tonga, Tuvalu
Tontemboan	150,000	Sulawesi
Toraja or Sadan	500,000	southern Sulawesi
WARAY-WARAY	2,400,000	Samar, Leyte, Sorsogon

Austro-Tai languages

A series of distant relationships may link the AUSTRONESIAN LANGUAGES, the KADAI LAN-GUAGES (including TAI LANGUAGES), the family consisting of MIAO and YAO, and possibly JAPAN-ESE. The family relationship was proposed by Paul Benedict in a series of papers beginning in 1942, reprinted in his *Austro-Thai language and culture, with a glossary of roots* (New Haven: HRAF Press, 1975).

Benedict's idea cut across the once generally accepted link between Tai and SINO-TIBETAN LANGUAGES, for which, however, there had never been much evidence beyond a number of close word resemblances between Tai languages and Chinese. It also denied the family link ('Austric') which had long been suggested between Aus-tronesian and AUSTROASIATIC LANGUAGES. This suggestion, made by Wilhelm Schmidt in 1906, is one which some Austroasiatic specialists still consider likely.

In both cases the resemblances, so Benedict argued, were due to borrowing. In later work he has argued more specifically for borrowing from an early Tai language into Chinese rather than from Chinese to Tai – an idea which in itself was controversial. Among the languages of his Aus-tro-Tai family, meanwhile, Benedict was cer-tainly able to list a surprising number of words that seem to go back to common roots: the monosyllabic words of the mainland languages corresponding, in most cases, to the stressed syllable of polysyllabic proto-Austronesian words.

Although few would consider it proved, the Austro-Tai hypothesis remains for the present the most promising of all the attempts to find distant family relationships among the languages of south-east Asia.

Languages of south-east Asia

By 3000 BC, rice cultivation was practised across the whole region from the Yangtze delta to northern Thailand. AUSTRONESIAN LANGUAGES, KADAI LANGUAGES, TAI LANGUAGES, MIAO and YAO may all derive from prehistoric languages spoken in this swathe of territory. Speakers of early SINO-TIBETAN LANGUAGES and AUSTROASIATIC LAN-GUAGES, with different cultures, may have bor-dered on these groups to the north and to the south-west.

Today the speakers of Austronesian languages have spread from island to island across the Indian and Pacific Oceans. The other language families mingle on the mainland in such a com-plex pattern, with such pervasive mutual influ-ences, that their relationships and their prehistory are extremely difficult to trace.

Avar

500,000 SPEAKERS

Russia, Azerbaijan

Avar is the best known of the North East Caucasian languages, spoken by mountain peoples in south-western Dagestan, the Russian republic beside the Caspian. There is an Avar-speaking minority in newly independent Azerbaijan.

Speakers call themselves *Ma'arulal*; in Russian they are *Avartsy*. Historically, standard Avar is the dialect of Khunzakh and was known as *Bol mats* 'military language'.

The name 'Avar' has a long history: a tribe called *Avares* was linked with the Huns, who invaded the Roman Empire in the 5th century. These Avares were still a threat to Byzantium a hundred years later, but disappear from the historical record after being defeated by Charlemagne in 796. Their connection with the people now called Avar is uncertain.

In the last few centuries, Avar speakers have dominated the communications and trade of multilingual southern Dagestan, and the Avar Khanate was already establishing itself as a force independent of the Golden Horde by around 1500. It came under Russian control between 1803 and 1821.

Nowadays Avar is an official and literary medium shared by speakers of a group of languages – the 'Avar-Andi-Dido languages' – which are distantly related to one another but have clearly been developing separately for many hundreds of years. Even the four main dialects of Avar itself – Khunzakh, Antsukh, Charoda and Gidatl – are so distinct from one another that they are not mutually intelligible. In areas of linguistic fragmentation a lingua franca is needed, and this role was played by the dialect of Khunzakh in western Dagestan, a centre of trade and military activity since the 16th century. Essentially, the lingua franca based on the dialect of Khunzakh is what is now known as Avar. It serves also as a second or third language for highland speakers of DARGWA.

The local languages consist of the Andi group (Andi, Botlikh, Godoberi, Akhvakh, Bagulal, Tlisi, the eight Chamalal dialects, Karata and Tindi), the Dido or Tsez group (Tsez, Khvarshi, Bezheta, Hinukh and Khunzib) along with the more distantly related Archi: for a map see CAUCASIAN LANGUAGES. None has more than ten thousand speakers: some have only a hundred or so. After the Russian Revolution these small linguistic communities were at first recognised as separate 'nationalities' of the Soviet Union. They have been gradually brought together under the name of Avar, and the local languages are now in decline.

Numerals in Avar, Andi and Dido		
Avar	**Andi**	**Tsez**
1 ts'o	se-	sis
2 k'igo	ch'ègu	qano
3 l'abgo	l'òbgu	l'ono
4 unqo	-òqogu	uyno
5 shugo	ìnshtugu	l'eno
6 ant'go	ònt'gu	el'no
7 ankgo	hok'ugu	ot'no
8 mik'go	bìyt'ugu	bit'no
9 īch'go	hòch'ogu	och'ino
10 ants'go	hòts'ogu	ōtsino

Avar speakers are in general Muslims: they claim that Islam was introduced to the region

as early as the 8th century. Until the beginning of this century the language of culture here was Arabic. Avar, occasionally written since the 17th century, traditionally used Arabic script. In 1928 the Latin alphabet was introduced. As with so many Soviet minority languages, Cyrillic script became standard in 1938.

AVESTAN

EXTINCT LANGUAGE OF IRAN

The sacred language of the Zoroastrian religion is known only from a single body of texts, the *Avesta*, with an adventurous history.

According to a late source Zarathustra (Greek *Zoroaster*) promulgated his religion three hundred years before the invasion of Alexander – thus around 630 BC. It soon achieved royal status, for the Persian Emperor Darius I (550–486 BC) and his successors were Zoroastrians.

Legends claimed that a vast body of sacred texts was written down at this early period: two copies existed, one of which was burnt accidentally while Alexander the Great destroyed the other. Some centuries after this first disaster a Parthian king, Vologeses, ordered all that could be found of the old sacred books to be collected; later still, around the 5th century, the Sassanian monarchs had a new edition made. The Islamic conquest of Persia led to fresh destruction, after which, once again, surviving fragments had to be pieced together.

Independently of these historical snippets we know that the surviving manuscripts of the *Avesta* texts are late and obviously incomplete. And we know the origin of the unique alphabet in which the *Avesta* is traditionally written: it is an enlargement of the kind of Aramaic alphabet used in Sassanid times. Finally, we know that the language of the older texts is much older than that – much nearer to proto-Iranian, ancestor of all the IRANIAN LANGUAGES. As we have them, the texts are accompanied by a translation and commentary, the *Zend*, in Middle Persian of the Sassanian period. By that time the real meaning was half-forgotten. Even if we cannot always understand the *Avesta* ourselves, we can tell that those Sassanian translations are, all too often, wrong!

Attempts to pin down the Avestan language geographically have not yet succeeded. It was not the language of the Persians of the Empire, for that was Old PERSIAN. It seems to have features of several of the Iranian dialects. No doubt the 'original' language will have been altered, repeatedly, in the course of oral transmission until, perhaps quite late in their history, the texts were fixed in writing.

There are modern Zoroastrian communities still surviving in the Iranian cities of Yazd and Kerman. A thriving Zoroastrian colony, the *Parsees*, has spread from its early centre of Bombay to

---

In praise of Mithra

Ahe raya xvarənaŋhača	For his glory and fortune
təm yazāi surunvata yasna	I will praise aloud
Miθrəm vouru-gaoyaoitīm zaoθrābyō.	Mithra of the wide pastures with libations.
Miθrəm vouru-gaoyaoitīm yazamaide	Mithra of the wide pastures we worship,
rāmašayanəm hušayanəm	giver of safe and comfortable dwellings
airyābyō daiŋhubyō.	to the Iranian lands.

Adapted from *The Avestan hymn to Mithra* ed.
Ilya Gershevitch (Cambridge: University Press, 1967) pp. 74–5

Gershevitch uses the transliteration preferred by Iranian scholars. It is followed here and in the text. He suggests that the often-recurring epithet *vouru-gaoyaoitiš* 'having/providing wide cattle-pastures' is the origin of the by-name *Cautes* given to Mithras when he was adopted as a god of the Roman army.

other cities of the west coast of India, to East Africa and to many parts of the world. To all these the *Avesta*, in its mysterious original language, is still a holy book. Modern Parsees say their household prayers in Avestan, in words that they understand through traditional Gujarati translations and commentaries.

The first ten numerals in Avestan are: *aivas, duvā, trāyaṣ, čatvāraṣ, panča, xšvaš, hapta, aštā, nava, dasa*.

AYMARA

2,000,000 SPEAKERS

Bolivia, Peru

One of the AMERIND LANGUAGES, Aymara is spoken on the high Andes plateaus near Lake Titicaca (see map at QUECHUA).

Aymara shows many similarities with neighbouring Quechua. An argument continues as to whether the languages have the same origin, or have grown together in the course of shared cultural development. Hermann Steinthal, at the 8th International Congress of Americanists in Berlin in 1888, asserted the former. J. Alden Mason, in the *Handbook of South American Indians*, argued that in their basis the languages had 'little in common' but that they shared a large number of words, 'perhaps as much as a quarter of the whole, obviously related and probably borrowed'. Some modern researchers favour Steinthal, positing a 'Quechumaran' grouping to include both Quechua and Aymara; the majority, probably, agree with Mason.

At any rate, there certainly has been cultural influence between the two. A hundred years before the Spanish conquest, Aymara territory had become part of the Inca empire. The west Peruvian dialects of Quechua show strong Aymara influence, as if Aymara had once been spoken there.

The Aymara language has a traditional form of picture writing, used until quite recently to produce versions of Christian religious texts. This seems to represent an early stage in the typical development of writing – an aid to the memory, used for fixed texts such as catechisms and the Lord's Prayer, in which the texts are at least half-remembered. In this picture writing the characters are not standardised or used in the same way in different places. There are often fewer signs than words: just enough to recollect to the user's mind what he needs to say. The majority of signs are pictures of people and things. Some others are symbolic, and the meaning of signs can be stretched by means of puns and homophones. Aymara in this traditional script was at first written on animal skins painted with plant or mineral pigments: later, paper was used.

In modern Bolivia, where the largest community of speakers is to be found, Aymara is now written in the Latin alphabet. The orthography, introduced in 1983, follows Spanish practice. Books and magazines are regularly published, notably by the Evangelical and Catholic churches.

Many Bolivians are trilingual in Aymara, Quechua and Spanish. Thus, besides its Quechua elements, Aymara has now many Spanish loanwords, though they are much altered to fit the sound pattern: *winus tiyas* for Spanish *buenos dias*, 'good day'; *wisiklita* for *bicicleta*, 'bicycle'. The first ten numerals in Aymara are: *maya, paya, kimsa, pusi, phisqa, suxta, paqallqu, kimsaqallqu, llatunka, tunka*.
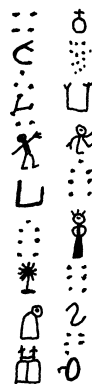
The Catholic sacraments in Aymara picture writing

AZERI

PERHAPS 14,000,000 SPEAKERS

Iran, Azerbaijan

One of the TURKIC LANGUAGES, Azeri is the national language of independent 'northern' Azerbaijan (once a republic of the Soviet Union) and is spoken in neighbouring districts of Georgia and Russian Dagestan. But the greatest number of speakers is in north-western Iran, in the province that will be called here 'southern' Azerbaijan (for map see TURKISH).

The place name *Azerbaijan* (local spelling *Azərbaycan*) derives from the Greek *Atropatene* – for this was once the kingdom of Atropates, a local Iranian ruler who established a dynasty on the south-western Caspian shores at Alexander the Great's death in 323 BC. *Azeri* has become the name of the Turkic language now spoken by the majority in the region. It is often called *Azerbaijani*, a term that misleadingly identifies the language with the political borders of Azerbaijan.

Speakers of a south-western Turkic dialect, the medieval ancestor of Turkmen, Azeri and Turkish, settled here between the 7th and 11th centuries AD. Some colonised the valleys and plains that border the Caspian Sea. Others turned to seasonal nomadism, their flocks grazing the Iranian and Caucasian highlands in summer and moving to the lowlands in winter.

Thus they range far beyond the borders of both halves of Azerbaijan. Apart from minorities in Georgia and elsewhere in Iran, there are 100,000 or more people of Azeri origin in Turkey, most of them belonging to the groups known as Karapapak and Terekeme, who migrated from northern Azerbaijan and Dagestan after 1828 and now live in Kars province. Their Karapapak language, originally a variety of Azeri, is gradually assimilating to the local dialect of Turkish.

Long disputed between Turks, Persians and Russians, Azerbaijan was divided between Russia and Persia (Iran) along the line of the River Araxes by the treaty of Turkmanchai, in 1828. In spite of an uprising against the Iran government in southern Azerbaijan in 1945, the border has held.

Briefly independent in 1918–20, northern Azerbaijan became independent again in 1991. Already Armenia's attempted annexation of the Armenian-speaking enclave of Nagorno-Karabakh, in the western mountains of Azerbaijan, has led to fierce fighting and eventually the expulsion of the hundreds of thousands of Azeris who formerly lived in Armenia. Meanwhile Azerbaijan has expelled the Lezghians from much of their former territory along its northern border, resettling this area with the refugee Azeris and with the Meskhetian Turks who have been driven out of Uzbekistan.

Some count the tales of *Dede Korkut* (see TURKISH) as the first classic of Azeri literature. In any case there has been an Azeri literary language, written in Arabic script, since the 14th century, and lively literary activity beginning in the 19th. Azeri oral epic poetry and romantic song is still a vigorous tradition – one that can be traced within Turkic cultures to the 11th-century performances of the *ozan* 'singer, storyteller'. Performances are now given by an *aşiq*, who improvises while playing on a stringed *kobuz*.

In Iran, Azeri education and publishing, outlawed under the Pahlavi monarchs (1925–79), revived in the 1980s: Arabic script is used. In the

1990s the Latin alphabet has become the standard in independent Azerbaijan. Azeri culture centres on Baku in the independent north and on Tabriz in the south, both cities of international importance.

Azari – confusingly – is the name of the Iranian language still spoken in parts of southern Azerbaijan (see map at GILAKI).

Information on local languages is hard to find in Iran. Recent estimates of Azeri speakers there (including the closely related Qāshqāy, spoken in the southern mountains) range from 500,000 (P. Oberling in the 1960s) through 1 million (N. Poppe, 1965) and 3 million (B. Comrie, 1981) to 8 million (*International Encyclopedia of Linguistics*, 1982/86, and other recent sources). In Iran, Azeri has no official status, but it is the major language of the north-west.

The Qāshqāy possibly number 100,000. They are traditionally seasonal nomads, keepers of horses, camels and cattle, whose lands are to the north-west and the south-east of Fars. They too had '*ashiqs*, who sang love poetry, battle songs and the epic tales of Köroghlu.

North of Azerbaijan lies Dagestan, a mountainous republic of Russia. In the scrambled terrain of Dagestan many peoples live and many languages are spoken. There had never been a single lingua franca for the region. High valley peoples learnt the languages of middle valley peoples; they in turn learnt lowland languages for trade and communication. 'Thus the Andi and Dido tribes used AVAR, while the Avars in turn used KUMYK or Nogai in their relations with the lowlands. In the same way, the Tabasarans in southern Daghestan used Lezgin [LEZGHIAN], and the Lezgins in turn used Azeri.'

In the 18th century classical ARABIC became the most-used lingua franca, and practically the only literary language, for Dagestan and neighbouring Chechnya. At the beginning of the 20th century the Young Turk movement, powerful here as well as in Turkey, urged the use of Azeri to replace Arabic. By the time of the Russian Revolution Azeri was in a strong position as lingua franca, especially in southern Dagestan.

Curiously, Soviet policy at first favoured Arabic as the second language of the region: it was seen as the choice of the 'masses' against the bourgeoisie. But this policy fell from favour, as it gave backing to traditional Islam and Sufism. Thus in 1923 Azeri was chosen as the preferred lingua franca, and the only school language, of Soviet Dagestan. Only five years later, policy changed again. There was now support for education and literacy in local languages. Dagestan (still staunchly Muslim) now has ten literary languages, and its usual lingua franca – naturally, since it is the language of government – is Russian.

Quotation from Alexandre Bennigsen, Chantal Lemercier-Quelquejay, 'Politics and linguistics in Daghestan' in *Sociolinguistic perspectives on Soviet national languages* ed. Isabelle T. Kreindler (Berlin: Mouton De Gruyter, 1985) pp. 125–42

Numerals in Turkmen, Azeri and Turkish			
	Turkmen	**Azeri**	**Turkish**
1	bir	bir	bir
2	iki	iki	iki
3	üç	üç	üç
4	dörd	dört	dört
5	beş	bäş	beş
6	altı	alty	altı
7	yeddi	yedi	yedi
8	səkkiz	sekiz	sekiz
9	doqquz	dokuz	dokuz
10	on	on	on

The Azeri alphabet

A B C Ç D E Ə F G H X İ I J K L M N O Ö P Q R S Ş T U Ü V Y Z

a b c ç d e ə f g h x i ı j k l m n o ö p q r s ş t u ü v y z

Writing in Azeri

Arabic script is still used by Azeris in Iran. In 1923 Soviet northern Azerbaijan adopted the Latin alphabet, which was called *Yanalif (yeni* 'new', *elifba* 'alphabet'). As elsewhere in the Soviet Union, the Cyrillic alphabet was imposed in 1939. In the 1990s the Latin alphabet has been reintroduced in independent Azerbaijan, in a form closely resembling Turkish, as shown in the box.

BAI

900,000 SPEAKERS

China

Bai is one of the SINO-TIBETAN LANGUAGES, or so most linguists believe. The older history of this language remains very mysterious. It is spoken in and around the city of Dali in Yunnan.

Speakers of *Bai* or *Pai* are called *Minchia* in older English writings, from local Chinese *min-jia* 'the common people'. The newly official term (from Chinese *bai* 'white') matches their own name for themselves, which is *pe-tsi* 'white people'.

Bai speakers, unlike most of the other minority peoples of Yunnan, are typically lowland dwellers and rice farmers, and they have adopted Chinese culture and Chinese beliefs quite comprehensively. Along with this, they have adopted Chinese words: most of the vocabulary of Bai is actually Chinese, some of it recent borrowing, some of it much earlier. Many Bai are bilingual in Chinese. Their cultural and economic links have been with Chinese, Tibetans and Burmese rather than with most of the hill peoples who surround them.

Can an older form of Bai be reconstructed? 'It would be difficult, because even those words which may be original Minchia vocabulary are difficult to interpret, as sound change seems to have proceeded radically to simplify the syllable structure of Minchia, thus eliminating contrasts that would have reflected proto-Tibeto-Burman and whichever subgroup of TB Minchia may belong to' (David Bradley).

It seems likely that the medieval kingdom of Nanchao, one of whose capitals was Dali, would have had Bai as one of its main languages. A Chinese text of the 9th century, *Man shu*, records several words in two Nanchao languages – but they do not coincide very much with Bai or with any of the other modern languages of Yunnan. Nanchao was conquered by Kublai Khan in 1253 and the Bai-speaking country has been part of China ever since.

Modern Chinese linguists regard Bai as a Burmese-Lolo language like YI. Scholars of Tai history sometimes take it to be a Tai language if they believe Nanchao to have been a Tai kingdom. The view was once popular that it was an Austroasiatic language of the Mon-Khmer group. In truth, no one knows the prehistory of Bai and no one at present can be certain with which language family to class it.

Bai has twenty consonants (a low number) and twelve vowels (a high number) including *v:* this is one of the few languages in the world in which *v* functions as a vowel. It has seven tones in open syllables. The first ten numerals are: *yi, go, sa, shi, ngur, fer, chi, bia, jiu, dser*.

South-east Asia: Bai and Naxi
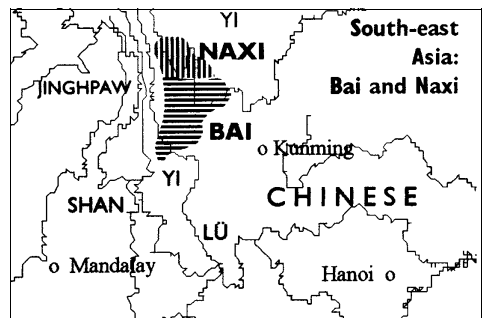
BALINESE

3,000,000 SPEAKERS

Indonesia

There was once a chain of Hindu and Mahayana Buddhist kingdoms stretching eastwards from India across south-east Asia and the Malay archipelago. Their traces remain – in stone temples, in statues of lions and gods, in languages that are full of Sanskrit loanwords, in names and forms of address that suddenly recall an Indian original.

Many centuries ago, most of these kingdoms went along other paths. Theravada Buddhism, Islam and Christianity have all won converts. Bali is one of the last outposts of Hinduism in Indonesia. Thus Balinese, one of the AUSTRONESIAN LANGUAGES, is at the same time one of a chain of cultural links, and a unique survival.

Balinese, the language of Bali in Indonesia, is linguistically closest to its two neighbours to the east, Sasak and Sumbawa (see map at SASAK). But Bali learnt its civilisation from the great island of Java, directly to the west. The earliest stone inscriptions on Bali, dating from 914, are in scripts indistinguishable from those used in Java. These ancient texts are in two languages – Sanskrit and (already!) Balinese. The language has a history of more than a thousand years.

Along with writing (see box) a whole literature, and a whole literary vocabulary, is shared between the two islands. Later Balinese inscriptions tend to be in Javanese, not Balinese. This interplay has its lucky side: if it were not for Bali's conservatism, the older Javanese literature would largely be lost and forgotten today. It was here that Javanese manuscripts were carefully preserved and recopied. It was through Balinese that western scholars, a century ago, discovered the old texts and learnt to read them.

Bali also learnt, and retained, a strictly hierarchical view of society, a view which is preserved more clearly in Balinese than in almost any other language. Lowland Balinese has very distinct 'formal' (*basa madia*) and 'informal' (*basa ketah*) registers, which speakers switch between as necessary, depending on their relation to the person addressed. Within the formal register, numerous detailed distinctions are made: *basa singgih* is the most elevated variety and consists largely of words borrowed from Javanese. The vocabularies of the three varieties are often quite different: 'eat' is *naar* in Basa ketah, *neda* in Basa madia, *ngadjengang* in Basa singgih.

Javanese influence fell away in the 16th century. More recently Balinese has been influenced by Malay, which has been the language of trade and travel in the archipelago for many hundreds of years. Through Malay, loanwords have come from Dutch, Portuguese and Chinese.

Balinese numerals are somewhat varied in form. With '2' and '3', different words are used in formal and informal registers. All the numerals have slightly different forms depending on whether they are used before or after a noun. The basic forms of the first ten numerals are: *sa, dua* or *kalih, talu* or *tiga, pat, lima, nam, pitu, kutus* or *ulu, sia, dasa. Dasa* '10' is a loanword from Sanskrit.

Kawi in Balinese

Written Balinese begins in the 10th century, with inscriptions recording royal decrees, some on stone, the majority on copper plates. The script looks as though it developed in writing on palm leaf, not on stone – but in tropical conditions palm leaf manuscripts do not last for a thousand years, so we may never know.

The Latin alphabet is now used for Balinese.

ha na ca ra ka da ta sa wa la ma qa ba nga pa ja ya nya

After C. Clyde Barber, *A Balinese-English dictionary* (Aberdeen: University Library, 1979)

Baltic languages

The Baltic languages (LATVIAN, LITHUANIAN and the extinct Old Prussian) are a compact group of INDO-EUROPEAN LANGUAGES, separate from but historically linked with their neighbours the Slavonic languages.

It is likely that as proto-Indo-European divided into dialects the linguistic ancestors of the Balts gradually moved northwards from an early location in what is now the western Ukraine. For a long period they must have remained in close touch with early Slavonic speakers. Eventually reaching the south-eastern Baltic shores, they settled in their present location without ever mingling with any large community of non-Indo-European speakers. Thus the Baltic languages – especially Lithuanian – seem closer than any other living tongues to the reconstructed proto-Indo-European of several thousand years ago.

Baltic languages on the ground

Latvian and Lithuanian descend from a hypothetical 'proto-East Baltic' language which will have begun to divide into dialects about AD 600 with a series of sound changes that have affected Latvian but not Lithuanian. They are the national languages of two republics both of which asserted their independence from the Soviet Union at the end of the 1980s.

Latvian has three main dialects. Livonian (*Tamnieks*) is spoken in the north-west; Central Latvian around Riga; Upper Latvian (*Augszemnieks*) in the east. Livonian is a confusing term as it is also used for an almost extinct Uralic language, related to Estonian, once the everyday language of north-western Latvia.

To the south-east, *Latgalian* is sometimes counted a separate language, midway between Latvian and Lithuanian, with a distinctive culture, Roman Catholic in religion.

Literary *Lithuanian* is based on the highland dialect of the *Aukstaiciai* of the south-east. The other major dialects are those of the *Zemaiciai* to the west and the *Suvalkieciai* south-west of the Nemunas river.

Two dialects of *Old Prussian* are known from a 14th-century glossary in Pomesan dialect and 16th-century catechism in Samland dialect. Soon after that the language must have become extinct, succumbing to the spread of German in East Prussia.

Numerals in proto-Indo-European and the Baltic languages

	Lithuanian	Latvian
1	vienas	viens
2	du	divi
3	trys	trīs
4	keturi	četri
5	penki	pieci
6	šeši	seši
7	septyni	septiņi
8	aštuoni	astoņi
9	devyni	deviņi
10	dešimt	desmit

Baltic languages

Baluchi

PERHAPS 4,000,000 SPEAKERS

Pakistan, Iran, Afghanistan

One of the IRANIAN LANGUAGES, Baluchi is a major language of south-west Pakistan, where it has over two million speakers, and can claim nearly a million speakers in south-east Iran.

There are about 200,000 Baluchi speakers in Afghanistan, where it is recognised as a 'national language'. (Quite distinct from them is the community of some thousands of Persian-speaking nomads in Afghanistan – called *Baluch* by themselves, *Jat* by others – who traditionally live by prostitution.) About 40,000 Baluchi speakers live around the Marw oasis in Turkmenistan: they migrated there from Afghanistan in the 19th century. Finally, there may be as many as half a million Baluchi-speaking migrant workers in Oman, Kuwait and the United Arab Emirates.

The name Baluchi or *Balōchī* is not found before the 10th century. It is believed that the language was brought to its present location in a series of migrations from northern Iran, near the Caspian shores: certainly it is more closely related to the north-west Iranian languages than to its neighbours.

For several centuries the speakers of Baluchi and BRAHUI have lived side by side. There are Brahui loanwords in Baluchi, and Brahui in turn is very heavily influenced by its neighbour. Baluchi naturally also contains loanwords from Persian – and from SINDHI, its eastern neighbour.

Written records of Baluchi date only from the 19th century. Quetta, in Pakistani Baluchistan, is now the centre of Baluchi culture and publishing. As with Sindhi and URDU, Baluchi is normally printed by lithography from handwritten copy in the *nasta'līq* form of Arabic script.

But quite apart from published prose and poetry there is a rich oral literature in Baluchi. Some of this has been collected in print – a major collection was made at Marw by the Soviet scholar I. I. Zarubin – but much remains to be done. Oral tradition preserves not only longer, anonymous tales and epics but also shorter lyrics by named poets of earlier centuries.

The first ten numerals in Baluchi are: *yak*, *dō*, *say*, *chīār*, *panch*, *shash*, *hapt*, *hasht*, *nuh*, *dah*.
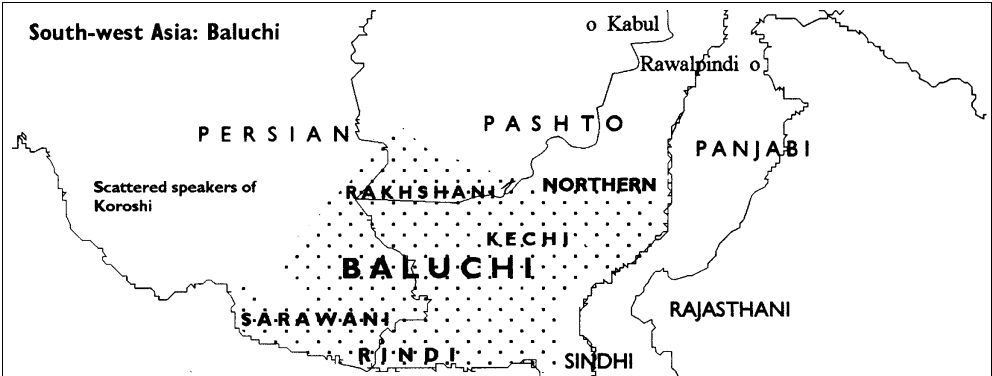
Dialects of Baluchi

Rākhshānī is the major dialect group in terms of numbers. *Sarhaddī*, a sub-dialect of Rākhshānī, is used by Radio Kabul. Other sub-dialects are *Kalātī*, *Chagai-Khārānī*, *Panjgūrī*.

The *Kechī* dialect of Pakistani Makran is heard in Pakistani radio broadcasts; *Sarawānī*, of south-eastern Iran, is used in radio broadcasts there. *Loṭūnī* is a third distinct dialect of Iran. Historically important is the *Coastal* or *Rindī* dialect, which extends along the Indian Ocean shore in Iran and Pakistan: this is the language of a good deal of classical Baluchi poetry.

Eastern Hill Baluchi or *Northern Baluchi* is very different from the rest. It was the first Baluchi dialect that became familiar to the British in India, and is thus particularly well documented in older linguistic work.

The camel drivers of the Turkic-speaking Qashqai, who live in the Fars province of Persia, are said to speak a language all of their own – *Koroshi*, most closely related to Baluchi.

Based on J. Elfenbein, *The Baluchi language: a dialectology with texts* (London: Royal Asiatic Society, 1966) and other sources

South-west Asia: Baluchi

o Kabul

Rawalpindi o

PERSIAN

Scattered speakers of Koroshi

PASHTO

PANJABI

RAKHSHANI · NORTHERN

KECHI

BALUCHI

SARAWANI

RAJASTHANI

RINDI · SINDHI

A riddle

Do gwahārāṇ dītha ambāzī,	I saw two sisters embracing,
ajab khush ant gwar-ambāzī,	Very happy at the embrace,
nēn-ī sūratā khamī,	Not the least difference in their looks,
yakē khor, dīgar chhamī.	But one is blind and the other can see.

The answer? A woman looking in a mirror.

M. Longworth Dames, *Popular poetry of the Baloches* (London: Royal Asiatic Society, 1907)

Banda

1,000,000 SPEAKERS

Central African Republic, Congo (Kinshasa)

One of the Ubangi group of NIGER-CONGO LAN-GUAGES, Banda is spoken across most of the central and eastern districts of the Central African Republic, by a mainly agricultural population.

Gbaya and Banda are the two major hinterland languages of the Central African Republic, now overshadowed linguistically by their relative, the national language, SANGO (see map there). Banda has probably been spoken in roughly its present location for as long as three thousand years, as the result of an expansion of early Ubangian languages from the west in the second millennium BC.

Its main dialects have thus gradually differentiated to the point where they are no longer mutually intelligible. But there is much travel and intermarriage among people of different dialects, so many in practice understand two or three. Most people are now bilingual in Sango; in towns, some men (but few women) speak French.

In Banda nouns a prefix marks the plural: *zu* 'man', *azu* 'men'. Banda is a three-tone language, and the tones are essential in word formation: *kāngà* 'to imprison', *kángà* 'prison', *kàngà* 'slave'. In Banda country, long distance messages are traditionally transmitted by three-tone drums, using formulaic phrases that match the tones of the spoken language.

Bantu languages

A very large group of related languages of central and southern Africa. Their interrelationship, as a language group comparable to Semitic or Germanic, was obvious to early explorers and was demonstrated linguistically by Wilhelm Bleek in 1857 and more fully by Carl Meinhof, whose comparative grammar appeared in 1899. Meinhof was the first to attempt a reconstruction of *Ur-Bantu*, proto-Bantu. Scholars had already argued that Bantu languages were related to the more diffuse family now called the NIGER-CONGO LANGUAGES. Diedrich Westermann proved this in 1927.

Within the huge and very old Niger-Congo family, Bantu belongs to the Benue-Congo group, whose other members are spoken in south-eastern Nigeria and Cameroun. It was clearly in this district, about four thousand years ago, that 'proto-Bantu' – the direct ancestor of all the Bantu languages of today – was spoken.

It was from this point that speakers of Bantu languages began to spread across the whole southern half of Africa. South central Africa is one of the very oldest areas of human habitation: Bantu speakers were newcomers at the end of a long history now difficult to trace, spreading new kinds of agriculture in the region. In archaeological terms they were at that early time a 'neolithic' group of peoples who made pottery but did not work metal. Later, after their migrations reached the Great Lakes region, the Eastern Bantu were almost certainly instrumental in the spread of metal-working technologies across Africa.

The Pygmy peoples of central and southern Africa, ethnically quite distinct, do not in general speak different languages from their neighbours. Thus *Aka*, a language of the south-western Central African Republic, is unusual: a Bantu language spoken by Pygmies alone. It has eleven noun classes, not unlike neighbouring Bantu languages, based on categories such as countable/not countable, long/round, diminutive/augmentative.

Most Bantu languages are tonal: in the majority, tones are at least partly predictable from the grammatical shape of verbs and nouns, but in others tones serve to differentiate words that are otherwise identical. A few, such as Swahili, have stress accents and no tones.

They are notable for simple syllable patterns, which often include prenasalised consonants (the *Mp-*, *Mb-*, *Nt-*, *Nd-*, *Nj-* familiar from personal and place names from the southern half of Africa). Some sound changes recur in many Bantu languages, such as 'Dahl's Law', in which *k-* is voiced if the next syllable begins with an unvoiced consonant: hence the local name *Gikuyu* for the language known to most outsiders as *Kikuyu*.

The numerals '1' to '5' and '10' in most modern Bantu languages are traceable directly to proto-Bantu. '6' to '9' are usually formed as compounds, '5 + 1' and so on; sometimes they are borrowed from other languages (see SWAHILI for an example). Probably proto-Bantu itself counted on a base of five, like the majority of the modern languages.

Bântu is the word for 'people' in numerous Bantu languages including LUBA (see box there): the *ba-* prefix is normal for the plural of a noun denoting human beings. 'Bantu' was first used as a name for the language family by Wilhelm Bleek in 1857. He applied it essentially to what we now know as the Niger-Congo family: later scholars narrowed the term.

At present the word *Bantu* tends to be avoided in South Africa because in the all-too-recent days of apartheid it carried political overtones. In that country, Bantu languages are therefore often called *Sintu languages*. This new word simply incorporates a different noun class prefix, the one that is appropriately used for a language. It appears in many Bantu languages as *si-*, *se-*, *ki-*: *siLozi*, *seTswana*, *kiSwahili*.

Bantu grammar for Bantu languages

Clement M. Doke, a linguist at the University of the Witwatersrand from 1923 to 1953, was perhaps the first scholar to see clearly that grammar in the Greek and Latin tradition, as adopted by most European linguists, was inadequate to describe Bantu languages. His thesis, *The phonetics of the Zulu language*, was presented in 1924 and his *Textbook of Zulu grammar* first appeared in 1927. This gave a new classification of the 'parts of speech' in a Bantu language, and a new terminology that has been gradually applied, by Doke and many others, to the description of Bantu languages. Doke himself worked on Khoisan languages as well as on Zulu, Shona, Southern Sotho, Ila and other Bantu languages. The Zulu grammar reached its sixth edition in 1961.

Bantu migrations

Proto-Bantu was perhaps spoken in or near the Grassfields of western Cameroun. The general region of origin can scarcely be disputed now that the membership of Bantu in the Bantoid group of the Benue-Congo branch of the Niger-Congo family has been demonstrated, for this is where the other Bantoid languages, siblings of proto-Bantu and as varied as might be expected after four thousand years of independent development, are still spoken.

The modern Grassfields languages, all spoken by quite small communities, are hard to classify but it has been suggested that they are the residue of the very earliest Bantu expansion.

The speakers of three or four early Bantu dialects had begun by 2000 BC to colonise the rainforest of Cameroun, Gabon, Congo (Brazzaville) and the Inner Basin of Congo (Kinshasa). One very early migration was by sea, to the Gabon estuary and the island of Bioko (where *Bubi* is still spoken: see map at EWONDO). The main movement, however, was probably a steady expansion south-eastwards from Cameroun and along major rivers. Living by fishing, by agricul-

ture and by river trade, and gradually adopting metal technology after about 300 BC, Bantu speakers established dominance over the remaining hunter-gatherer peoples of the rainforest, including the so-called 'pygmies', who now speak Bantu languages. The Bantu migrations into the rainforests were complete by about AD 300.

Meanwhile, perhaps as early as 500 BC, perhaps later, Eastern Bantu languages were spreading southwards from the region of the Great Lakes in modern Uganda, Tanzania and Kenya. They were eventually to reach the eastern half of what is now South Africa (the western half remaining the territory of speakers of KHOISAN LANGUAGES). They certainly interacted with, and perhaps mingled with, some of the Bantu languages of the rainforest, though the details here are much disputed. This migration reached the Atlantic along the coast of modern Angola. The Eastern Bantu expansion is usually linked with the spread of ironworking and of pastoralism across southern Africa, roughly two thousand years ago. A later archaeological break (the 'Later Iron Age' of southern Africa, with new pottery styles) at roughly AD 800 to 1000 may represent a second series of migrations.

What is not clear is the route by which early Eastern Bantu speakers migrated from the western plateaus to north-eastern Congo (Kinshasa) and the Great Lakes region in the first place. No solid archaeological evidence has been found for this migration, yet somehow it took place. Some think that it went almost due east, into the rainforest, and then up the Zaire and Uele rivers. Others believe that it was a gradual expansion of migrant cattle herders whose route kept to the north of the rainforest.

Linguists have made many sub-classifications of Bantu languages and their immediate relatives. One may see *Equatorial Bantu* (the rainforest groups), *Western Bantu* (sometimes the rainforest Bantu groups, sometimes others too), *Lacustrine* (including Luyia, Gusii, Nyankole, Ganda), the *East Highlands Group* (most of what is called Eastern Bantu above), *Narrow Bantu* or *Traditional Bantu* (excluding the Grassfields languages), the *Ungwa Group* (most Bantu lan-

guages and TIV), the *Bin Languages* and *Bane* and *Wide Bantu* (groupings of Bantu with Grassfields and some other Niger-Congo languages, separately known as *Bantoid* and *Semi-Bantu*). The continuing debate is highly important for the light that it may throw on African prehistory.

The Bantu spectrum

Detailed classification work on Bantu languages was done by Malcolm Guthrie. He assigned the five hundred languages and dialects of this group to lettered subdivisions A to S. Although superseded by later work, his classification is still sometimes used for reference so it is given for the languages listed here. A 'J' subdivision was introduced by Meeussen in 1953, incorporating some of the languages in Guthrie's subdivisions D and E; some other renumberings have been tried, but there is no standardisation about them, so it is Guthrie's own numbers (from his *Comparative Bantu* [Farnborough: Gregg, 1967–71]) that appear here.

The present list includes all Bantu languages with over a hundred thousand speakers – and gives cross-references to those which have a separate entry in this volume. More information on some of the smaller languages can be found through the index.

Guthrie was not in sympathy with modern comparative and historical linguistics. He believed that the Bantu language family, whose study he

BANTU LANGUAGES: THE SUBGROUPS
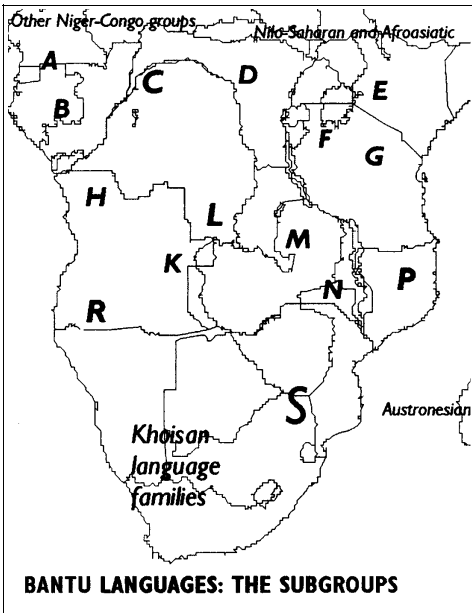
revolutionised with his work of reconstruction, was independent of other families and that modern Bantu languages represented various levels of decay from an original, fully logical proto-Bantu, spoken somewhere near the centre of the modern Bantu-speaking area. As an explanation of language change this approach was already outdated. Vestiges of it remain, however, not only in Guthrie's (indispensable) proto-Bantu reconstructions but in quite recent attempts by researchers to define what is 'a Bantu language'.

Asu	G22	315,000	Tanzania
Bangubangu	D27	120,000	Congo (Kinshasa)
Basaa (or Mbene) and			
Bakoko	A43	280,000	Cameroun
BEMBA	M42	1,850,000	Zambia and Congo (Kinshasa)
Bembe of Congo	D54	252,000	Congo (Kinshasa)
Bena	G63	490,000	Tanzania
Bera (or Bira), Komo and			
Nyali	D30	300,000	Congo (Kinshasa)
Bukusu	E31c	565,000	Kenya
Chagga or Chaga	E60	800,000	Tanzania
CHOKWE	K11	1,500,000*	Angola and Congo (Kinshasa)
Chopi or Lenge	S61	333,000	Mozambique
Chwabo or Cuabo	P34	665,000	Mozambique
Comorian	G44	450,000*	Comores

DUALA	A24	87,700	Cameroun
Embu	E52	242,000	Kenya
EWONDO, Bulu and Fang	A70	1,374,000	Cameroun, Equatorial Guinea and Gabon
Fuliru	D63	266,000	Congo (Kinshasa)
GANDA, Soga and Gwere	E10	3,542,000	Uganda
Gogo	G11	1,000,000	Tanzania
GUSII	E42	1,390,000	Kenya
Hehe	G62	630,000	Tanzania
Herero	R31	76,000	Namibia
Hunde	D51	200,000	Congo (Kinshasa)
Jita and Kwaya	E25	319,000	Tanzania
Kagulu	G12	217,000	Tanzania
Kalanga	S16	220,000	Botswana and Zimbabwe
KAMBA	E55	2,460,000	Kenya
Kami	G36	315,000	Tanzania
Kanyok	L32	200,000	Congo (Kinshasa)
Kaonde	L41	217,000	Congo (Kinshasa)
Kela	C75	180,000	Congo (Kinshasa)
Kele	C55	160,000	Congo (Kinshasa)
Kerebe	E24	100,000	Tanzania
KIKUYU	E51	4,360,000	Kenya
Komo	D23	150,000	Congo (Kinshasa)
KONGO	H16	4,720,000	Congo (Kinshasa), Angola and Congo (Brazzaville)
Konjo or Konzo	D41	250,000	Uganda
Kunda	N42	100,000	Zimbabwe
Kuria or Koria	E43	345,000	Kenya and Tanzania
Kwanyama	R21	150,000	Namibia
Lala and Bisa	M50	354,000	Zambia and Congo (Kinshasa)
Lamba and Seba	M50	170,000	Zambia and Congo (Kinshasa)
Langi	F33	275,000	Tanzania
Lega	D25	400,000	Congo (Kinshasa)
Lenje	M61	136,000	Zambia
LINGALA and Bangala	C36d	12,000,000 as a first or second language	Congo (Kinshasa), Congo (Brazzaville), Central African Republic
Logooli, Idakho, Isukha and Tirikhi	E41	503,000	Kenya and Uganda
LOMWE and Ngulu	P32	2,000,000	Mozambique and Malawi
LOZI	K21	450,000	Zambia
LUBA	L30	7,810,000	Congo (Kinshasa)
Luchazi	K13	125,000	Angola, Zambia
LUNDA	L52	550,000*	Congo (Kinshasa), Zambia, Angola
LUVALE	K14	600,000	Zambia, Angola and Congo (Kinshasa)
LUYIA, Nyore and Saamia	E30	3,734,000	Kenya
MAKONDE	P23	1,060,000	Tanzania and Mozambique
MAKUA	P30	3,540,000	Mozambique, Malawi and Tanzania

Mambwe-Lungu or Rungu	M15	307,000	Zambia and Tanzania
Masaba or Gisu	E31	500,000	Uganda
Matengo	N13	150,000	Tanzania
Mbala	H41	200,000	Congo (Kinshasa)
Mbola	D11	100,000	Congo (Kinshasa)
Mbunda	K15	102,000	Zambia and Angola
Benguela MBUNDU	R11	3,000,000	Angola
Luanda MBUNDU	H21	1,820,000	Angola
Mbwela	K17	100,000	Angola
MERU	E53	1,230,000	Kenya
Mijikenda	E72	988,300	Kenya, Tanzania
Mongo and Ngando	C60	216,000	Congo (Kinshasa)
Mpuono	B84	165,000	Congo (Kinshasa)
Mwanga	M22	223,000	Zambia
Mwera	P22	345,000	Tanzania
Nandi or Ndandi or Shu	D42	903,000	Congo (Kinshasa)
NDEBELE	S44	1,550,000*	Zimbabwe and South Africa
Ndengereko	P11	110,000	Tanzania
Ndonga	R22	240,000	Namibia and Angola
Ngando	C63	121,000	Congo (Kinshasa)
Ngindo	P14	220,000	Tanzania
Ngombe	C41	150,000	Congo (Kinshasa)
Ngoni	N12	205,000	Tanzania and Mozambique
Ngulu	G34	132,000	Tanzania
Nilyamba	F31	440,000	Tanzania
NKORE, Nyoro, Tooro, Kiga, Haya and Zinza	E10	4,668,000	Uganda, Tanzania and Congo (Kinshasa)
Nsenga	N41	250,000	Zambia
Ntomba	C35	100,000	Congo (Kinshasa)
Nyakyusa and Ngonde	M31	820,000	Tanzania and Malawi
Nyamwezi	F22	904,000	Tanzania
NYANJA or Chichewa	N30	4,000,000	Malawi, Zambia, Mozambique and Zimbabwe
Nyaturu or Remi	F32	490,000	Tanzania
Nyemba	K18	100,000	Angola
Nyiha	M23	306,000	Tanzania and Zambia
Nyungwe	N43	262,500	Mozambique
Pangwa	G64	185,000	Tanzania
Phende, Samba, Holu and Kwese	L10	492,000	Congo (Kinshasa)
Pogolo	G51	185,000	Tanzania
Rufiji	P12	200,000	Tanzania
Ruguru	G	506,000	Tanzania
RUNDI, Rwanda and Ha	D60	12,248,000	Burundi, Rwanda, Tanzania, Uganda and Congo (Kinshasa)
Safwa	M25	158,000	Tanzania
Sanga	L35	431,000	Congo (Kinshasa)
SENA	N44	1,200,000	Mozambique

Shambala	G23	485,000	Tanzania
Shi or Nyabungu	D53	654,000	Congo (Kinshasa)
SHONA	S10	7,950,000	Zimbabwe and Mozambique
Songe	L23	938,000	Congo (Kinshasa)
Northern and Southern			
SOTHO	S30	7,400,000*	South Africa and Lesotho
SUKUMA	F21	4,000,000	Tanzania
Sumbwa	F23	191,000	Tanzania
SWAHILI	G42	41,400,000 as a first or second language	Tanzania, Congo (Kinshasa), Kenya and Uganda
SWAZI	S43	1,600,000*	South Africa, Swaziland and Mozambique
Taabwa	M	250,000	Congo (Kinshasa)
Taita	E74	153,000	Kenya
Teke dialects	B70	267,800	Congo (Brazzaville)
Tetela	C71	750,000	Congo (Kinshasa)
Tharaka	E54	100,000	Kenya
Tonga of Malawi	N15	200,000	Malawi
TONGA of Zambia	M64	880,000	Zambia and Zimbabwe
Tonga or Shengwe of Mozambique	S62	225,000	Mozambique
TSONGA, Ronga and Tswa	S50	4,095,200	Mozambique and South Africa
TSWANA	S31	4,500,000*	South Africa and Botswana
TUMBUKA	N21	1,500,000	Malawi and Zambia
VENDA	S21	850,000*	South Africa and Zimbabwe
XHOSA	S41	6,900,000*	South Africa
Yaka	H31	150,000	Congo (Kinshasa) and Angola
YAO (AFRICA)	P21	1,160,000	Malawi, Tanzania and Mozambique
Zalamo	G33	450,000	Tanzania
Zigula	G31	336,000	Tanzania
ZULU	S42	8,800,000*	South Africa and Lesotho

Dialect shades into language imperceptibly across this wide region. Some languages often spoken of as distinct are grouped into single entries in the present list.

The above population figures are in general adapted from the fuller list of Bantu languages in the *International encyclopedia of linguistics* (New York: Oxford University Press, 1992) vol. 3, pp. 31–53. Those marked * have been revised on the basis of other recent sources.

Bantu noun classes

Very striking in practically every Bantu language is an elaborate system of noun classes, differing for singular and plural. Typically these classes are marked by noun prefixes, with matching prefixes for adjectives and for verbs to mark agreement with noun subjects. In this book there are sample tables of these noun classes, with some examples, at EWONDO, LUBA, MAKONDE and SOTHO. Twenty-one classes are reconstructed in one version of proto-Bantu. Whichever Bantu language they are working on, scholars still usually label these noun classes in accordance with Wilhelm Bleek's numbering system, worked out in 1851. The twenty-one classes are set out in the box, with three alternative reconstructions of their proto-Bantu prefixes.

Bantu noun classes

Conventional number and probable area of meaning	Meinhof's 'Ur-Bantu'	Guthrie's 'Common Bantu'	Later work
1: person	umu	mu	mo
2: people (plural of 1)	aβa	ba	ba
3: animate, agent	umu	mu	mo
4: plural of 3	imi	mi̧	me
5: singular of 6	ili	di	de
6: paired things, multiples	ama	ma	ma
7: custom, method, tool	iki	i̧	ke
8: plural of 7	iβi	bi̧	be
9: animal	ini		ne
10: plural of 9	ilini		dine
11: one of many things	ulu	du	do
12: plural of 13 and 19	utu	ka	to
13: small thing	aka	tu	ka
14: abstraction	uβu	bu	bo
15: verb infinitive	uku	ku	ko
16: 'on'	apa	pa	po
17: 'outside'	uku	ki	ko
18: 'in'	umu	mu	mo
19: diminutive	ipi	pi̧	pi
20: pejorative	uγu	gu	go
21: augmentative	iγi	gi	gi

Carl Meinhof, *Grundzüge einer vergleichende Grammatik der Bantusprachen* (Berlin: Reimer, 1906) pp. 1–27; P. Alexandre in *Les langues dans le monde ancien et moderne* ed. Jean Perrot, Parts 1–2 (Paris: CNRS, 1981) pp. 355–6; and other sources

BASHKIR

1,000,000 SPEAKERS

Russia

One of the TURKIC LANGUAGES, Bashkir is spoken by about a quarter of the population of Bashkortostan, one of the self-governing republics within Russia, in the southern Urals (see map at TATAR). It is now classed with Tatar in the Western or Kypchak group of Turkic languages, but this classification belies its complex history.

A people called *Bashqurt* were ruled by the Volga Bulgars (see CHUVASH) in the 10th century. Their present territory, after Bulgar, Mongolian and Tatar domination, came under Russian rule in the late 16th century. Its capital, Ufa, was founded as a Russian fort in 1574. Bashkirs are traditionally Muslim, and Ufa is now a centre of religion and religious education for Muslims in many parts of Russia.

Bashkir shows strong evidence of its history in the form of loanwords from Mongolian, from one of the Ob-Ugric languages (see HUNGARIAN) and from Russian. In spite of long term Tatar influence, both medieval and modern, Bashkir remains quite distinct from Tatar.

Russian, Tatar, Bashkir and Chuvash are used in education and the media in Bashkortostan. The Bashkir literary language and its local adaptation of the Cyrillic script are quite new developments, dating only from 1923 and 1940 respectively. In the late 19th century Tatar was used, alongside Russian, as a literary language in the region, and it still has that role. But there is now a growing tendency to adopt Russian as the everyday language: only two-thirds of those who called themselves Bashkir, in the last USSR census, claimed Bashkir as their first language.

Numerals in Bashkir and Tatar		
Bashkir		**Tatar**
ber	1	ber
ike	2	ike
ös	3	öch
dürt	4	dürt
bish	5	bish
altı	6	altı
ete	7	jide
higedh	8	sigez
tughıdh	9	tugız
un	10	un

Basque

PERHAPS 660,000 SPEAKERS

Spain, France

Spoken in France and Spain, on either side of the Pyrenees, Basque is the only living language of Europe that has no known linguistic relatives.

Basque was once thought of as belonging historically to the Iberian peninsula; many viewed it as a descendant of ancient Iberian, the language spoken in parts of Spain at the time of the Roman conquest in the 2nd century BC. But Iberian inscriptions and coins show it to have been a quite different language. The tribal name *Vascones* is actually first met with in Latin texts concerning not the south but the north – Gaulish or 'French' – side of the Pyrenees. *Vascones* is clearly the equivalent of modern *Basque*, but it is also the ancient form of modern *Gascogne*, 'Gascony', the region of France that extends well to the north and east of the modern Basque country, the same region that the Romans called *Aquitania*. From the few sources for the ancient Aquitanian language it seems certain that it is, in fact, the ancestor of Basque.

In Basque itself, the Basque language is called *Euskara* and the Basque country *Euskal herria*. This may possibly be linked with the tribal name *Ausci* of Latin times, which also survives as the name of the French town *Auch*. Other theories exist, however.

There is no doubt that Basque was once spoken over a wider area of France and Spain than it is now. But it was never the language of a major state. Even the kingdom of Navarre, which once straddled the Pyrenees in the Basque country, gave more prestige to Latin and to local forms of Romance than to Basque.

The first book printed in Basque – though with a Latin title – was *Linguae Vasconum primitiae*, 'Elements of the language of the Basques', by Bernard Etxepare (1545). Serious research on the language may be dated from that time. Publishing in Basque flourished in the 17th century. Basque and its culture have suffered from periodic recrudescences of nationalism in both France and Spain, which have resulted from time to time in the outlawing of Basque in official contexts, in schools, and even in all public places.

Basque now has about 660,000 fluent speakers: the estimate results from adding the number of speakers in Spain to the guessed number in France, where linguistic censuses are not taken. As many as 500,000 others know something of the language.

Souriquois: Basque pidgin of the Canadian coast

Basque whalers and cod fishermen regularly spent their summers around the mouth of the St Lawrence river from the 16th century onwards. The French explorer, Jacques Cartier,

Europe: Basque
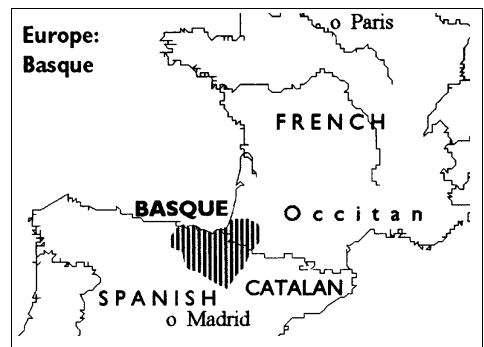
found in 1542 that the Amerindians of the St Lawrence shores could speak a kind of Basque. A similar report is made for Newfoundland by the historian Esteban de Garibay in 1571. More details of this pidgin language were recorded by later travellers. Having learnt it in contact with the early Basque mariners, the Indians naturally used it in speaking to the French and English explorers who came along later.

The name *Souriquois*, first reported in 1612, looks Indian. Far from it: it is a French spelling of Basque *zurikoa* 'language of the whites'.

Through these early voyages, which extended from Iceland across to New England, Basque pidgins may have had a significant influence on American English, as argued by J. L. Dillard: see his *A history of American English* (London: Longman, 1992) pp. 1–8.

The first ten numerals in Basque are: *bat*, *bi* (or *biga*), *hiru*, *lau*, *bost*, *sei*, *zazpi*, *zortzi*, *bederatzi*, *hamar*. A base of '20' is used to form higher numbers: *hogei* '20', *hogeitamar* '30', *berrogei* '40', *berrogei eta hamar* '50', *hirurogei* '60', *larogei* '80', *ehun* '100'.

There is no *f* in Basque. It is interesting that in a wide area around where Basque is now spoken, from the Garonne to Burgos, Latin *f*- has turned into *h*-. Does this mean that Basque was once spoken over all this area?

Souriquois on the St Lawrence river

A kind of jargon, *un certain baragouin*, assisted conversation between French and Algonquian language speakers in 17th-century Quebec. 'The Frenchmen who spoke it thought that it was good Indian, and the Indians thought that it was French,' said the missionary Paul le Jeune in a letter in 1632. L'Escarbot's *Histoire de la Nouvelle France* ('History of New France', 1612) explains the origin of this mixed language: 'The local people, to help us along, speak to us in a language with which we are more familiar, one in which a good deal of Basque is mixed. Not that they really want to speak our language – as they sometimes say, it was not *they* who came to look for *us* – but having been in such prolonged contact they are bound to remember some words.'

Batak languages

3,500,000 SPEAKERS

Indonesia

The Batak group of AUSTRONESIAN LANGUAGES is spoken in the highlands of north central Sumatra. The best known, with well over a million speakers, is Toba, usually called Toba Batak.

As the language of a longstanding independent written culture it has been much influenced by Sanskrit: see the section on 'The Batak script' for the word *si*, Sanskrit *śrī*, 'honoured', and the phrase *si raja*, 'honoured king, His Majesty', direct Sanskrit loans that are found in many languages of Indonesia. Other influences are Minangkabau, language of Sumatra's west coast,

and Malay, the lingua franca of the whole archipelago. Arabic, Portuguese and Dutch words are also to be found in Batak, many of them borrowed indirectly by way of Malay.

The Batak Lutheran Church, independent and locally organised since the 1930s, is the legacy of the work of the Rhenish Mission Society, begun in 1862. A translation of the Bible into Batak was completed in 1894, and schools spread rapidly.

The first ten numerals in Toba Batak are: *sada*, *dua*, *tolu*, *ɔpat*, *lima*, *ɔnɔm*, *pitu*, *ualu*, *sia*, *sappulu*.

The language of lamentation

At a death-bed and at a burial it is proper in Toba Batak to use a special form of language in chanted laments which narrate the life and achievements of the person mourned. Here are some examples of the traditional phraseology:

They are orphaned	Songon anak ni manuk na sioksiok i	They are like young chickens crying for help
I cannot speak	Songon sangge hinuntam, songon gansip na niodothon i	I am like a purse kept shut, like tongs squeezed tight
A scholar and speaker	Parjagajaga di bibir, parpustaha di tolonan i	One who is always watchful when speaking, and has the books in his throat
A headman or king	Sigongkonon bodari, sialopon manogot	The one who must be invited at night and met early in the morning

Examples from J. P. Sarumpaet, 'Linguistic varieties in Toba Batak' in *Papers from the Third International Conference on Austronesian Linguistics* ed. Amran Halim and others (Canberra: Australian National University, Research School of Pacific Studies, 1982) vol. 3, pp. 27–78

Toba Batak and its relatives

The Batak group consists of seven very similar languages or dialects of northern Sumatra: *Alas* or Kluet, *Karo Batak*, *Dairi* or Pakpak, *Simalungun* or Timur, *Toba Batak*, *Angkola*, *Mandailing*.

The Batak script

The Batak script is one of the offshoots of the ancient Brahmi alphabet once used for Sanskrit. Manuscripts (*pustaha*, from Sanskrit *pustaka* 'book') are written on bark. Texts may be in several of the dialects – Toba, Dairi, Angkola, Mandailing – and sometimes in Malay. Literacy in this traditional script is confined largely to *datu* 'priests', who use it for magical texts and diagrams, including calendars, which are important astrologically.

These texts are composed in a special language full of metaphor. A cooking pot is *si boru na birong panuatan*, *sitabo utauta*, 'her dark lady-

Insular South-east Asia:
Batak languages
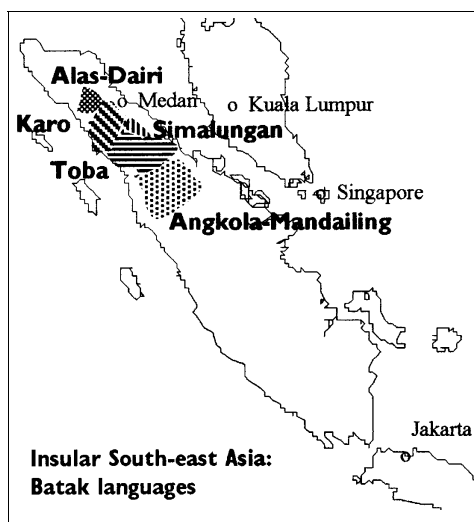
ship, source of things, whose emptying is tasty'. A lizard is *si raja ongkat di ruangruang, parbaju-baju bosi*, 'his majesty the king who lives in holes, who wears an iron coat'.

a ha ma na ra ta sa pa la ga ja da nga ba wa ya nya i u

BAULE

2,000,000 SPEAKERS OF BAULE AND ANYI

Ivory Coast and Ghana

Baule is one of the Kwa group of NIGER-CONGO LANGUAGES. Spoken by about one and a half million people of Ivory Coast, it is one of the two most important regional languages of the country (the other being Dyula, a variety of MANDEKAN). It was the mother tongue of the founding president of independent Ivory Coast, Félix Houphouët-Boigny (*ufue bwanyi*, literally 'white ram').

Until recently Baule was not a written language. It is now written in a 22-letter variety of the Latin alphabet, omitting *c h j q r x* but adding *ɛ ɔ*. As in French, an *n* following a vowel marks nasalisation.

Baule has twelve vowels, including five nasal vowels. All syllables end with a vowel. It has two basic tones, but modified in context, making for a fairly complex sound pattern. Two-tone drums are traditionally used to send messages between villages, mimicking the tone patterns of spoken phrases. Baule is more preponderantly monosyllabic even than its close relative Anyi: 'village' is *kulo* in Anyi, *klo* in Baule.

The first ten numerals in Baule are: *kun, nnyon, nsan, nnan, nnun, nsiɛn, nso, nmɔtyuɛ, ngwlan, blu*. Baule has borrowed heavily from French: *ekoli* 'school'; *loto* 'car' from French *l'auto* 'the car'; *amindi* 'noon' from French *à midi* 'at noon'.

Based on M. Malherbe,
Les langues de l'humanité (Paris:
Laffont, 1995) pp. 589–94 and other sources

Baule and Anyi on the map

Baule or Baoulé is a language of inland Ivory Coast, spoken in the regions surrounding Bouaké and Yamoussoukro and extending into western Ghana.

Anyi or Anyin is the eastern dialect of Baule, spoken on the left bank of the Comoé river. Anyi (under the name *Aowin*) and Sehwi (*Sefwi*) are the two Ghanaian dialects of Baule. There are perhaps 500,000 speakers of these dialects, including 200,000 in Ghana. Anyi includes the sub-dialects Nzema (or Nzima) and Ahanta.

Birthdays and names

In Baule, as in AKAN (see text there) and EWE, children are customarily named in accordance with the day on which they were born:

	Boys	Girls
Sunday	Kuame, Kuain	Amuin
Monday	Kuasi	Akisi
Tuesday	Kuadyo	Adyua
Wednesday	Konan	Amlan
Thursday	Kuaku	Au
Friday	Yao	Aya
Saturday	Kofi	Afue
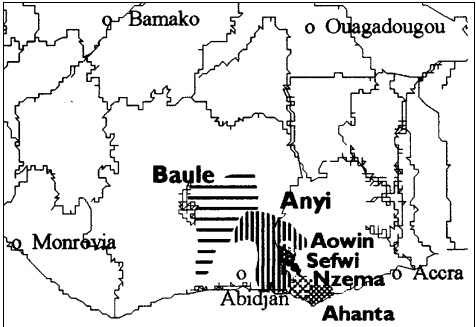
West Africa: Baule and Anyi

Beja

PERHAPS 500,000 SPEAKERS

Sudan, Eritrea

Beja is spoken by a nomadic people of whom rather little is known. They live between the Nile valley and the Red Sea, in the arid hills of north-eastern Sudan (see map at AFAR); one tribe, the *Beni Amir*, are to be found in northern Eritrea. Few details of their history are recorded, but it seems likely that Beja speakers have been in this region for at least two thousand years, and that they are the *Blemmyes* of ancient Greek geographical texts.

Beja is the Arabic name for this people and their language. Their own name is *ti-Bedaawye*. The English term *Bedouin* derives from it, by way of Arabic and French. The Bedouin of the Arabian peninsula are actually Arabic speakers, but many of those of the western shore of the Red Sea speak Beja.

In Eritrea speakers call their language *Hadareb*. The tribal names *Bishari* and *Hadendiwa* have also been used to denote Beja.

At the end of the 19th century two linguists worked on Beja: Herman Almqvist published a description of the Sudanese dialect, Leo Reinisch of the dialect of the Beni Amir. Few outsiders have studied the Beja or their language since then, and too little is known of the boundaries between dialects to place them reliably on a map. In Eritrea, however, two dialects are said to be spoken – one by the aristocracy, the other by the serf class.

Beja speakers in Eritrea, the Beni Amir, are apparently fairly recent migrants, who have intermarried with Tigre speakers and are now largely bilingual in Tigre. The *Beni Kanz* were a Beja-speaking group who went through a similar development several centuries ago. By agreement with the Nubian kingdom of Makouria, they settled on the Nile, between the First and Second Cataracts, at a point where nowadays not Beja but a NUBIAN language, *Kenzi*, is spoken. As they settled, they evidently adopted – probably by intermarriage – the language of Makouria whose northern sentinels they had become: this language was *Dongolawi* Nubian, of which Kenzi is a dialect.

Beja has five vowels, a distinction of vowel length, and (according to a recent analysis by Richard Hudson) a pattern of word stress which is predictable but highly complex in its realisation in actual speech. The first ten numerals are: *ngaal, mhaloo-b, mhay, fadig, áy, asagwir, asaramaa-b, asamháy, as's'adig, tamin*.

BELORUSSIAN

7,500,000 SPEAKERS

Belarus

Belorussian is one of the three Eastern SLAVO-NIC LANGUAGES, and has generally been over-shadowed by its neighbours (for map see RUSSIAN).

> *Belorussian* or *Byelorussian* means *White Russian*, and that name (or *White Ruthenian*) has sometimes been used for the language. It has nothing to do with the White Russian faction that attempted counter-revolution against Communism in 1918–22.

Belarus is divided from Ukraine by the vast Pripet Marshes. It seems likely that early Slavonic speakers slowly spread northwards from Ukraine to settle Belarus and Russia in the first millennium AD. The written Old Russian of Kiev (see UKRAINIAN) may be regarded as ancestor equally of Russian, Ukrainian and Belorussian. Modern Belorussian is part of a dialect continuum that links Ukrainian to the south and Russian to the east.

Little is known of the history of Belarus before it became part of the dominions of the pagan Lithuanians under Prince Gedymin in 1315. *Western Russian* or *Ruthenian*, an early form of Belorussian mixed with Old Slavonic, was the official language of Lithuania, which in due course became a Christian state and one half of the Polish-Lithuanian kingdom. This variety of Belorussian thus spread wider than any later form of the language, being spoken and written – in some contexts – all the way from the Baltic coast to Ukraine, where, gradually modified by Ukrainian, it was used administratively even in the 17th-century Cossack state in eastern Ukraine. It was often called *prostaya mova*, 'common tongue', to distinguish it from Church Slavonic.

However, Polish gradually supplanted Belorussian as the ruling language of Belarus and Belorussian peasants were increasingly subject to Polish landowners. At the end of the 18th century Russia annexed Lithuania, including Belorussia, and Russian became the new language of prestige.

A Belorussian translation of the Bible, by F. Skaryna, had been printed in Prague in 1517–19. There are important texts from the 16th century, including chronicles of Lithuania. The language and literature flourished in the 19th century, and even more with the establishment of an autonomous Belorussian Republic within the Soviet Union.

Belarus declared its independence in 1991 but remains on good terms with Russia, although Russia was blamed for the Chernobyl nuclear disaster which has left part of south-eastern Belarus uninhabitable. Most Belorussians speak Russian fluently, and some still regard their mother tongue as little more than a rustic dialect of Russian (as Russians themselves tended to do). The Belorussian press is more than half Russian in language, though the Russian minority forms less than a quarter of the population of the country.

The language differs from Russian not only by its characteristic sound pattern – there are examples in the table of numerals – but also because of the large number of Polish loanwords.

> The Belorussian version of the Cyrillic alphabet is easily recognisable. It uses I i (in place of Russian И и) for *i* and Ў ў for *w*. Until the early 20th century, under Polish influence, some Catholic Belorussians wrote their language in the Latin alphabet.

Не зыч ліха другому, каб не давялося самому

Nye zich lyikha drugomu, kab nye davyalosya samomu

Don't wish ill on another, lest it fall on you

Numerals in East Slavonic languages			
Belorussian	**Ukrainian**	**Russian**	**Russian in Cyrillic**
1 adzyin	odin	odin	один
2 dva	dva	dva	два
3 tri	tri	tri	три
4 chatiri	chotiri	chetıre	четыре
5 pyats'	pyat'	pyat'	пять
6 shests'	shist'	shest'	шесть
7 syem	sim	sem'	семь
8 vosyem	visim	vosem'	восемь
9 dzyevyats'	devyat'	devyat'	девять
10 dzyesyats'	desyat'	desyat'	десять

In this table all three languages are given in Latin transliteration: Russian also appears in the original Cyrillic script.

Bemba

1,850,000 SPEAKERS

Congo (Kinshasa), Zambia

One of the BANTU LANGUAGES, Bemba is the mother tongue of the largest single linguistic group of Zambia, where it is one of the eight official languages of the country and is spoken by many as a second or third language.

In its own region in north-eastern Zambia, Bemba is surrounded by related dialects. Its older history is uncertain: traditions speak of a migration from the west and of an origin among LUBA-speaking peoples. At any rate, the speakers of Bemba were, in the 19th century, the most powerful among their immediate neighbours. When the British South Africa Company took control here, in 1900, the Bemba king ruled the whole area between lakes Nyasa, Tanganyika, Mweru and Bangweulu, and even further to the east and the south-west.

The first ten numerals in Bemba are: *-mo, -bili, -tatu, -ne, -sano, mutanda, cine-lubali, cine-konse-konse, funai, ikumi*. English loanwords in Bemba include *shitoolo* 'store', *petulo* 'petrol'; from Portuguese *carreta* comes Bemba *iceleeta* 'cart'.

> The language is known to its own speakers as *chiBemba*; they refer to themselves as *ba-Bemba*. In older sources the form *Wemba* is found.

Languages of the Copperbelt

The earliest lingua franca of the Zambian Copperbelt was the pidginised ZULU known as Fanakalo. This was introduced at the end of the 19th century by the first European and Asian immigrants, who had learnt to use it in South Africa.

To the labourers in the mines, however, Fanakalo was both a foreign language and a colonial language: it was neither easy to learn nor socially acceptable. It continued to be used until the 1940s, by which time the ethnic profile of the mines had changed. Once, most workers had come from far away; now the majority of mine employees were speakers of Bemba and related languages. What was required was a language in which the minority of migrant workers could converse with this new majority.

Thus a simplified form of Bemba developed, known in English as *Town Bemba* (a term introduced by Irvine Richardson) or *Broken Bemba*; to its own speakers it is *ciKopabeeluti* 'Copperbelt language'.

Town Bemba is a good term, for this is a language of urbanism, a young people's language, the passport to the bustling life of Ndola, Mufulira, Chingola and neighbouring towns in the crowded Copperbelt. Here, speakers use it in work and leisure. Rural migrants learn it when they come to work in the city – and may cease to use it if they retire to their home villages. But with increased labour mobility it is now also spoken by significant minorities in Lusaka, Livingstone and Harare.

Town Bemba developed as a local lingua franca, easy to learn for speakers of a Bantu language (they take a few months to master it) but tending to exclude Europeans – and no less popular for that. It has superseded Fanakalo and was never seriously challenged by NYANJA (the police language). English, the fourth lingua franca of the Copperbelt, is used only by a minority.

Bemba and its relatives: the map

BEMBA has 1,850,000 speakers in Zambia and Congo (Kinshasa). *Taabwa*, with 250,000 speakers in Congo (Kinshasa), may be regarded as a northern dialect of Bemba.

Lala and Bisa, with 350,000 speakers, and Lamba with 175,000, are more distantly related to Bemba and are also spoken in Zambia and Congo (Kinshasa).

A third related group is formed by TONGA of Zambia, with 900,000 speakers in Zambia and Zimbabwe (divided into two main dialects, Plateau or Northern Tonga and Valley or Southern Tonga) and two smaller Zambian languages, *Lenje* (140,000 speakers) and *Ila* or *ciIla* (60,000). These are grouped as the 'three peoples', *Bantu Botatwe*.

Southern Africa
Bemba, Tonga and their relatives
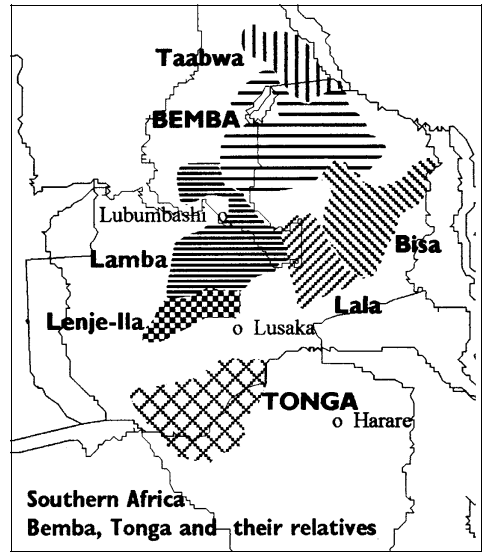
BENGALI

180,000,000 SPEAKERS

Bangladesh, India

Bengali, the best known of the Eastern INDO-ARYAN LANGUAGES, is the national language of Bangladesh (formerly East Pakistan) and is also spoken by the great majority in the Indian province of West Bengal: for map see ASSAMESE. Differences are beginning to emerge between the colloquial standard Bengali used in the press and media in the two capitals, Calcutta and Dacca.

The country at the mouth of the Ganges was already called *Vaṅgāla*, 'Bengal', in an 11th-century inscription.

The region was apparently occupied by Indo-Aryan speakers for the first time in the course of the later first millennium AD.

The Middle Bengali period is dated to the 14th to 18th centuries. Hinduism was by now the religion of Bengal – soon to share its predominance with Islam – and literature of the period is religiously inspired. The *Brajabuli* literary dialect, a mixture of Maithili and Bengali, was used conventionally by Vaishnava poets in a wide area of India, as if it were thought somehow specially appropriate to the Braj homeland of Krishna.

Among the classics of modern Bengali are the historical novels of Bankimchandra Chatterji, which are set in a period of Bengali cultural renaissance in the 15th century, just before the spread of Islam: his first novel was in English (*Rajmohan's wife*, 1864), but he then turned to Bengali. Modern Bengali has had two literary standards, *śadhu bhāṣā*, which looked back to 14th- and 15th-century litera-ture and to Sanskrit culture, and *colit bhāsā*, based on the modern colloquial of Calcutta. Rabindranath Tagore, the greatest modern Bengali author, preferred *colit bhāṣā*. A range of spoken and written styles, more and less formal, has developed from it, and *śadhu bhāṣā* is now little used.

Bengali, like the other Eastern Indo-Aryan languages, has no grammatical gender. The verb has three separate inflexions in the 2nd and 3rd persons ('you, he, she') to indicate the relative status of speaker and subject.

There are marked regional dialects, including some which may be considered separate languages (see map at ASSAMESE).

Sylheti has about 5,000,000 speakers in Sylhet District of Bangladesh, a hundred miles north-east of Dacca. There is a large community, perhaps as many as 100,000, of Sylheti speakers in Britain – they are usually called *Bengalis* or even *Pakistanis*. The biggest concentrations are in the boroughs of Tower Hamlets and Camden in London.

Bishnupriya Manipuri (or Mayang) is historically a form of Bengali once current in Manipur. Its speakers were driven from there in the early 19th century, and it is now spoken in Tripura State, in Cachar District of Assam, and in Sylhet District of Bangladesh. There are thought to be 150,000 speakers. The two dialects – once divided geographically, but no longer – are *Mādai gāng* and *Rājār gāng*, 'Queen's village' and 'King's village'.

Caryāpada

Around 1000 AD, when Bengali, Oriya and Assamese were not yet distinguishable as separate languages, the remarkable, mystical Buddhist *Caryāpada* songs were composed. They were discovered in a manuscript at Kathmandu and first published in 1916. They are claimed as the foundation of the literary tradition of all three languages.

Gaṅgā Jaūnā majheṃ re bahai nāī
tahiṃ buṛilā mataṅgī yoiā līle pāra karei.

Between Ganges and Jumna another river flows:
the outcaste woman easily ferries a sinking scholar
across it.

Bāhatu Ḍombi, bāha lo Ḍombi; bāṭata bhaila uchārā;
Sadguru pāapasāeṃ jāiba puṇi Jinaurā. . . .

Row on, woman, row on, woman! It darkened
as we went;
by the grace of the Guru's feet I shall go again
to the City of the Jinas . . .

Kabaḍī na lei boḍī na lei succhaṛe pāra karei;

She takes no cowries, she takes no coppers: she
ferries whomever she pleases;

jo rathe caṛilā bāhabā ṇa jāi kuleṃ kula buṛai.

But if one mounts and cannot drive, he is set
back on the bank.

Caryāpada 14

Numerals in Bengali and related languages

	Bengali	Oriya	Assamese	Bishnupriya Manipuri
১	ek	eka	ɛk	ā
২	dui	dui	dui	dū
৩	tin	tini	tini	tin
৪	cār	cārī	sāri	sāri
৫	pānc	pañca	pās	pāz
৬	chay	chaa	say	soy
৭	sāt	sāta	khāt	hād
৮	āṭ	āṭha	āth	āt
৯	nay	naa	na	nau
১০	daś	daśa	dah	dos

The Bengali alphabet

অ আ ই ঈ উ ঊ এ ঐ ও ঔ কখগঘঙ চছজঝঞ টঠডঢণ তথদধন পফবভম যৱল শষসহ

a ā i ī u ū e ai o au k kh g gh ṅ c ch j jh ñ ṭ ṭh ḍ ḍh ṇ t th d dh n p ph b bh m y r l ś ṣ s h

The Bengali alphabet is one of the local developments of India's early Brahmi. Like the others, it has numerous conjunct characters for doubled and adjacent consonants. When this alphabet is used for Sanskrit texts, the symbol ব serves for both *b* and *v*.

BERBER LANGUAGES

In the 18th and early 19th centuries linguists became interested in the distant relationships of well-known language families. The Semitic group, including Arabic and Hebrew, was found to have pervasive similarities with two language groups of north Africa: Egyptian and Berber. This was the beginning of the recognition of the 'Hamitic' or 'Hamito-Semitic' family, now most often known as the AFROASIATIC LANGUAGES.

Theories crop up from time to time linking Berber with the ancient Iberian language of Spain, with Basque, and with other linguistic survivals in Europe. They are generally short on evidence and probability.

Beraber is the name used for themselves by the speakers of TAMAZIGHT, one of the Berber languages of Morocco. The word has a long history: it was first used in northern Africa by Greeks or Romans, who referred to the local inhabitants vaguely as *barbari*, 'barbarians, foreign language speakers'.

Before the spread of Islam, and of the Arabic language that came with it, Berber languages were almost universally spoken across northern Africa and the Sahara from the Atlantic to the borders of Egypt: at least, this is what all surviving evidence suggests, for no other indigenous language group seems to be represented in early inscriptions or place names. Only limited inroads were made by Greek, the language of the founders of Cyrene and neighbouring cities in the 7th century BC; Punic, the Semitic language spoken by the Phoenician settlers at Carthage, near modern Tunis; and Latin, language of the Roman Empire. If a Romance language began to develop in Roman north Africa, it did not last long.

The mysterious *Guanche*, indigenous language of the Canary Islands, is generally thought to have been a Berber language, though perhaps long separated from the main group. It became extinct in the 16th century.

As the Islamic conquest swept across north Africa and onwards to Spain, Berber speakers were soon dominant in the armies, in government and culture. Through this channel Berber can be shown to have influenced modern Spanish and Portuguese, as well as the local Arabic of modern north Africa. Some of the greatest medieval Arabic scholars, such as Ibn Khaldūn and the traveller Ibn Baṭṭūṭa, were of Berber origin. Ibn Khaldūn, the great philosophical historian of the 14th century, wrote a history of the Berbers.

But, almost as quickly, Arabic, language of the prevailing culture and religion, became the lingua franca, and then the mother tongue, of an increasingly large proportion of the population.

Since that time, Arabic has been the language of government, of education and of all written culture in north Africa: knowledge of the Berber script soon died in most places. The survival of the Berber languages of the north, still spoken by millions in north Africa though faced with official lack of recognition and sometimes active discrimination, is a tribute to the tenacity of traditional culture. It also has something to do with the seclusion of women, which ironically is a legacy of invading Islam: for, even today, while many Berber-speaking men are bilingual in Arabic, women are not. With the increasing reach of modern media, this situation is likely to change, and the position of these historic languages is under more serious threat.

It can be legitimately argued that one should still speak today of 'the Berber language', not of 'Berber languages'. The dialects spoken from Mauritania to Egypt remain identifiably alike. But it does not at present seem likely that a single literary standard will emerge; on the contrary, regional standards are tending to crystallise, with or without official support. That is why the major varieties of Berber have separate entries in this book.

In the Sahara, TAMASHEQ is the vehicle of a still vigorous nomadic culture – and the ancient script

is still remembered. TAMASHEQ is a national language in Niger, Mali and Burkina Faso.

In all there are perhaps 12,000,000 speakers of Berber languages. Across all of north Africa, place names are largely of Berber origin. Many have the easily recognised double mark of feminine nouns, *t- -t* (Tamasheq *barar* 'son', *tabarart* 'daughter'). This characteristic shape is seen in the language names *Tamazight* and *Tashelhet* and in place names such as *Tamanrasset, Touggourt, Tanezrouft.*

Berber languages naturally have many loan-words from Arabic. Latin loanwords, from an earlier period of cultural influence, are also easy to find (see box at TASHELHET).

The Berber languages

Owing to the lack of official recognition of Berber languages in Morocco and Algeria, population figures are largely guesswork. The number of speakers in Libya and in the Siwa oasis of Egypt is also unknown (see map at TAMAZIGHT).

TAMASHEQ is the language of the *Tuareg* of the Sahara. It has perhaps 1,000,000 speakers in Niger, Mali, Burkina Faso, Algeria and Libya.

Chaoui or Shawia is the language of about 250,000 speakers in the Aurès mountains in south-eastern Algeria.

KABYLE has at least two million speakers (one recent estimate is 7,000,000) in north-eastern Algeria.

Mzab has about 80,000 speakers, centring on Ghardaia in the Mzab region of Algeria.

RIFIA or Northern Shilha has 1 to 2 million speakers near the northern coast of Morocco and in north-western Algeria.

TAMAZIGHT or Central Shilha is spoken by about 3,000,000 speakers, the *Beraber*, in north central Morocco.

TASHELHET or Southern Shilha has perhaps 3,000,000 speakers in southern Morocco, Algeria. The few speakers in Mauritania are a reminder that Berber speech was once widespread in that country, now supplanted by Hassaniya Arabic.

Siwa and *Zenaga*, at opposite extremities of the Sahara region, are the most isolated of today's Berber languages, each with a few thousand speakers.

Libyan and Berber scripts

Over a thousand stone inscriptions have been found in North Africa, dating from the 2nd century BC and into Roman times, in a script and language which is neither Punic nor Latin. Although they are mostly very short, consisting largely of names, it is clear that the language concerned (usually called 'Old Libyan') is a form of Berber.

The script is still used by TAMASHEQ speakers. The native name for it is *tifinagh*. Removing the *t*-prefix, this seems to derive from Latin *Punicae* 'Punic, Carthaginian'. The script shows Punic influence – and the Romans appear to have observed the fact.

· ⵙ ⵉ ⵏ ⵣ ⵣ # ⵜ ⵅ ⵥ ⵣ ⵞ ⵗ ⵏ ⵝ ⵏ ⵏ ⵔ ⵙ ⵏ][ⵙ ⵙ ⵙ ⵙ +

ʼ b ǧ d h w z ž ẓ ḫ ṭ y k l m n ǵ f q g r š t

Bhojpuri

PERHAPS 40,000,000 SPEAKERS

India, Trinidad, Guyana, Fiji, Mauritius, Suriname and other countries

One of the INDO-ARYAN LANGUAGES, Bhojpuri is spoken in the middle Ganges valley (for map see HINDI), both north and south of the historic city of Benares. Variant forms of Bhojpuri are the majority languages of Trinidad, Guyana and Fiji.

> The ruined city of Bhojpur, near Shāhābād, was once the capital of a powerful Rajput principality. It fell foul of the Emperor Akbar, and later of the British Raj, and disappeared from the scene after 1857. Even at its climax, Bhojpur never ruled all the country where *Bhojpurī* is spoken. The language has also been called *Purbī* or *Purbiyā*, the language 'in front' or 'to the east' of Delhi.

Bhojpuri is not one of the established literary languages of India. The poet Kabir (1399–1518) claimed to write *Banārsī-bolī*, the 'language of Benares', his native city, but in fact his writing shows features that belong to Braj (see HINDI). Then and later, education here has been traditionally in Sanskrit and Hindi. Speakers of Bhojpuri nowadays will generally say that they speak 'Hindi'. There are relatively few Sanskrit loanwords; rather more are drawn from Bengali and Hindi, which are spoken in the cities from which government has emanated, Calcutta (the old capital of British India) and Delhi.

However, Bhojpuri has a strong tradition of oral literature. The plays called *Bidesiyā* are a reminder of the long history of emigration from the region, for they tell the distress of a wife whose husband has travelled far away to find work.

For a table of numerals see MAITHILI.

> *Sadānī* or *Nāgpuriā* is sometimes considered a dialect of Bhojpuri, sometimes a separate language. It is the lingua franca of the eastern Chota Nagpur plateau, and is strongly influenced by the Austroasiatic language SANTALI that is still spoken there. There may be as many as 1,200,000 speakers. The name *Sadānī* means 'headquarters language', a reminder of how the language has gradually spread from local administrative centres in this jungle region.

Bhojpuri across the world

The Indians who were recruited to work in tropical plantations in the 19th century came, in their majority, from the middle Ganges region. There were speakers of Hindi dialects, of Bengali, of Maithili and Magahi, and of other Indian languages too. But in each major community a single language soon stabilised, to serve as a lingua franca among all Indian émigrés. Linguists who work on the languages of this diaspora today find that they have more of the special features of Bhojpuri than of any of its relatives.

The everyday name of these languages, however, is 'Hindi' or 'Hindustani'. If speakers are able to study their community language at school, standard Hindi is what they study. If they publish, standard Hindi is usually what they write. When their ancestors left India, Bhojpuri had hardly been identified as a separate language – and Hindi is still the Indian language of prestige, in literature, in education and in religion.

Mauritian Bhojpuri has about 300,000 speakers in Mauritius. Emigration took place between 1834 and 1900. Standard Hindi is used in school and in the media.

South African Bhojpuri is a threatened language, with fewer than 100,000 speakers: few children learn it. It contains loanwords from English (*afkaran* 'half a crown'), Fanakalo (*bagasha* 'visit') and Afrikaans.

Sarnami is the language of the Indian community of Surinam and the Netherlands, with about 180,000 speakers. There is some printed literature in Sarnami.

Guyanese Bhojpuri, locally best known as Hindi and Urdu, is the majority language of Guyana, with about 400,000 speakers.

Trinidad Bhojpuri or Hindi has been a major language of Trinidad, now in decline.

Fijian Hindustani (350,000 speakers, narrowly a majority of the population) seems to differ somewhat from the other Bhojpuri dialects of the diaspora, having more noticeable elements of 19th-century Bazaar Hindustani. Standard Hindi is sometimes used in Fiji in formal contexts: here English is tending to take its place. There is also a *Pidgin Hindustani*, commonly used in Fiji when Hindustani speakers converse with others who do not know the language.

The people down river

भागलपुर के भगोलिया ǀ	Bhāgalpur ke bhagoliyā	Bhagalpur are runaways,
कहलगाँव के ठग ǁ	Kahalgānv ke ṭhag	Kahalgaon are thugs,
पटना के देवलिया ǀ	Paṭanā ke devaliyā	Patna are sharpsters,
तीनू नमजद ǁ	tīnū namajad.	All three are known for it –
सुनि पावे भोजपुरिया ǀ	Suni pāvē Bhojpuriyā	If a Bhojpuri hears of them
त तीनू के तूरे रग ǁ	ta tīnū ke tūrē rag.	He'll break the heads of all three.

Maithili is spoken at Bhagalpur; at Patna Magahi is spoken. Kahalgaon is where the Bengali language area begins. The English word *thug* is a loanword from Indian languages.

After U. N. Tiwari, *The origin and development of Bhojpuri* (Calcutta: Asiatic Society, 1970)

BIKOL

3,000,000 SPEAKERS

Philippines

One of the AUSTRONESIAN LANGUAGES, Bikol is the major language of the Philippine provinces of southern Luzon – unofficially referred to as *Bicolandia*. It is closely related to Tagalog, the *de facto* national language, spoken to the north, and to the Bisayan dialects of the Visayas islands to the south (see maps at TAGALOG and CEBUANO).

The language name can be spelt *Bikol, Bicol* (the Spanish form) or *Vicol*.

A fragment of Bikol oral epic, *Handiong*, was transcribed by the Franciscan priest Bernardino Melendreras (1815–67) and published in Spanish in 1895. Sadly, the Bikol text does not survive. It told of the Flood and the rising of three volcanoes, Hantic, Colasi and Isarog.

Numerals in Tagalog and Bikol		
Tagalog		**Bikol**
isa	1	saro'
dalawa	2	duwa
tatlo	3	tulo
apat	4	apat
lima	5	lima
anim	6	anom
pito	7	pito
walo	8	walo
siyam	9	siyam
sampû	10	sampulo'

Bislama

PERHAPS 60,000 SPEAKERS OF BISLAMA AS A LINGUA FRANCA

Vanuatu

Like TOK PISIN of Papua New Guinea, Bislama is a modern form of *Beach-la-mar*, the English pidgin of the 19th-century western Pacific (see ENGLISH CREOLES AND PIDGINS). The influence of the old China Coast Pidgin can still be traced in the vocabulary and structure of the modern descendants of Beach-la-mar.

Bislama, the local pronunciation of *Beach-la-mar*, is now the preferred name; in French the modern form is *Bichelamar*. The word is said to derive from *bêche de mer*, the French term for the sea cucumber, which was gathered in the South Seas and marketed as a delicacy in China.

Bislama is the lingua franca of Vanuatu, formerly the New Hebrides. These islands were ruled as a 'condominium' by Britain and France between 1906 and 1980.

Originating in the mid 19th century as a language of labourers on European-owned plantations and of sailors on European-owned ships, Bislama became rapidly more important during the Second World War with increased population movement and the arrival of American military bases. In the 1960s New Hebrideans took part in local government, and Bislama was the available lingua franca. It is now an essential medium of communication among a population of 150,000 who speak 105 different local languages – though its English base makes it more accessible to the English-educated than to the French-educated Vanuatuans.

French and English remain official languages of independent Vanuatu, but Bislama now ranks as the national language. It is much used in government, press and broadcasting, as well as in trade.

The first ten numerals in Bislama, still close to English, are: *wan, tu, tri, fo, faev, sikis, seven, et, naen, ten*.

BODO-GARO LANGUAGES

PERHAPS 2,000,000 SPEAKERS

India

This group of SINO-TIBETAN LANGUAGES is spoken by hill peoples to the north and south of the Brahmaputra valley in Assam. The three major languages are Bodo, Garo and Tripuri. Two thousand years ago – before ASSAMESE arrived from the west and AHOM from the east – proto-Bodo was probably the language of the Assamese lowlands, the middle Brahmaputra valley.

Bodo or *Bāṛā* is the name for themselves of the Mec, Dimasa and Kachārī tribes. They are said long ago to have established an independent Buddhist kingdom: this was conquered by the AHOM of Assam. The Assamese language shows considerable Bodo influence; in turn, Assamese has influenced Bodo, particularly in its word structure.

Bodo peoples occupy what is in subcontinental terms a strategic location to the north of the Brahmaputra at the point where it narrows to a corridor linking Assam with the rest of India. They have taken full advantage of this location in agitating for political recognition. They won local autonomy in the early 1990s but a struggle for full independence continues, and many deaths resulted from the bombing of an Assam–Delhi train in December 1996.

Garo is spoken to the south of the Brahmaputra in the Garo Hills, at the western end of the Shillong plateau. Its speakers call themselves *Mande* 'man' or *A'chik* 'hillside': their neighbours to the east speak the quite unrelated Austroasiatic language KHASI.

'The Garo have long had market and headhunting relations with the surrounding plains people' (R. Burling in *Ethnic groups of mainland southeast Asia* ed. Frank M. LeBar and others (New Haven: HRAF Press, 1964) p. 56). Weekly markets are still a central feature of Garo culture. Garo speakers had no political units larger than the village, but they were generally independent until the British conquest in 1867: after that the Garo Hills formed a district within Assam, though with little outside interference, until the imposition of Assamese as official language led to demands for autonomy and the creation of the new state of Meghalaya in 1972. Garo and KHASI are its two main languages. As many as a third of Garo speakers are Christians, the majority of these being Methodists.

Speakers of Tripuri (or *Tipuri* or *Kok Borok*) are identified with the old state of Tripura, independent and sometimes powerful before the Mughal conquest in the 17th century. Tripura became a separate state of India in 1972. The majority of its inhabitants speak Bengali, a language of culture which was gradually spreading in Tripura under Mughal and British rule.

Bodo, Garo and Tripuri on the map

Bodo has 1,000,000 speakers in the Bodo hills of north-western Assam.

Garo has 500,000 speakers in the Garo Hills. *Atong* or Koch, to the south-east, with 50,000 speakers, is not mutually intelligible with Garo.

Tripuri (including the *Riang* dialect) is one of the three major languages of Tripura State, India, where it has 350,000 speakers. There are also 50,000 speakers in eastern Bangladesh.

Numerals in the Bodo-Garo languages

	Bodo	**Garo**	**Tripuri**
1	sè	sa	kai-sa
2	nè	gin-i	kun-nui
3	tham	git-tam	kā-thām
4	brè	bri	kai-brui
5	bā	bong-a	bā
6	ṛā	dok	dok
7	sni	sin-i	shini
8	zat	cet	cār
9	skhō	sku	cikuk
10	zi	ci-king	ci

Numerals are combined with classifiers: Robbins Burling, *A Garo grammar* (Poona: Schools of Linguistics, Deccan College, 1961) lists seventy-five of these. A few examples:

Classifier	**Range of meaning**	**Example**	**Example translated**
te-	hollow objects	me-dik te-sa	one rice-pot
dik-	potful	mi dik-sa	one pot of rice
kap-	cupful (English loanword)	ca kap-gin-i	two cups of tea
pak-	half of something	ku-mir-a pak-gin-i	two halves of an orange
cang-	times	cang-git-tam re-ang-a-ha	he went three times

South Asia:
Bodo, Garo and Tripuri
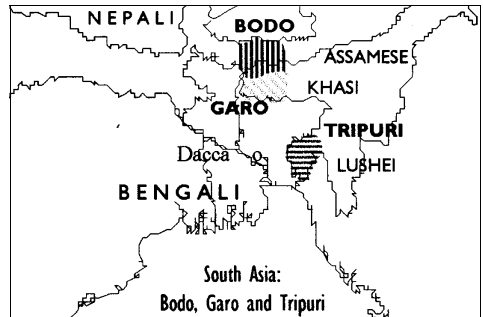
BRAHUI

1500,000 SPEAKERS

Pakistan

Brahui, a major language of western Pakistan, is an astonishing survival – utterly different from the Iranian languages that surround it. Most Brahui speakers are bilingual in Baluchi. There is no doubt that Brahui belongs to the family of DRAVIDIAN LANGUAGES of south India, though it is separated from the nearest of them by many hundreds of miles (see map at KURUKH). The relationship was first pointed out by the German linguist Christian Lassen in 1844.

One common guess is that early Dravidian speakers migrated southwards into the Indian peninsula thousands of years ago, and that Brahui (and perhaps Elamite) remain as a clue to show the route they followed. But it cannot be as simple as this: Brahui shares some innovations with northern Dravidian languages such as KURUKH so they must have separated from the other Dravidian languages before they themselves began to differentiate. In reality the early history of Brahui is quite unknown.

It is at least certain that the speakers of Brahui and BALUCHI have long lived side by side. Tradition tells of Hindu rule in medieval Kalat, the centre of Brahui and Baluchi culture, followed by a Brahui dynasty, followed by Mughal rule as elsewhere in India, followed by renewed Brahui domination. This last period contains a date – 1660, the accession of Mīr Ahmad as Khan of Kalat, ruling both Brahui and Baluchi tribes.

The pastoralist Brahuis traditionally migrate in large numbers in winter, the people called Sarāwān to Kacchī, the people called Jhalawān to Sind, where they have hereditary winter quarters. It has been suggested that their nomadic routes used to be westwards towards Afghanistan in earlier centuries, and that this explains why there are some 20,000 Brahui speakers now settled in Afghanistan, with a few further north in Tajikistan. At all events, Brahui is heavily influenced by Baluchi and by Sindhi, languages in which many Brahui speakers are necessarily bilingual. Although its Dravidian descent is still obvious, Brahui now has rather few inherited Dravidian words in its lexicon: Iranian, and specifically Baluchi, words predominate. Its sound pattern is most like that of Indo-Aryan languages such as Sindhi, without the short *e* and *o* and the multiple *r* sounds typical of Dravidian.

The first ten numerals in Brahui are: *asiṭ, iraṭ, musiṭ, chār, panch, shash, haft, hasht, noh, dah*. Only '1' to '3' are Dravidian – the higher numbers are borrowed from Baluchi or Persian.

O hilārki daun e ki ginjishk tūtaki – 'he's as fond of dates as a sparrow is of mulberries'.

BRETON

500,000 SPEAKERS

France

Celtic languages were once spoken all across the country that we now call France. They gave way before Latin, the language of education and government in the western Roman Empire. But as the empire faded and Saxon and other raiders attacked its northern provinces, Celtic speakers from Britain migrated in great numbers to northern France to fight and to settle.

These migrants were naturally called *Britanni*, 'Britons'. The great majority came to live in the north-western peninsula that was known to the Romans as Armorica – and was now renamed *Brittany*, or 'Little Britain', after the new inhabitants. This is how the island from which they set out gained its fuller name of *Great Britain*.

Perhaps this migration brought Celtic speech back to a country where it had already died out. More likely, it reinforced a surviving Celtic speech community and provided a new standard language, closer than before to Cornish and Welsh (see map at CELTIC LANGUAGES).

There was close contact between Wales, Cornwall and Brittany in early medieval times. Their religious communities were interwoven: students and missionaries were ever crossing the English Channel. Gildas, the 6th-century author of *The Ruin of Britain* (see box at WELSH), is commemorated as a saint in Brittany.

Even while Brittany was an autonomous duchy, in medieval times, French gradually became important there as a language of culture, of the nobility, and of the church hierarchy. As Norman dukes and French kings eroded the eastern frontiers of Brittany, the area of Breton speech gradually receded. In recent centuries French has been imposed as essential throughout Brittany by the highly centralised and nationalistic educational and legal systems of France. All speakers of Breton are now bilingual: it is said that there are 1,200,000 of them, but fewer than half of these use it daily as their first language.

The early literature of Brittany is in Latin – a series of religious texts and local saints' lives which contribute to the history of western France, south-west England and Wales. The 'Breton *lais*' written in French in the 12th and 13th centuries, by Marie de France and others, may be versions of oral poetry in Breton: as tales of Arthurian heroes they parallel the prose of the Welsh 'Mabinogion' and they inspired the better-known romances of the *Table Ronde* that were soon to be written in French, English, German and other European languages.

The first surviving Breton texts are some scraps of song inserted by a lonely Breton monk of the 14th century in a Latin manuscript that he was copying, Vincent de Beauvais' *Speculum historiale*, 'Mirror of history'. One of these reads:

Mar ham guorant va karantit
da vont in nos o he kostit . . .

> If my lover promises me
> I may lie beside her tonight . . .

A 15th-century text, a conversation between King Arthur and the prophet Guinclaff, long believed lost, was rediscovered in 1924 in an attic near Morlaix by the Breton scholar Francis Gourvil. The first printed text, a passion play, appeared in 1530. Breton literature revived in

the 19th century (partly inspired by a collection of none too authentic folk poetry, *Barsaz Breiz*, published by Hersart de la Villemarqué in 1839) and flourishes today, though with no official encouragement.

Since its earliest speakers came from lowland Britain and from western Gaul, in both of which Latin had been widely spoken and Roman cul-

ture had long been in fashion, Breton has more Latin loanwords than any other Celtic language: *laer* 'thief', *koan* 'supper', *eured* 'wedding'. It is naturally full of French loanwords, some of which, to judge by their form, go back to a very early period: *fresk* 'cool', modern French *frais*, which was a late Latin borrowing from Germanic *frisk*; *brau* 'beautiful' from Old French *brave*.

Numerals in the Brythonic languages		
	Welsh	**Cornish**	**Breton**
1	un	onen, un	unan
2	dau, dwy	deu, dyw	daou, diou
3	tri, tair	try, tyr	tri, tair
4	pedwar, pedair	peswar, peder	pevar, péder
5	pump	pymp	pemp
6	chwech	whegh	c'hwec'h
7	saith	seyth	seizh
8	wyth	eth	eizh
9	naw	naw	nav
10	deg	dek	dek
20	dau ddeg, ugain	ugent	ugent

Welsh has two ways of counting from 11 to 20: the old-fashioned way, still heard in some dialects, makes 16 '1 on 15' and 18 'two 9s'. Here are the older forms: *un ar ddeg, deuddeg, tri ar ddeg, pedwar ar ddeg, pymtheg, un ar bymtheg, dau ar bymtheg, deunaw, pedwar ar bymtheg, ugain.* This can be compared with the 'North Country score' (see box at ENGLISH).

In Breton, counting from 21 to 99 goes not by tens but by scores, just as it does in French from 61 to 99. So 79 in Breton is *naontek-ha-tri-ugent* '19 on three 20s'; in French it is *soixante-dix-neuf* '60 + 19'.

Bugis

3,600,000 SPEAKERS

Indonesia

One of the AUSTRONESIAN LANGUAGES, Bugis or Buginese is spoken in the southern peninsula of the spider-like island of Sulawesi (Celebes).

About three million Bugis speakers live in their historical homeland: many are rice farmers in the well-watered lowlands. They are also well known as seamen, and many speakers have settled in towns and cities elsewhere in the archipelago – other parts of Celebes, Borneo, Singapore, Java and Irian (New Guinea). The former principalities of Bugis have, at times, exerted power far beyond their linguistic boundaries: in the 18th century they dominated the seaways from the straits of Malacca eastwards.

Bugis, language of a Muslim culture, has loanwords from Arabic, Malay and the Dutch of the former colonial power. Like neighbouring MAKASAR (see script table there) Buginese has traditionally been written in *lontara* script, derived from the ancient Brahmi alphabet of India.

Many languages avoid the second person singular (older English 'thou') in polite address to strangers. The usual method is to substitute the second person plural (English *you*, French *vous*) or a third person form (German *Sie*, Italian *Lei*). Bugis uses the first person plural 'we': thus *idi'maning* means both 'we' and 'you [polite]', while *iko* and *iko'maning* are the singular and plural forms for 'you [familiar]'.

Bugis and Makasar on the map

These are the two major languages in a dialect chain covering much of south Sulawesi.

The principal dialects of *Bugis*, language of the former kingdoms of the centre of the peninsula, are Bone, Enna' or Sinjai, Camba, Soppeng,

Barru, Sidrap, Wajo, Luwu and – in an enclave among Mamuju speakers – Pasangkayu. The north-western dialect *Sawitto*, spoken at Pinrang, is distinct from the rest. In *Pangkep* Bugis and Makasar are closely mingled, but their speakers live in separate villages.

Makasar is the language of the once influential Sultanate of Gowa, whose capital was at Sungguminasa near modern Ujungpandang (still a major trading port, called 'Makassar' or 'Macassar' on most maps). Gowa is the dialect of the old capital and its district. Other dialects are Maros-Pangkep, Turatea and Bantaeng.

Konjo is spoken in the south-east corner of the peninsula. Coastal Konjo, with about 125,000 speakers, and Mountain Konjo, with 75,000, are the main dialects.

Selayar has 90,000 speakers and is the language of Selayar island, off the southern tip of Sulawesi.

Mandar, a language of 250,000 speakers in coastal districts around Majene in western Sulawesi, is not part of the Buginese-Makasarese dialect group but has, likewise, been written in Lontara script. Napo was once the capital of the
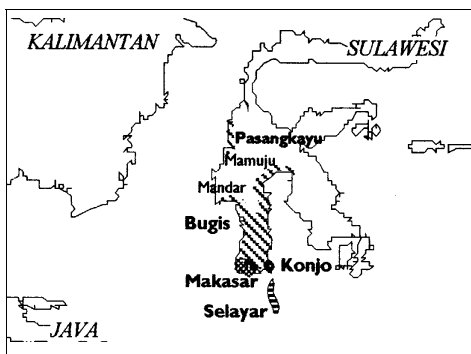
Mandar kingdom, and its dialect, Balanipa, is still regarded as the standard. *Mamuju*, with 95,000 speakers, is the neighbour of Mandar to the north.

Based on Charles E. Grimes, Barbara Dix Grimes, *Languages of south Sulawesi* (Canberra: Australian National University, 1987)

	Numerals in Bugis and neighbouring languages				
	Makasar	**Konjo**	**Selayar**	**Bugis**	**Mandar**
1	se're	se're	se're	seddi	mesa
2	rua	rua	rua	duwa	da'dua
3	tallu	tallu	tallu	tɪllu	tallu
4	appa'	'a'pa'	'appa'	'ɪppa'	'ape'
5	lima	lima	lima	lima	lima
6	annang	annang	anang	ɪnnɪng	'anang
7	tuju	tuju	tuju	pitu	pitu
8	sangantuju	karua	karua	aruwa	'aruwa
9	salapang	salapang	ka'assa	asera	amessa
10	sampulo	sampulo	sampulo	sɪppulo	sappulo

BULGARIAN

8,500,000 SPEAKERS

Bulgaria, Ukraine, Moldova

Bulgarian, one of the South SLAVONIC LAN-GUAGES, belongs to a dialect continuum that includes Macedonian, Serbian, Croatian and Slovene. It is the official language of Bulgaria.

> The people known as *Bulgars*, led by their king Asparuch, conquered the eastern Balkans in AD 680. Their Turkic language was soon forgotten (see CHUVASH), but their name survives in *Bulgaria*. Whether Slavonic speakers arrived alongside them, or at a different time, is not clear. Shortly afterwards, at any rate, the usual language here was a Slavonic dialect, and it was naturally called *Bulgarian* after the name of the country.

The Slavonic speech of Bulgaria and Macedonia (the two are hardly to be distinguished at this period) was first recorded in the 9th century in the form of OLD SLAVONIC or *Old Bulgarian*, in Christian texts that were circulated throughout the Slavic-speaking lands. *Middle Bulgarian* is the local language that emerges – in grammatical slips and spelling mistakes by scribes – in the later Old Slavonic texts of the 11th to 14th centuries. These are all in Cyrillic script, which has continued to be used for Bulgarian ever since. There is also plenty of evidence of Middle Bulgarian in later Old Slavonic texts from Wallachia and Moldavia (modern Romania), though Bulgarian was never the everyday language there.

Between the 15th and 19th centuries Bulgaria was under Turkish rule. Over this period the three great influences on the language were Turkish, Greek (as the language of Christian education) and Old Slavonic (still used in the Church). In particular, Bulgarian has been heavily influenced by Turkish. For five hundred years the local government, the elite and the army were Turkish-speaking. Through all this period Islam was spreading in Bulgaria, and Turkish was the everyday language of that religion. Many Turkish speakers settled in the country: some, in spite of recent discrimination, are still there. This, too, brought loanwords into Bulgarian.

Meanwhile there was regular interchange between Bulgaria and the nearest Christian lands to the north. In the 18th and 19th centuries Bulgarians emigrated in this direction in some numbers. There are now 275,000 speakers of Bulgarian in southern Ukraine, centring on Odessa, and 100,000 in Moldova.

> Pomaks trace their origin to Bulgarian/Macedonian speakers who converted to Islam in the 18th century. Some Pomak Muslims retain their distinct culture and dialect in Macedonia, Bulgaria and Greece. Others fled from discrimination in Europe and settled in Turkey, where the *Pomakça* dialect of Bulgarian is still spoken, though it is not likely to survive into future generations.

Modern Bulgarian texts, in a language already significantly different from that of the medieval manuscripts, begin to emerge in the 16th century. There was great interest in the language from the beginning of the 19th century, with grammatical studies by the Serbian Vuk Karadžić (1822) and the Bulgarian Neofit Rilski (1835).

Modern Bulgarian has several striking differences from other Slavonic languages. One is that nouns are no longer declined. As in Roma-

nian, however, there is a definite article suffixed to the noun: *grad* 'city', *gradăt* 'the city'; *gradove* 'cities', *gradovete* 'the cities'. For a table of numerals see SLOVENE.

The Cyrillic alphabet for Bulgarian

А Б В Г Д Е Ж З И Й К Л М Н О П Р С Т У Ф Х Ц Ч Ш Щ Ъ Ь Ю Я

а б в г е ж з и й к л м н о п р с т у ф х ц ч ш щ ъ ь ю

a b v g d e zh z i y k l m n o p r s t u f kh ts ch sh sht ă ' yu ya

The latest spelling reform, abolishing the letters Ѫ ѫ and Ѣ ѣ (see OLD SLAVONIC), took place in 1945.

BURMESE

21,000,000 SPEAKERS

Burma

One of a recognised Burmese-Lolo group of SINO-TIBETAN LANGUAGES, Burmese reached its present predominance in coastal and central Burma relatively recently.

Burma, Burman (now less used) and *Burmese* are English approximations to the colloquial Burmese name for the speakers, their country and their language, *Bamā*. The stress falls on the final long *ā*, hence the old-fashioned English rendering *Burmah*.

In Burmese formal speech and writing *Myanmā* is preferred to *Bamā. Myanmar*, the form at present encouraged in English and French, is a misspelling of the Burmese official name.

Burmese has been spoken in the middle Irrawaddy valley from at least the 9th century: its arrival there must have been the result of migrations from the north-east, where related languages are still spoken.

Anawrata, who ruled at Pagan from 1044 to 1077, established a lasting Burmese power one of whose effects was to spread the language westwards to Arakan (an independent kingdom until the 18th century) and southwards to the Irrawaddy delta region. Here it gradually replaced the mysterious Pyu language and eventually challenged MON. The relatively few remaining Mon speakers are now probably bilingual in Burmese.

But Burmese is pervasively influenced by Mon. Its script, its Buddhist religious terminology, its political vocabulary and its phonetics all demonstrate the long cultural supremacy of the Mon kingdoms of the Indian Ocean coast. Some Pali loanwords in Burmese have clearly been transmitted through Mon. At times Shan-speaking rulers held much of inland Burma, but the linguistic effect of this was relatively slight.

The Burmese cultural focus generally remained on the middle Irrawaddy – Pagan being eventually succeeded as capital by Ava, Amarapura and then Mandalay. Far beyond the area where it was a mother tongue, Burmese was a language of government and diplomacy in tributary states to west, east and south. The British, having annexed Arakan in 1826 and coastal 'Lower Burma' in 1852, retained Rangoon as their capital after conquering Upper Burma in 1885. Burma regained independence in 1947. This event has been followed by lengthy warfare with former tributary peoples, notably speakers of KAREN and SHAN. Burmese is now the official language of Burma, though primary education is sometimes in minority languages. It serves as national lingua franca – but only to the limits of government and army control.

The oldest written record of Burmese is the four-language Myazedi inscription at Pagan, dated to 1112. Royal inscriptions and the sacred texts of Buddhism were soon joined by a growing original literature, partly Buddhist in inspiration (there is a rich, often illustrated literature of *jataka* tales of the Buddha's former births), partly secular poetry and prose. At times Thai influence on Burmese literature has been powerful. Historical texts include the famous 'Glass Palace Chronicle', *Hmannan yazawindawgyi*, compiled from earlier sources in 1829–32. Printing in Burma began at the American Baptist Mission Press in 1816–17; it was also an American missionary, Adoniram Judson, who was responsible for the first major Burmese–English dictionary, published in 1852. The first non-Christian Burmese press began to issue a newspaper in Rangoon in 1868.

South-East Asia:
The Burmese-Lolo languages
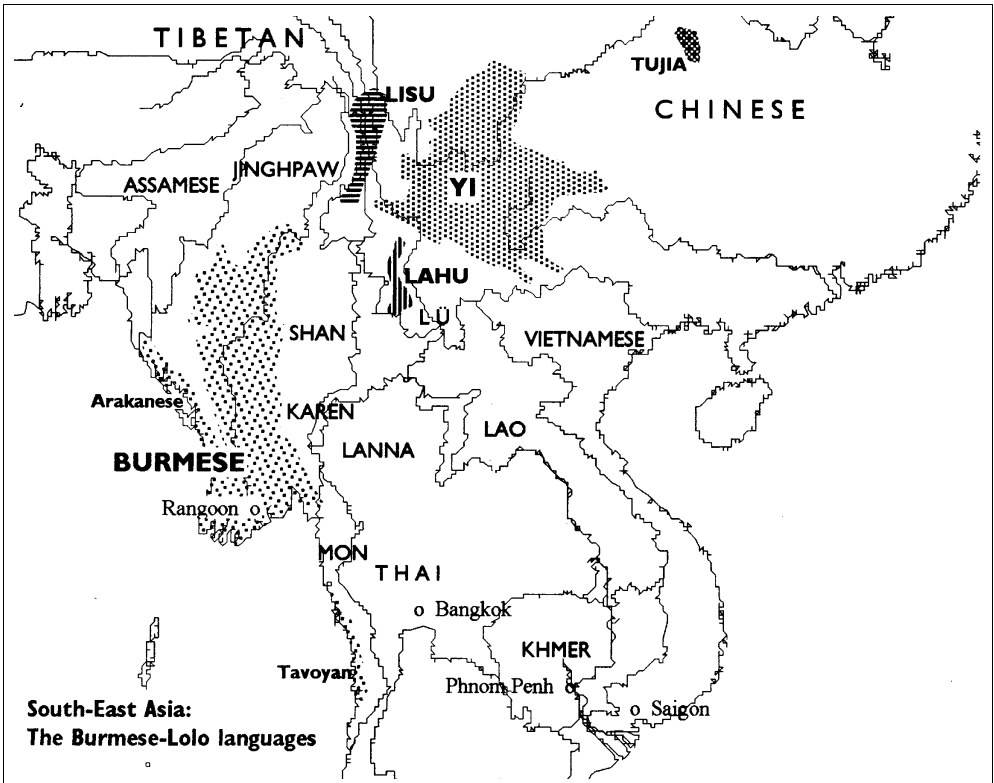
There is a considerable difference between colloquial or 'spoken' Burmese – to be found in the dialogue passages of novels and on television – and the formal literary language, 'written' Burmese, used in most printed books, journalism and radio news bulletins. The two styles are no longer kept apart so completely as they used to be.

Burmese is a tonal language, but in a completely different way from Chinese or Thai. It has a high and low tone; also a 'creaky' tone resembling the 'breathy' register of Mon and Khmer, and a 'checked' tone representing a short vowel followed by a glottal stop.

'Pali, the language of Theravada Buddhism, was known to the Burmese since before our first records of the language and Pali loan words are found from the earliest times. The obvious need for loans was in the field of religion (for example, words for *nirvana*, *karma*, monk, hell) . . . Influence from English was inevitable during British rule. Loan words are found in the predictable fields (car, telephone, radio, plug, com-mittee, cadre, coupon) . . . Loans from Mon are long-standing and are not nowadays generally perceived as loans by native speakers. They cover a wide range of fields, including flora and fauna, administration, textiles, foods, boats, crafts, architecture and music' (J. Okell in *South-East Asia languages and literatures: a select guide* ed. Patricia Herbert and Anthony Milner (Whiting Bay, Arran: Kiscadale, 1988) p. 5).

Burmese and the Burmese-Lolo languages: the map

Hani, Lahu, Lisu, Yi and others make up the 'Loloish' subgroup. Burmese and some minor hill languages make up the 'Burmish' subgroup.

HANI or *Akha* is spoken in south-western Yunnan, centring on Mojiang. There has been significant migration into mountain districts of Vietnam, Laos, Burma and Thailand.

The principal area of LAHU speech lies to the west of Akha territory, between the Salween and

the Mekong, in Yunnan and north-eastern Burma. There are Lahu villages in several south-east Asian countries.

LISU originates to the north of Lahu, extending along the upper Salween and Mekong valleys from Burma and western Yunnan into Sichuan. There has been extensive migration into southeast Asia and even into India.

Burmese, after southward migration and expansion in medieval times, has become the major language of the Irrawaddy plain. Apart from the main dialect, spoken with little variation in Rangoon, Mandalay, and the country between, some significant regional dialects are known. *Arakanese* retains the *r* sound which in standard Burmese has coalesced with *y*. This is why the British, who had learnt their Burmese in Arakan, called their capital *Rangoon*: its name in standard Burmese, now official in English too, is *Yangon*. A 19th-century migration has spread Arakanese across the Indian border to the Chittagong Hills: the dialect of the 125,000 speakers here is known as *Marma* or *Magh*. *Tavoyan* is a long-established southern dialect, historically separated from the rest by an area of Mon speech. *Intha*, *Taungyo* and *Danu* are minor dialects, influenced by Shan, spoken in the south-western Shan State.

There are several 'hill languages' more closely related to Burmese than to the Loloish group. One, *Maru*, has about 100,000 speakers in the eastern districts of Kachin State and in neighbouring parts of Yunnan.

Tujia, a little-known language now in steep decline, still has about 200,000 speakers in Hunan and Hubei provinces of China. It is thought to belong to the Burmese-Lolo group: if so, it lies far to the north-east of all its relatives. As many as 3,000,000 people of the area identify themselves as Tujia, but most now have Chinese as their mother tongue.

Poet in exile

မဲဇာ တောင်ခြေ

From Me'-za mountain's jungle foothills,

စီး တွေတွေရတည်၊ မြစ်ရေ ဝန်းလည် မြင်တောစည်က်

Washed by the circling river constantly,

ရွှေပြည်ကိုသင၊ တရှာတောမိ॥

My heart, against my will,

yearns for the Golden City . . .

Exiled to the jungle, far from the royal court, because of misbehaviour by one of his servants, Let-wè Thondará (1723–1800) wrote a series of poetic laments to his wife. The king was so moved by the verses that the poet was immediately recalled.

Hla Pe, 'Mind-bending Burmese poems and songs' in *Journal of the Burma Research Society* vol. 59 (1976) pp. 1–47

Burmese script
The thirty-three characters

ကခဂဃင စဆဇဈည ဋဌဍဎဏ တထဒဓန ပဖဗဘမ ယရလဝ သဟဠအ

k kh g gh ng c ch z zh ny ṭ ṭh ḍ ḍh ṇ t th d dh n p ph b bh m y y l w th h ḷ a

The twelve vowels shown with K

က ကာ ကိ ကီ ကု ကူ ကေ ကဲ ကော ကို ကား က်

ka kā ki kī ku kū kè ké kò kó ka: kā

Font: *Suu Kyi Burma* by Soe Pyne

Burmese numerals		
	Transliteration	**Actual pronunciation**
၁	tac	ti'
၂	hnac	hni'
၃	sum:	thō'
၄	le:	lè
၅	nga:	ngà
၆	hkrok	chau'
၇	hku' hnac	hkú hni'
၈	hrac	shi'
၉	kui:	kò
၁၀	hcay	she

Writing in Burmese

Burmese script is a development of the Mon alphabet, Indian in origin, and is perfectly adapted for writing with a stylus on palm leaves: rounded shapes are necessary because straight strokes with a stylus would split the leaf.

The thirty-three characters are shown in the box in dictionary order (but some dictionaries begin with the vowel character, အ, and some end with it). There are many more than twelve possible vowel-tone combinations with each character – but those shown are the twelve that schoolchildren chant as they learn to read and write, beginning with *k* and running through each character in turn. The same vowel combinations, with each character in turn, are used in numbering manuscript pages from 1 to 396.

Owing to its complex and multilingual history, the match between characters in this script and the sounds of modern Burmese is very different from the match between the ancestral script and the sounds of Sanskrit and Pali. This is why such very different transliterations into Latin letters will be found. Some scholars use the original values of the characters: others are led by the modern pronunciation of Burmese.

BUYI

2,000,000 SPEAKERS

China

Buyi is one of the most northerly and, to scholars outside China, one of the least known of the TAI LANGUAGES. It is spoken by a minority population in south-western Guizhou province.

Buyi (older transliteration *Pu-yi*) is the official Chinese name. The language has been given many names by linguists and ethnologists: *Jui, Dioi, Yoi, Yay* are attempts at reproducing the speakers' own name for it, while alternative Chinese names have come across as *Zhongjia* or *Chung-chia* or *Yi-jen* or *I-jen*.

In their own heartland the speakers of Buyi are scarcely distinguishable culturally from Chinese – and most of them are bilingual in Chinese. They tend to be lowland farmers, living in stone-built villages off the main roads; but many settle in towns, where they gradually become Chinese. In this region the hill farmers are generally speakers of MIAO: they are quite distinct culturally as well as linguistically.

Buyi is a typically Tai language, with six tones. It is similar to its southern neighbour, ZHUANG (see map there), which is spoken by a much larger minority group. There is no old-established written form of Buyi: those who learnt to write used Chinese, and a form of Chinese is used in local forms of spirit worship. Buyi love songs, traditionally important in courtship, often take the form of a question-and-answer dialogue.

Some hill peoples now living in Yunnan, northern Vietnam and Laos speak Tai dialects closely related to Buyi: they include *Nhang* (or Nyang or Giay or Yay) and the very archaic *Saek*, the recently discovered language of a small community in central Thailand. Linguists have occasionally worked on these displaced dialects of Buyi, but seldom on the language as spoken within China.

The first ten numerals in Buyi are: *nēu, lōng, lām, lī, hā, lɔk, shɛt, pēt, kū, ship*.

CATALAN

6,500,000 SPEAKERS

Spain

Like the other major ROMANCE LANGUAGES, Catalan grew out of the Latin of the Roman Empire, and it shows strong similarities both with Spanish to the west and with Occitan to the north. After centuries of decline Catalan has now emerged as the language of the autonomous region of Catalunya, established in 1979.

When the Kingdom of Aragon and the County of Catalonia were united in 1137 Catalan became the language of a major state, one with interests and connections far across the Mediterranean. The language remained important administratively until the 15th century, when the centre of power moved successively to Aragon and then to Castile, where Spanish was spoken. Since then the prestige of Catalan has depended on the success of movements for local autonomy, which first gained momentum in the late 19th century.

The language of lyric poetry in early medieval Catalonia was Provençal (Occitan). Catalan literature developed later, and is important for its prose, including the chronicles of Bernat Desclot and Ramon Muntaner and the political writings of Francesc Eiximenis.

Catalan and Occitan

Standard Catalan is based on the language of Barcelona, an *Eastern* dialect: others in this group are Roussillonais and Mallorquí. The dialect of Alghero in Sardinia also belongs here: Alghero has been a Catalan-speaking enclave since the 14th century. The major *Western* dialect is Valencian (see map at OCCITAN).

OCCITAN (Provençal) has no single widely accepted standard, in spite of the efforts of many activists. The numerous dialects range from *Provençal* and *Languedocien* in the south to *Limousin* in the neighbourhood of Limoges.

Quite distinct is Gascon, notable for its unusual sound changes, such as the replacement of Latin *f* by *h*, a feature shared with BASQUE.

A group of dialects transitional between French and Occitan, *Franco-Provençal*, is spoken in south-eastern France near the Swiss and Italian borders. Its territory covers the Italian autonomous district of Aosta, where there are about 70,000 speakers.

Numerals in the Iberian languages

Galician	Portuguese		Spanish	Catalan
un, unha	um, uma	1	uno, una	un, una
dous, duas	dois	2	dos	dos, dues
tres	tres	3	tres	tres
catro	quatro	4	quatro	quatre
cinco	cinco	5	cinco	cinc
seis	seis	6	seis	sis
sete	sete	7	siete	set
oito	oito	8	ocho	vuit
nove	nove	9	nueve	nou
dez	dez	10	diez	deu

Caucasian languages

The Caucasus mountains rise between the Black Sea and the Caspian. In geopolitical terms they help to divide the traditional spheres of influence of Russia to the north and Turkey to the south.

Like some other mountain ranges (the Hindu Kush, the Himalayas, New Guinea) the Caucasus has fostered the survival of a remarkable number of languages, most of them spoken by very small communities. Why? Because narrow mountain valleys hinder long distance communications, which help languages to spread; and because they provide excellent defences behind which relatively small communities can survive and prosper.

Since the maps in this book are at a standard scale, the map of the Caucasian languages is the most overcrowded of them all. In the Caucasus four completely independent language families are to be found – though evidence is accumulating that North Central and North East Caucasian are related, while some scholars, on evidence that is so far rather weak, add North Central and North West Caucasian to this grouping. Interspersed among these are also to be found members of other families well known elsewhere: Indo-European languages including Armenian, Russian and Tat (a variety of Persian), and Turkic languages including Azeri and Ossete.

If it is true, as many believe, that the ancient languages Hurrian and Urartian belong to the North East Caucasian family, the recorded history of Caucasian languages can be traced back over three thousand years. Otherwise, records begin with the invention of Armenian and Georgian scripts in the 4th century AD.

The mountain of tongues

'Caucasian languages' is a geographical grouping. It is customarily applied to the four independent

Caucasian languages

language families of the Caucasus, and not to the languages spoken there that belong to other well-known families. For completeness, some of the latter are marked on the map.

The *Kartvelian* or South Caucasian language family includes GEORGIAN (see map there), Svan (35,000 speakers in Georgia), MINGRELIAN and Laz.

For *Nakh* or North Central Caucasian see map at CHECHEN.

North West Caucasian includes Abkhaz, Abaza and the CIRCASSIAN languages (Adyge and Kabardian). For map see ABKHAZ.

The *North East Caucasian* or Eastern Caucasian or Dagestanian family includes the following languages, marked by numbers on the map and listed here with the number of speakers of each:

Agul	1	14,000
Akhvakh	2	7,000
Andi	3	12,000
Archi	4	1,000
Avar	5	500,000
Bagulal and Tlisi	6	7,000
Bezheta or Kapucha	7	2,000
Botlikh	8	2,000
Budukh	9	500
Chamalal dialects	10	6,000

Dargwa	11	300,000
Dido or Tsez	12	10,000
Dzhek or Gek or Kryz	13	1,000
Godoberi	14	2,000
Hinukh or Ginukh	15	200
Kaitak or Qaidaq	16	20,000
Karata	17	8,000
Khinalugh	18	2,000
Khunzib or Hunzal	19	1,000
Khvarshi	20	800
Kubachi or Ughbug	21	5,000
Lak	22	120,000
Lezghian	23	400,000
Rutul	24	20,000
Tabasaran and Khanag	25	90,000
Tindi	26	8,000
Tsakhur	27	12,000
Udi	28	5,000

For more on Dargwa, Kaitak, Kubachi and Lak, see DARGWA.

For more on the Avar-Andi-Dido group, see AVAR. It includes Andi, Botlikh, Godoberi, Akhvakh, Bagulal, Tlisi, the eight Chamalal dialects, Karata, Tindi, Dido, Khvarshi, Bezheta, Hinukh and Khunzib.

The remaining languages in the North East Caucasian family belong to the Lezghian or Samurian group. They have a particularly complicated political status.

LEZGHIAN itself is an official language in Russian Dagestan but not in Azerbaijan.

Russian has been imposed as literary language for speakers of *Agul* and *Rutul*. Agul is spoken by a fiercely independent and isolated group of clans, the Aguldere, Kurkhdere, Khushandere and Khpuikdere, on the upper Kurakh and Gyulgeri rivers. Rutul is spoken on the upper Samur, and Soviet policy had earlier been to assimilate its speakers to the Azerbaijani.

Tabasaran has separate official status in Dagestan. Its northern dialect, *Khanag*, is not mutually comprehensible with standard Tabasaran.

For any remaining speakers of the Shahdag languages *Budukh*, *Khinalugh* and *Dzhek* (which includes *Kryz*), Azeri is the official language. So it is for the *Tsakhur* speakers of Azerbaijan. For the 10,000 Dagestani speakers of Tsakhur, linguistic policy changed repeatedly. Tsakhur was briefly recognised as a literary language in the 1930s: Avar then took its place. Russian was finally imposed as their literary language in the late 1950s.

For *Archi*, Avar is the official language. Archi has a few hundred speakers only: it is the language of the single village of Ruch Archi in western Dagestan.

Udi, alone among North East Caucasian languages, is spoken by a people most of whom are Christians. Attached for thirteen hundred years to the Armenian church, this gradually shrinking community is settled around Vartashen and Nidzh in Azerbaijan.

CEBUANO

12,000,000 SPEAKERS

Philippines

One of the AUSTRONESIAN LANGUAGES of the Central Philippine group, Cebuano is the best known of the Bisayan languages.

These are the languages of *Visayas*, the group of mountainous islands, divided by narrow arms of the sea, lying between Luzon to the north and Mindanao to the south. The people call themselves *Bisayaq* and often call their language, collectively, *Binisayaq*, 'Visayan'. To eastern Muslims *bisayaq* came to mean 'slave', for the sultanates of the Malay archipelago raided these islands regularly. To its inhabitants now, the word signifies 'local, indigenous, native', with the implied warmth and pride of a phrase like 'home-grown' or 'home-made'. In Cebuano a local breed of chicken is *manuk bisayaq*; in Aklanon a local variety of rice is called simply *bisayaq*.

Cebuano, language of the island of *Cebu*, has also been called *Sebuano* (a simple spelling variant) and *Sugbuhanon*.

Bisayan language speakers have been fishermen and traders among the islands of the Philippine and Malay archipelagos since before the arrival of the Spaniards in the 16th century. Cebuano, Hiligaynon, Tausug and the rest, as vehicles of trade and culture, have influenced the languages of coastal and inland peoples of Mindanao, Palawan, Borneo, Mindoro, Luzon and many of the smaller islands on their trade routes. The effect of Bisayan oral literature has been traced in the *urukay* and *ambahan* songs of the Hanunoo speakers of Mindoro.

In its turn, Cebuano written literature developed under Spanish inspiration. A Christian manual of behaviour, *Lagda sa pagca maligdon sa tauong Bisaya*, published in 1734, is a landmark of Cebuano prose. Everyday language, too, is redolent of Spanish influence even to the greeting *kumusta*, Spanish *como esta*, 'How are you?' It has been calculated, by John U. Wolff, that a quarter of Cebuano vocabulary is of Spanish origin.

The Bisayan languages are certainly long-established in the Visayas region, but they have spread well beyond their original borders. The Cebuano of northern Mindanao, for example, shows the underlying influence of the various local languages from which the population has gradually shifted in order to adopt the more prestigious Cebuano. The same process can be seen at work still: smaller local Bisayan dialects, such as the *Porohanon* of the Camotes between Cebu and Leyte, gradually give way to the dominant Cebuano.

The effect of these long term influences is so pervasive, and the number of loanwords so great, that it is far from easy to ignore the later accretions and work out the original 'genetic' affiliations of the languages of the southern Philippines.

Based on David Zorc, *The Bisayan dialects of the Philippines: subgrouping and reconstruction* (Canberra: Australian National University, 1977) and other sources

Bisayan languages on the map

Seaborne trade is central to the culture of Visayas, and sea travel here is easier than land travel. This can be seen from the patchwork of languages and dialects, which tend to be

The Philippines: Bisayan languages
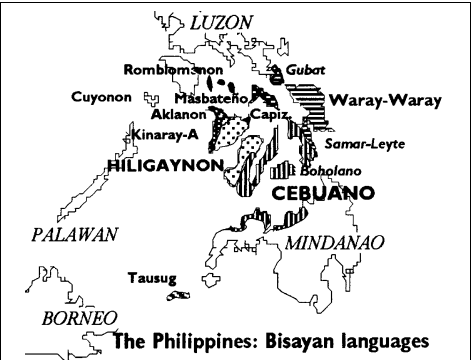
not divided but united by the arms of the sea that interpenetrate the archipelago.

Aklanon has about 350,000 speakers on the north-western shore of Panay.

Cebuano has spread more widely than the rest – in fact it has more native speakers than Tagalog, though this balance is likely to change. The mountain spine of Cebu itself faces south-eastern Siparay on one side, western Leyte and Bohol on the other (but some consider *Boholano* a separate language). Across the Mindanao Sea, much of the northern shore of Mindanao is also Cebuano-speaking.

Cuyonon, with 90,000 speakers, is the language of the Cuyo islands, also spoken on the coast of Palawan.

HILIGAYNON is spoken on eastern Panay and north-western Siparay. It is close to *Capiznon* of the north-eastern shore of Panay, and to *Masbateño*, the main mother tongue and lingua franca of Masbate island, which faces Panay and Siparay across the Visayan Sea.

Kinaray-a has about 300,000 speakers. It is the language of south-western Panay.

Romblomanon, with 200,000 speakers, is the language of the small island of Romblon and much of Tablas and Sibuyan islands on either side of it.

Tausug or Sulu, one of the two languages of Jolo and the Sulu islands, has a distinct history. The Muslim sultanate of Sulu was not conquered by Spain until 1878. There are nearly 500,000 speakers of Tausug as mother tongue, and perhaps as many again speak it as a second language. It is an important lingua franca of western and southern Palawan and the north-eastern coasts of Borneo.

WARAY-WARAY forms a dialect group including *Samar-Leyte*, *Northern Samar* and *Gubat*. These dialects are spoken on Samar, eastern Leyte and the south-eastern tip of Luzon.

Numerals in Cebuano and Tausug (Sulu)		
Cebuang		**Tansug**
'usa	1	'isa
duha	2	duwa
tulu	3	tuu
'upat	4	'upat
lima	5	lima
'unum	6	'unum
pitu	7	pitu
walu	8	walu
siyam	9	siyam
napulu'	10	hangpuu'

CELTIC LANGUAGES

The Celtic group of INDO-EUROPEAN LANGUAGES was once spoken across Ireland, Britain, Gaul (France), parts of Spain and Italy, much of southern Germany and the Danube valley. The Celts, farmers and brave warriors, were widely feared. Massive invasions had taken their language to Spain and northern Italy. There was even a Celtic-speaking enclave in classical Asia Minor: St Paul's Letter to the Galatians was addressed to the young Christian churches of this distant province. Place names across the Roman Empire testify to Celtic migrations: *Mediolanum*, Milan; *Singidunum*, the Latin name for Belgrade; *Laccobriga*, Lagos in southern Portugal.

Yet the four modern Celtic languages (BRETON, GAELIC, IRISH and WELSH) are now spoken only in scattered coastal and mountainous districts of north-western Europe – 'the Celtic fringe' in the jargon of centralist politics. All four survive only precariously. Languages of ancient and rich culture, they have been subject to centuries of social and political pressure from French and English. In the last hundred years two more Celtic languages, Cornish and Manx, have become extinct.

Manx

The Celtic tongue of the Isle of Man (*Ellen Vannin*) belongs to the group of *q*-Celtic languages like Irish: thus Manx *quig* (compare Welsh *pump*) 'five'.

Man was successively under Irish, Viking and Scottish domination, and was afterwards ruled by the Stanley family. The lordship of Man was sold to the English Crown in 1765. Under increasing English influence, Manx disappeared as a language of everyday life during the 19th and early 20th centuries: the last native speaker, Ned Maddrell, died in 1974. But it is still an official language: new laws must be promulgated in Manx and English.

Bishop John Phillips published the first Manx book, a prayer book, in 1611. He used English spelling rules as he devised a written form for Manx. The language showed strong English influence and also contained many Viking words: the Manx parliament, *Yn Kiare-as-feed* 'the Twenty-Four' (known in English as the House of Keys), meets annually at *Tynwald*, the same word as Icelandic *Thingvellir*, 'assembly ground'. The vocabulary of the English dialect of Man still includes Norse words: *clet* 'rock', *burrow* 'hill'.

Cornish: the once and future language?

The native speech of Cornwall is closely related to Welsh. It is closest of all, however, to Breton, the Celtic language of Brittany. Frequent contacts persisted, at least till the 16th century, between the priesthoods and the seagoing communities of these two peninsulas that face one another across the English Channel. Many place names are duplicated in Cornwall and Brittany, and many obscure local saints are commemorated in both.

The Celts of south-west England were geographically divided from those of Wales by the Saxon victory at Dyrham, near Bath, in 577. The West Saxons eventually advanced through Somerset and Devon. Cornwall itself was conquered in 936, but the Cornish language, without any official status, survived in daily use for nearly a thousand years after that date. From the 16th century come the *Ordinalia* and other religious plays in Cornish. But its survival was by now threatened: by 1600 it was said that nearly all Cornish speakers were bilingual in English. The last Cornish speaker who knew no English was said to be Dolly Pentreath: she died in 1777.

The very last native speaker of Cornish, John Davey of Zennor, died in 1891. Yet, in a modern revival movement, many Cornish children and adults now learn something of the language. Can it regain its position as a language of everyday life?
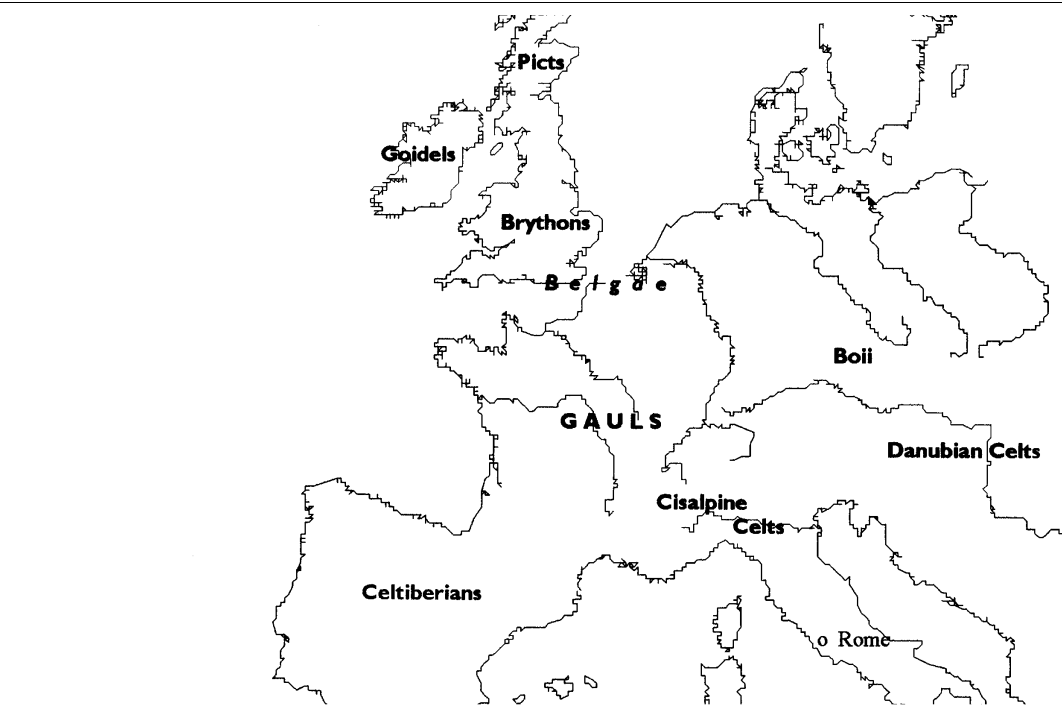
Twenty-three centuries of recorded Celtic

Dates AD	Goidelic	Continental	Brythonic	Dates AD
100–600	Primitive Irish	Gaulish	British	to 550
600–725	Archaic Irish	–	Primitive Welsh/Cornish/Breton	550–800
725–950	Old Irish	–	Old Welsh/Cornish/Breton	800–1150
950–1250	Middle Irish	–	Middle Welsh/Cornish/Breton	1150–1550
1250–	Modern Irish/Gaelic	–	Modern Welsh/Cornish/Breton	1550–

Brythonic and Continental Celtic are sometimes grouped together as *p*-Celtic, because Indo-European *q* became *p* in this group of dialects. Goidelic, more 'conservative', may be called *q*-Celtic.

For numerals in the Brythonic languages see table at BRETON; for numerals in the Goidelic languages see GAELIC.

Two worlds of Celtic languages

1st century BC

When the Roman Empire spread across Europe, two thousand years ago, three groupings of Celtic dialects can be distinguished – the evidence comes from place names, from the few words recorded in early inscriptions, and from reconstructions based on later texts. Scholars sometimes call *Goidelic*, or Gaelic, the language that was then spoken in Ireland (which remained entirely outside the Empire); *Brythonic* is the usual name for the Celtic of Britain; *Continental Celtic* is the general term for the dialects stretching from Spain to Galatia, of which the main group was *Gaulish*, spoken in what we now call France. *Pictish*, the language of the Scottish highlands in ancient times, may

Probable areas of Celtic speech in the 1st century BC

Galatians

have been a fourth subdivision of Celtic, but this is uncertain.

19th century AD
Continental Celtic disappeared completely: it was last spoken somewhere in northern France not long after the fall of the Roman Empire.

> *Cabmdhavas e metten, glaw yu etn* – 'a rainbow in the morning, rain is in it'. This rhymed Cornish proverb of about 1700, noted by William Gwavas of Paul, is close in rhythm to the English 'Red sky in the morning, shepherd's warning'.

Goidelic speakers invaded western Scotland in early medieval times, and their language replaced Pictish. Both Irish and Scottish Gaelic,

Europe:
Celtic languages

as well as the extinct Manx, are thus Goidelic languages.

In its heyday *Irish* had two main dialects, roughly corresponding to the legendary division of Ireland between *Leath Chuinn* and *Leath Mhogha*, Conn's half and Mogh's half. Others (such as Richard Stanyhurst, see box at IRISH) saw dialect boundaries corresponding to the four provinces, Ulster, Munster, Leinster and Connacht. Ulster Irish was continually influenced by the Scottish Gaelic of southward migrants. Unfortunately, very little of the Irish dialects was recorded before most of them died out.

In the 5th and 6th centuries Brythonic speakers migrated in large numbers from southern England to north-western France. This region, called *Brittany* after them, maintained close links with Cornwall for a long time. Welsh, Cornish and Breton are the three modern Brythonic languages. Welsh grew apart from Cornish and Breton, which are said to have been mutually intelligible. Cornish is no longer a mother tongue.

Celtic loanwords

Though they retreated before Latin, English and French, Celtic languages left their mark on the everyday vocabulary of all three.

camminus	road	French *chemin*, Italian *cammino*, Spanish *camino*
cerevisia	beer	Spanish *cerveza*
leuca	league [distance]	French *lieue*, Spanish *legua*, Portuguese and Occitan *legoa*, Italian *lega*

CHADIC LANGUAGES

The Chadic language group has around six hundred members, only one of which – Hausa – is at all well known. Half a dozen others are spoken by a hundred thousand people or more: most of the rest have quite small numbers of speakers. They are languages of Nigeria, Chad and Cameroun. The one Chadic language that is important on a world scale is HAUSA (see map there).

How do Chadic languages come to be spoken where they are? They form crucial evidence for the early history of one of the most important language families in the world: for (as was first observed by F.W. Newman in 1844) Chadic languages are related to the well-known Semitic group of languages. In fact Chadic is one of the six branches of the Hamito-Semitic or AFROASIA-TIC language family, whose other speakers are to be found in the Near East (e.g. Arabic, Hebrew), north Africa (e.g. Ancient Egyptian, Berber) and north-eastern Africa (e.g. Amharic, Somali). If it were not for Chadic, scholars might be able to argue that the Afroasiatic family originated in the Near East, spreading westwards and southwards. Some ignore Chadic and do argue this.

On this theory, the evidently ancient existence of Chadic languages at roughly their present location is almost impossible to explain: where would they first have differentiated from the other five branches, and why would their early speakers have migrated south-westwards across the Sahara? Chadic almost forces the adoption of a wholly different theory – the origin of Afroasiatic languages *in Africa* to the south and east of the Sahara. From there, the migrations that would produce Chadic, Berber, Egyptian and the Semitic languages are rather easier to explain.

In spite of differences resulting from some thousands of years of separate development, Chadic languages generally share some distinctive characteristics. They often have two classes of verbs, distinguished by the final vowel of the basic form, which may be –ə or –a: as in numerous other languages of the world, transitive verbs agree with the object, intransitive verbs with the subject. Verbs typically have distinct plural forms: singular *muri*, plural *mute* 'die'. Plural verb stems are often formed by replacing an internal vowel with -a- or by doubling an internal consonant. Intransitive verbs may be marked by a suffixed 'intransitive copy pronoun': *na ta-no* 'I went', *a ta-to* 'she went', *mə ta-mu* 'we went'. For the 'verbal extensions' reminiscent of English phrasal verbs, see HAUSA.

CHAM

230,000 SPEAKERS

Cambodia, Vietnam

Cham is unusual among AUSTRONESIAN LANGUAGES – for it is clearly long established on the Asian mainland. There, nearly two thousand years ago, it was the main language of the Hindu kingdom of Champa, founded, according to Chinese sources, in AD 192.

When this kingdom fell, as a result of a Vietnamese victory in 1471, many of its people migrated inland: this is the origin of the Cambodian speakers of Cham, now the largest group. Cham speakers are traditionally fishermen and traders along the waterways of Cambodia and Vietnam.

Islam had been adopted here in the 14th century and the speakers of Cham are still Muslims, though they have little contact with others of the religion. Malay is their language of religious education. A few still learn Arabic, but even fewer are able to make the pilgrimage to Mecca. Among the Vietnamese Cham, Muslim beliefs are mixed with Hinduism, and numerous gods have been recognised including *Po Haova* 'Eve' and *Po Adam* 'Adam'.

Proto-Chamic, linguistic ancestor of several scattered languages (see map), was influenced long ago by an early form of Khmer and perhaps other Austroasiatic languages. Sanskrit was the learned language of Champa, as of all the Hindu kingdoms of south-east Asia, and there are now many Sanskrit loanwords in Cham. More recently Malay, Arabic, Vietnamese, modern Khmer, French and English have all contributed loanwords.

Cham is one of the languages in which men's and women's speech differs most noticeably. Men, traditionally literate, use expressions from the older Cham literary language; women, traditionally not taught to read, speak in a 'modern' style.

Among the languages of the Malay Archipelago the closest linguistic relative of Cham is ACHEHNESE (see table of numerals there).

The Aceh-Chamic languages on the map

The languages related to Cham are widely scattered, and their early history is not fully known. Although Cham was the language of an early Hindu kingdom, the first four groups listed below are all now predominantly Muslim.

Western Cham is the modern form of the language as spoken in parts of Cambodia, along the Mekong and near Kompong Thom, by about 150,000 speakers.

Land of eagle-wood

Ves, corre a costa que Champá se chama,
  Cuja mata é do pao cheiroso ornada,

wrote Luis de Camões in his 16th-century epic of the Portuguese discoveries, 'see, here runs the shore called Champa, whose jungle is adorned with aromatic wood'. This is *Aquilaria agallocha*, aloes-wood or eagle-wood – an English term based on Portuguese *pao de aquila*, 'eagle-wood', which is borrowed from an Indian or south-east Asian language (compare Sanskrit *aguru*). When they go to the forest to gather eagle-wood, the Cham and Roglai are said to use a secret language, just as BATAK speakers do when collecting camphor.

South-East Asia: the Aceh-Chamic languages

Eastern Cham is the language of the Phanrang district of southern Vietnam, with about 80,000 speakers.

Huihui or Utset is spoken by a migrant group whose linguistic ancestors settled, long ago, in Hainan (China). There are now only about 3,000 speakers.

ACHEHNESE is the language of a former independent kingdom of northern Sumatra (Indonesia).

Besides these, there are further related languages spoken by mountain peoples – non-Muslims – of central Vietnam. They include *Chru* (15,000), *Haroi* (15,000), *Jarai* (200,000), *Rade* (120,000) and *Roglai* (45,000). Speakers of some of these languages were once subject to the old Cham kingdom: others, including Rade, were its enemies.

Cham script

The beautiful Cham script is one of the less-known offshoots of the Brahmi script once used for Sanskrit in India. It is known from inscriptions of the first millennium AD, in both Cham and Sanskrit, and some eastern Cham speakers still use it. Among speakers of western Cham, whose Islam has been closer to orthodoxy, Arabic script has been used. Under French rule the Latin alphabet was introduced for both Cham communities.

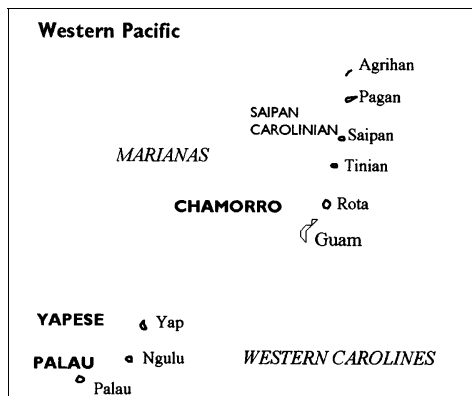k kh g gh ṅ c ch j jh ñ t th d dh n p ph b bh m y r l v ś s h a

CHAMORRO

72,500 SPEAKERS

Guam, Marianas

Unlike the languages of Micronesia, the three AUSTRONESIAN LANGUAGES of the Marianas and western Caroline islands are quite close linguistically to those of the Philippines – which lie fifteen hundred miles due west.

The islands came under Spanish influence in the 17th century. They passed successively to Germany, to Japan and then to United States suzerainty – Guam itself under direct US rule as a strategic air base, the other islands as a Trust Territory. The indigenous language of Guam, still a United States External Territory, is Chamorro, and it is still spoken by the majority of the inhabitants – but English is the island's only official language.

Chamorro shows strong influence both of Spanish and English. In vocabulary it is not very close to the Philippine languages, and its phonetics are unusual owing to a complex pattern of vowel harmony. All the modern Chamorro numerals are borrowed from Spanish: *un, dos, tres, kuatro, sinko, sais, siette, ocho, nuebi, dies*. Older sources, however, give native Chamorro forms: *hacha, hugua, tulo, futfat, lima, gunum, fiti, gualu, sigua, manot*. There are now so many Spanish loanwords in the vocabulary as a whole that some consider Chamorro a Spanish creole. However, Chamorro grammar remains very close to its Philippine relatives.

Western Pacific

Agrihan
Pagan
SAIPAN
CAROLINIAN Saipan
MARIANAS
Tinian
CHAMORRO Rota
Guam

YAPESE Yap
PALAU Ngulu WESTERN CAROLINES
Palau

Chamorro, Palau and Yapese: the map

Chamorro or Tjamoro has about 60,000 speakers in Guam and over 10,000 in the Northern Marianas.

Palau or Palauan (15,000 speakers) is the language of Belau or Palau, a group of islands in the western Carolines, and also has some speakers in Guam.

Yapese has 5,000 speakers on Yap, one of the western Carolines.

Migrants from the Caroline Islands settled on Saipan, in the northern Marianas, in the 19th century. A Micronesian language, *Saipan Carolinian*, is now spoken there alongside Chamorro. The Northern Marianas are now a United States 'Commonwealth Territory', with three official languages: Saipan Carolinian, Chamorro and English.

Chechen

900,000 SPEAKERS

Russia

Chechen belongs to the small family of Nakh languages (North Central CAUCASIAN LANGUAGES). It is the language of a people who have resisted Russian expansion in the Caucasus for over two hundred years.

Chechen speakers call themselves *Nwokhchi* (singular *Nwokhchuo*) and their language *Nwokhchīn mwott*. The name *Chechen*, used in Russian and internationally, comes from a village in the Chechen lowlands; similarly, Russian *Ingush* comes from the lowland Ingush village name *Angusht*.

Chechnya is a firmly Muslim country: Arabic is still important in education. Russia has fought a bitter struggle for the conquest of Chechnya, a conquest which was still far from assured after the hundred years' Caucasian War of 1760–1860 and was challenged by fierce rebellions in 1860–1, 1864, 1877–8, 1917–18 and 1940–4. In that year of mass deportations, many Chechen speakers were killed and the whole of the surviving population was dispatched to central Asia: there were then about 400,000, of whom a quarter died in the first five years of exile. The personal dislike of Stalin, a Georgian, for Georgia's Muslim neighbours contributed to the merciless treatment of Chechens and some other Caucasian peoples in 1944.

Chechen speakers returned to their traditional lands after 1957. The old Chechen-Ingush republic was re-established, still within Russia, and gradually enlarged. The powerful Sufi brotherhoods of Chechnya fomented a new independence struggle at the collapse of the Soviet Union. In the course of the still continuing warfare of the 1990s, Chechnya and Ingushetia were separated administratively in 1992.

Arabic used to be the only literary language of Chechen speakers. The first orthography for Chechen itself was devised in 1923, in the Latin alphabet. This was replaced with Cyrillic script, after a shift in Soviet language policy, in 1938. Even with the addition of the letter I (not used in Russian) the present script is short of vowel symbols for Chechen: changes, including the use of double letters for long vowels, are talked of.

In general the dialects of Chechen are mutually intelligible. Slightly more distinct, though still closely related, is the neighbouring language Ingush. Many speakers of Ingush know enough Chechen to be able to understand it when spoken: some Chechen speakers are equally competent in Ingush. Ingush, too, first became a literary language in the 1920s.

Ingush was the language of 90,000 deportees to Siberia in 1944, who suffered a similar death toll to that of the Chechen speakers. They gradually returned to the Caucasus from 1957 onwards, and now number over 200,000: but much of their territory had been occupied by Ossete speakers meanwhile. The Ingush capital was once right-bank Vladikavkaz, while left-bank Vladikavkaz was the capital of North Ossetia.

The Nakh languages on the map

Chechen is the principal language of Chechnya. Literary Chechen is based on the central lowland dialect. There were once speakers in Georgia, in the upper valleys of the Assa, Argun and Alazani

Numerals in Chechen, Ingush and Bats			
	Chechen	**Ingush**	**Bats**
1	tsa'	tsḥa'	tsḥa
2	shii	shi'	shi
3	qwoa	qo'	qo
4	vii	-i'	-'iū
5	p'khii	p'khi'	p'khi
6	yalkh	yɛlkh	yetkh
7	vworh	vuorh	vorl'
8	barh	barh	barl'
9	is	īs	is
10	it'	it	īt

The Caucasus: Nakh languages

rivers: few remain. Some Chechen speakers still live in Kazakhstan, to which they were exiled en masse in 1944.

Ingush has about 200,000 speakers in Ingushetia, which has now become a separate self-governing republic of Russia.

Bats (or Batsbi or Tsova-Tush) is the language of 2,500 cattle-farming people in the Akhmet region of Georgia. In winter and spring they are in the village of Zemo Alvani on the Kakhetian Alazani river; in summer they are in the mountain meadows on the upper Kakhetian Alazani and Tush Alazani.

CHEROKEE

10,000 SPEAKERS

United States

One of the AMERIND LANGUAGES, Cherokee is related to the Iroquoian group, as was first observed by Benjamin Smith Barton in *New views of the origin of the tribes and nations of America* (1797), though the Cherokee were not one of the original Five Nations of the Iroquois.

In the 17th century the Cherokee lived in the southern Appalachians, from which the spreading colonists expelled them in 1838. A few hid in the mountains, emerging in 1849 to settle in what is now the Qualla Reservation, North Carolina, but the great majority trekked to north-eastern Oklahoma. In 1821 a Cherokee, Sequoyah, devised a script for the language: this came at just the right time to be used by the first missionary to the Cherokee, Samuel Worcester, who began his work in 1825 and published translations, religious texts and almanacs in the script. It was widely used for a century or more. In the 19th century several short-lived newspapers – *Cherokee phoenix, Cherokee advocate, Cherokee messenger* – were published in the language.

Fewer Cherokee know the script now, though even in the 20th century there has been a *Cherokee newsletter* and a series of publications in the language sponsored by the Carnegie Cherokee Project.
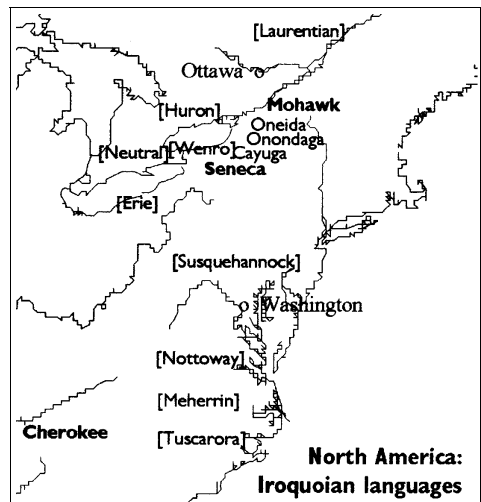
Oklahoma Cherokee has become a tonal language: linguists recognise three level tones, a rising tone and two falling tones. Cherokee is important for its indigenous literature of myths and sacred ceremonial, recorded by ethnologists in the late 19th century. The first ten numerals in Cherokee are: *sawu, ta'li, tsō, nv̄g, hīsg, sudal, galquōg, tsunēl, sonēl, sgo*.

The Iroquoian languages

Cherokee is classed as the only Southern Iroquoian language. It is now spoken in North Carolina and Oklahoma.

The surviving *Northern Iroquoian* languages are Cayuga, Mohawk, Oneida, Onondaga, Seneca and Tuscarora. They are spoken by small communities in reservations and reserves in New York state, Ontario and Quebec.

Laurentian is the usual name for the extinct Iroquoian language spoken on the shores of the St Lawrence when Jacques Cartier explored the river in 1534 and 1535. The language had disappeared completely when later explorers visited the region. It was, however, closely related to *Huron*, also now extinct, once spoken in central Ontario between lakes Ontario and Huron. Samuel Champlain made notes on Huron in 1615 and a grammar was published in 1632.

North America:
Iroquoian languages

The Cherokee syllabary

	a		e		i		o		u		v
D	a	R	e	T	i	Ꭳ	o	Ꮎ	u	i	v
Ꮪ	ga	Ꮐ	ge	Ᏹ	gi		go	J	gu	Ꭱ	gv
Ꮽ	ka		ke		ki	Λ	ko		ku		kv
W	ta	Ꮦ	te	Ꭻ	ti		to	S	tu	Ꮼ	tv
Ꮮ	da	Ꮷ	de	Ꮣ	di	V	do		du		dv
Ꮖ	ha	Ꭾ	he	Ꭽ	hi	Ꮀ	ho	Γ	hu	Ꮂ	hv
W	la	�	le	Ꮅ	li	G	lo	M	lu	Ꮑ	lv
Ꮉ	ma	Ꮋ	me	H	mi	Ꮝ	mo	�孓	mu		
Θ	na	Λ	ne	h	ni	Z	no	Ꮕ	nu	Ꮕ	nv
Ꮮ	hna										
G	nah										
Ꮖ	qua	Ꮻ	que	Ꮯ	qui	Ꮺ	quo	ꮃ	quu	Ᏼ	quv
Ꮀ	sa	Ꮞ	se	Ꮿ	si	Ꮆ	so	Ꮎ	su	R	sv
Ꮜ	s										
Ꮪ	dla	L	dle	C	dli	Ꮬ	dlo	Ꮲ	dlu	P	dlv
Ꮭ	tla		tle		tli		tlo		tlu		tlv
G	tsa	Ꮴ	tse	Ꮶ	tsi	K	tso	J	tsu	Ꮳ	tsv
G	wa	Ꮍ	we	Ꮎ	wi	Ꮼ	wo	Ꮄ	wu	Ꮾ	wv
Ꮝ	ya	Ᏸ	ye	Ꮿ	yi	Ꮖ	yo	G	yu	B	yv

Font: *Cherokee.ttf* by Joseph LoCicero IV

Sequoyah's syllabary for Cherokee, invented in 1821, is quite independent of all other scripts, though some of its characters resemble Roman ones. In the transliteration *v* stands for a nasal vowel like the French *un*.

CHHATTISGARHI

PERHAPS 7,000,000 SPEAKERS

India

Although it can be called a dialect of Hindi in the widest sense (see map there), most linguists agree that Chhattisgarhi, like Bvhojpuri to its north, is best regarded as a separate language. It is one of the INDO-ARYAN LANGUAGES, and is spoken in south-eastern Madhya Pradesh in the districts of Raipur, Bilaspur, Raigarh, eastern Balaghat and northern Bastar.

Chhattisgarh is literally the 'country of the 36 forts'.

The first ten numerals in Chhattisgarhi are: *ek, dui, tīn, cār, pāṃc, che, sāt, āṭh, nō, das*.

Chinese

PERHAPS 1,125,000,000 SPEAKERS OF CHINESE LANGUAGES

China, Taiwan, Hong Kong, Malaysia, Singapore and many other countries

Chinese is one of the SINO-TIBETAN LANGUAGES – or rather, it is a group of different languages, all descended from proto-Sino-Tibetan.

In origin, the Chinese languages are the speech of the farming communities of northern and south-eastern China and of the peoples of the coast as far as the borders of Vietnam. But one of them is also the language of a great and very long-lived empire. Fifteen hundred years ago, the northern Chinese – speakers of the language that is directly ancestral to 'Mandarin' – were already spreading their rule southwards and inland. In the 20th century the southern Chinese languages, like the unrelated minority languages of the south and south-west, are all of them threatened by the inexorable encroachment of the speech of the capital.

Chinese through history

Although the Chinese languages definitely belong to the same family as Tibetan, Burmese and their relatives, they differ radically. Proto-Sino-Tibetan must be dated many thousands of years ago. Possibly it was spoken in what is now southern China. Proto-Chinese, most likely, developed along the coast and the coastal valleys south of the *Yángzě* delta – the area of Chinese speech where dialect diversity is still at its peak.

The oracle bones of about 1400 BC, discovered at archaeological sites near the lower Yangze, are the oldest written documents of Chinese, which thus has a recorded history as long as that of Greek. True Chinese literature begins later: with collections of poetry, and with the works of Confucius and Mencius, around 500 BC. By that time Chinese was the language not only of the lower Yangze but of the wide lands to the north, in which the historic Chinese capitals, Beijing not the least of them, grew to eminence.

The classical literature of China is extremely rich, including short and long poetry, history, memoirs and many other prose genres. Chinese can claim the oldest printed literature in the world, dating back to a Buddhist sutra printed from wood blocks in AD 868.

Prose fiction has been cultivated in Chinese for much longer than in any other language. *The Water Margin* and *The Story of the Stone (Dream of the Red Chamber)* are, in their different styles, landmarks of world literature.

It is not surprising that there are several major languages extending across the vast and heavily populated region that is modern China. It is more surprising that one language alone – *Pǔtōnghuà*, the 'common language' or Mandarin – has so many hundreds of millions of speakers across northern China. Elsewhere in the world, single languages have not managed to extend themselves so far, except for rather brief periods. Examples are the Latin of the Roman Empire, the English of the British Dominions and the United States, the Russian of the Soviet Union: but not one of these languages has yet achieved such an enormous number of first-language speakers. Here Putonghua is unique, and its uniqueness must come from three factors: North China has been under a single government, most of the time, for well over a thousand years; administration and culture have shown impressive stability and uniformity over all this time; and, probably, northern peasants have found it more necessary than those

in the south to travel and to resettle in order to escape famine and to find work.

In the modern context Putonghua has every reason to spread further and faster. It is propagated by the administration of a centralised state, by its media, by its education and by ever increasing nationwide travel and migration. In the past, very few people in southern China knew the language of the capital: nowadays, very many do, especially the young.

The 'internal' history of Chinese – the history of its words and its sounds – is very difficult to trace, since the script does not directly reflect the sounds of the language. A 'rhyming dictionary' compiled in AD 601 contributed to the researches of Bernhard Karlgren, who reconstructed the sounds of Ancient Chinese in *Grammata Serica* (1940) and *Grammata Serica recensa*. Karlgren's work has revolutionised many aspects of Chinese studies. Thanks to him we know far more about the patterns and forms of ancient poetry, the origins and histories of the other Chinese languages, and, much further back, the phonology and grammar of proto-Sino-Tibetan, the ancestral language out of which, many thousands of years ago, Chinese, Karen and the Tibeto-Burman languages all developed.

The best known of all monosyllabic languages, Chinese has about 1,600 possible syllables in its sound pattern. How can all the concepts of the modern world be specified in only 1,600 words? They cannot: although each syllable can be regarded as an independent word, Chinese strings them together, as would any other language, to specify ideas.

Varieties of Chinese

Together, the Chinese languages have more mother-tongue speakers than any other language on earth. Even on its own, Putonghua probably achieves this first place.

The Chinese languages are not mutually intelligible, but they share the Chinese writing system – and thus Chinese literature belongs to all of them equally, though in informal writing plenty of local variations may occur. That is why all the languages are treated here under the single heading 'Chinese'.

Putonghua or Mandarin first spread as the language of the capital, Beijing, and thus of administrators and scholars throughout the vast empire. China has for many centuries had a highly centralised system of higher education and centrally organised recruitment to the higher echelons of the civil service. The speech of Beijing has for all this time been the natural language for communication among Chinese from different regions, and the obvious first language to learn for non-Chinese living or working in China.

Putonghua is distinguished among the Chinese languages in having only four tones (high; high rising; falling-rising; falling), no voiced stops and no syllable-final consonants.

Wú is named after the old state of Wu, whose capital, two thousand years ago, was Soochow. Shanghai is nowadays the centre of Wu speech. The language has three ranges of stops – voiced, voiceless unaspirated, voiceless aspirated – as did middle Chinese. Like Putonghua it is one of the few languages of the world in which *z*, or a sound very like it, functions as a vowel.

Gàn is named after the river Gan, flowing through the province of Jiangxi.

Xiāng is the language of Hunan province: but while 'Old Xiang' of the mountains and valleys of the south and east is a quite distinct language, 'New Xiang' of the north-west and the cities is becoming a mixed language, not too different from neighbouring Putonghua dialects.

Yuè or Cantonese, named after the old southern state of Yue, is the major language of southern China, centred on the great trading city of Canton. Yue preserves all the middle Chinese syllable-final consonants, *p t k m n ng*. It sounds markedly different from Putonghua.

Kejia or Hakka is the language of northerners who moved south in medieval times. *Hakka* is in fact a Yue term meaning 'guests', and *Kèjiā* is the Putonghua reading of the same word. Kejia is close to Gan but with phonological and lexical differences.

Mǐn, Hokkien or Teochew is a group of nine mutually unintelligible dialects of Fújiàn and Taiwan. Min is the typical language of the Chinese fishing community of Singapore.

Abroad it is sometimes known as Swatownese, Swatow being the major emigration port for the speakers of this language. Swatow itself is in Guangdong province, but its cultural links are with Fujian to its north. The whole region is geographically isolated from the rest of China – except by coastal trade – and this has maintained its linguistic diversity.

North China and south China differ in more than their relative linguistic diversity. *Nán chuán, běi mǎ*, according to the Chinese proverb: 'in the south the boat, in the north the horse'. Travel, south of the Yangze, was traditionally by water.

Chinese in the Latin alphabet

Pinyin originated (as *Latinxua*) in the Soviet Union in 1931, at a time when most minority Soviet languages were being given Latin orthographies. A slightly revised form was named as China's official romanisation system in 1958. It is now very widely used both inside China, whenever transliteration is needed, and by people who write about China in other languages. It has taken longer to catch on in Hong Kong, Taiwan and Singapore. Some well-known Chinese place names are still familiar in an otherwise outdated mid-19th-century transliteration: *Peking* (Pinyin *Beijing*), *Fukien*.

The Wade-Giles system, developed by two British scholars in the late 19th century, is the earliest that is still in general use. It is often used in historical and literary texts about China in English, especially older ones. Those who study China and Chinese need to be able to convert between Wade-Giles and Pinyin. The most obvious difference between the two is that while Pinyin writes the commonest stop consonants *b d g – p t k*, Wade-Giles writes them *p t k – p' t' k'*.

The Yale system was devised in the 1940s when there was a sudden upsurge in Oriental language study in the United States. Other systems exist, such as *Gwoyeu Romatzyh*, but they are seldom met with except by specialists.

All the systems offer ways of writing the tones of Chinese, but most users omit the tone marks.

Periods of Chinese

Proto-Chinese: before 500 ʙᴄ
Archaic Chinese: 500 ʙᴄ to ᴀᴅ 1
Ancient or Middle Chinese: ᴀᴅ 1 to 1000
Modern Chinese: ᴀᴅ 1000 onwards
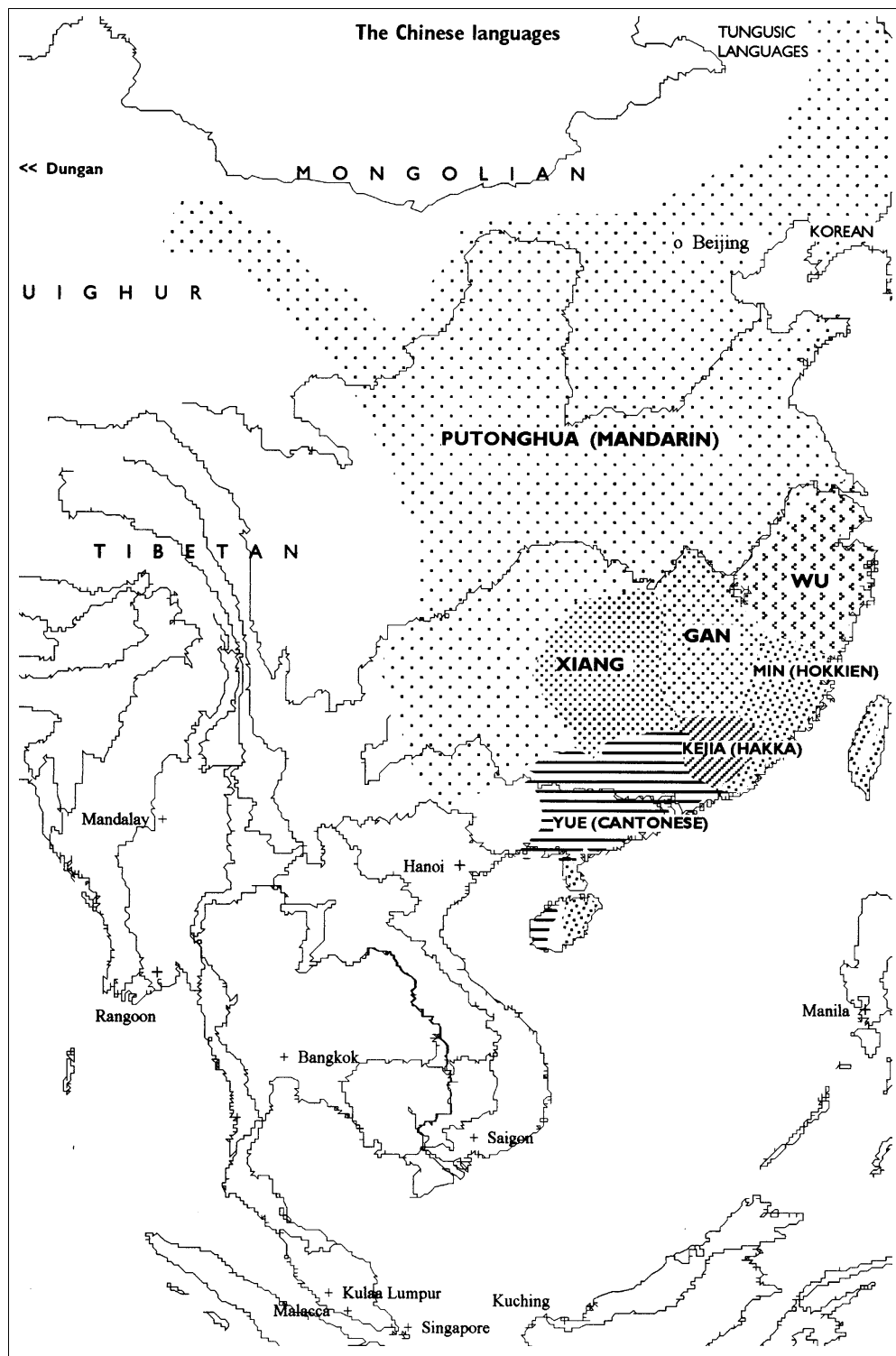
The Chinese languages on the map

Pǔtōnghuà or Mandarin (800,000,000 speakers) is in origin the language of the neighbourhood of Beijing. It is spoken in the north of China generally; by the 7,000,000 Chinese Muslims, an official nationality of Chinese Mongolia; and by rapidly increasing numbers throughout China. Overlying the regional languages, there are regional dialects of Putonghua. In its own homeland the major dialect divisions are *northern* (within which there is a distinct *north-western* dialect), *southern* or eastern – spoken around Nanking and the lower Yangze, and *south-western*, the language of Sichuan. Putonghua has also been called *Kuanhua, Pekingese, Kuo-yü* and by other names.

A medieval offshoot of what is now called Putonghua, the Chinese that is the language of traditional learning in Vietnam is pronounced on a system established in the 10th century: that was when Vietnam became independent and thus ceased to draw on the administrators and educators of the empire. More recently, Vietnam's Chinese community has used Yue (Cantonese) as its lingua franca.

Gàn (25,000,000 speakers) is the language of Jiāngxī.

Kèjiā (40,000,000 speakers), once well known as Hakka, is spoken throughout south-eastern China in agricultural communities in Yue and Min-speaking areas. It is the language of the New Territories of Hong Kong.

Mǐn (50,000,000 speakers) is the language of Fújiàn, also spoken across the water in Taiwan. It is the language of many Chinese speakers abroad, including some very important communities in south-east Asia, notably Singapore and

The Chinese languages

<< Dungan

TUNGUSIC
LANGUAGES

M O N G O L I A N

o Beijing

KOREAN

U I G H U R

T I B E T A N

PUTONGHUA (MANDARIN)

WU

GAN

XIANG

MIN (HOKKIEN)

KEJIA (HAKKA)

Mandalay +

YUE (CANTONESE)

Hanoi +

Rangoon

Manila +

+ Bangkok

+ Saigon

+ Kuala Lumpur

Kuching

Malacca +

+ Singapore

Bangkok. It has had many names: Hokkien, Fukien, Teochew, Teochiu, Chiuchow, Cháozhōu, Taeciw, Swatownese. The coastal dialects of Swatow, Amoy and Foochow are the best known abroad: inland dialects are said to be quite distinct. *Mĭn Nán* ('Southern Min') extends along the coast of Guăngdōng and is also spoken on Hăinán island.

Wú (90,000,000 speakers) is the language of Shanghai, the heavily populated coastal districts nearby, and Zhejiang. With the recent vast growth of Shanghai, the dialect of the old Wu capital of Soochow is now 'old-fashioned' rather than standard.

Xiāng or Hunanese (55,000,000 speakers) is the language of Hunan.

Yuè or Cantonese (65,000,000 speakers) is spoken in a large area of southern China. A form of Yue, influenced by English and Malay, is the lingua franca of Hong Kong, and is the majority language of the Chinese communities in the United States and in Britain.

Dungan or Tung-an has about 50,000 speakers in Kyrgyzstan and Kazakhstan and was a national language of the Soviet Union. Unlike all other varieties of Chinese, Dungan is written in Cyrillic script.

The Chinese script

The oldest known passages of written Chinese are 'oracle bones' and inscriptions of 3,500 years ago. The brief texts scratched on bones were clearly used for fortune-telling.

In those 3,500 years Chinese script has developed massively: yet some characters are still recognisable. As shown in the box, they name concrete and familiar objects.

Chinese characters
Ancient form

Modern form

sun moon man river cow
After Richard Newnham, *About Chinese*
(Harmondsworth: Penguin, 1971) pp. 35–6

A thousand years later, when Chinese literature had begun to be written down, all words – however abstract or concrete their meaning – could be recorded. The stock of characters had grown in several ways, but the most productive was by the doubling of characters: often the right-hand element was borrowed to give a sound or a 'sounds-like', the left-hand element to suggest a meaning.

Single or doubled, each character makes up a square block of text and each represents a single syllable of the spoken language.

In communist China a simplification of the individual characters and a reduction in the total number of characters have both been promulgated. Here Chinese is now normally printed from left to right across the page, like the Latin alphabet. The older pattern, still widespread in Taiwan, Hong Kong, Singapore and elsewhere, is to lay out the page with lines read vertically and following from right to left.

Most Chinese dictionaries arrange characters

Numerals in Chinese

yī	一	1
èr	二	2
sān	三	3
sì	四	4
wŭ	五	5
liù	六	6
qī	七	7
bā	八	8
jiŭ	九	9
shí	十	10

not by sound but according to the number of strokes (from 1 to 17) used to write the character, or used to write the 'radical' element of a doubled character.

CHOKWE

1,000,000 SPEAKERS

Angola, Congo (Kinshasa), Zambia

Chokwe, one of the BANTU LANGUAGES, is spoken in the upper Kasai valley and along the left bank tributaries of this river.

Chokwe speech has spread to its present territory in the last two hundred years. Tradition speaks of a Chokwe origin in the old Northern LUNDA state of Mwata Yamvo, perhaps in the 16th century, and of a migration from there to the area of Kangamba in south-eastern Angola: but this may be the history of a ruling group rather than of the whole people. At any rate, speakers of the related Mbwela, Luchazi and Mbunda (see map) are now to be found around Kangamba, and Luimbi is the dialect of a smaller group to the north-west. Chokwe itself has spread along the river valleys northwards, westwards and southwards, for its speakers were travellers and traders, particularly active in the slave trade. To the west they traded with the MBUNDU of Luanda. In the north they eventually conquered their traditional point of origin, the Lunda empire, in 1885–8: there is now a belt of Chokwe speech dividing Northern Lunda from Southern Lunda.

Soon afterwards, with the colonial partitions, Chokwe territory was divided between Angola and the Congo Free State, now Congo (Kinshasa). Some eastward migrants later settled in Northern Rhodesia (now Zambia). LUVALE, language of the upper Zambezi, is important as a second language for many speakers of Chokwe, Mbunda and Luchazi who inhabit the surrounding hills.

The first six numerals in Chokwe are: *-mu, -ali, -tatu, -wana, -taanu, sambano*. It is unusual among Bantu languages in the number of nouns that require no class prefixes. In some verb forms, tone is used to distinguish between negative and positive – not an easy feature for the foreign learner.

'There are almost as many variations on the word *Chokwe* as there are writers on this area,' wrote Merran. McCulloch in 1951 'Practically every combination of the prefixes *ba-, ka-, u-, tu-, va-,* or *wa-* with the roots *choko, cokwa, chiboque* (used by Livingstone), *chivoque, chiboqwe, chioko, chiokwe, djok, jok, kioko, kioque, quioco, tsioko, tshioko, tshiokwe, tshoko, tsiboko, tschibokwe, tschiwokwe* or *tschivoque* has been used. Reasons for this variation include the adoption by writers of the names used by neighbouring tribes in referring to the Chokwe, and the correspondence of some of the forms with dialectal differences within Chokwe itself . . . The most correct rendering of the name is *Kachokue* in the singular, *Achokue*, or in the north *Tuchokue*, in the plural. Chokwe in Angola are occasionally included in the term *Ganguella*, and Chokwe immigrants in [Zambia] are included by the LOZI in the term *Wiko*, and by Europeans as well as by other tribes in the term *Balovale*' (M. McCulloch, *The Southern Lunda and related peoples* (London: International African Institute, 1951) p. 28)
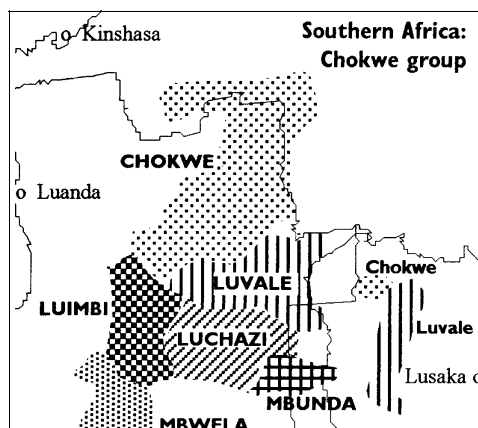
The speakers of Luimbi, Luchazi, Mbunda and Mbwela were despised by the MBUNDU-speaking slave traders of the Benguela highlands to their west. They were given the general label *oviNgangela*, also applied to the smaller Nyemba community. 'As a grass hut is not a house, so an *ociNgangela* is not a person,' the proverb said. This derogatory term recurs as *Ganguella* in Portuguese ethnographical sources.

Luvale, Chokwe, Luchazi and Mbunda: the map

Chokwe territory is in Lunda province of Angola and Kasai Occidental and Bandundu provinces of Congo (Kinshasa).

Mbwela, *Mbunda*, *Luchazi* and *Luimbi*, with an additional 350,000 speakers in Angola and Zambia, may be regarded as co-dialects with Chokwe. Luchazi has two dialects, those of the *vaka-ntunda* 'river people' and *vaka-ndonga* 'bush people', differing in phonetics and in vocabulary. The smaller Mbunda and Luchazi population in Zambia was reinforced by refugees in 1917, after a Mbunda-Luchazi uprising in Angola was brutally suppressed.

LUVALE is the language of the headwaters of the Zambezi. Politically, Luvale speakers are divided between the North Western province of Zambia and México province of Angola.

Southern Africa: Chokwe group

CHUVASH

## 1,500,000 SPEAKERS

Russia

Chuvash is one of the TURKIC LANGUAGES – but so different from the rest that a few linguists prefer to count it as a separate branch of the ALTAIC family. It is spoken in Chuvashia, a self-governing republic of Russia.

Modern Bulgarian is a Slavonic language. But the Bulgars, after whom Bulgaria is named, spoke a language related to Turkic. While one group established themselves in the southern Balkans – and eventually lost their language – a second group remained on the banks of the Volga. The Volga Bulgar state disappeared from history. But it is now believed by many scholars that one dialect of the Volga Bulgars' language of medieval times survives. It is what we call Chuvash (*Chăvash*).

Much earlier than the Turkic migrations and conquests that spread other Turkic languages across Asia, the 'Bulgars' first appear in historical records, living on the steppes north of the Black Sea. They were allies or subjects of the mysterious Khazar Empire of the 7th to 10th centuries AD. While the Khazars ruled the lower Don and Volga valleys and the steppes between, the Bulgars' homeland was to the north, in the middle Volga valley, roughly where the Chuvash now live. The ruins of their old capital city, Bolgar, are still to be seen just south of the confluence of the Volga and the Kama.

The Volga Bulgars used Arabic script. Their language is known from tomb inscriptions and from the many Volga Bulgar loanwords in modern HUNGARIAN. Early Hungarian speakers clearly lived and fought beside the Bulgars in the Russian steppe in the centuries before they reached their own new home in the middle Danube valley – modern Hungary.

Moscow

Cheboksary
Kazan

CHUVASH

Omsk

Russia:
The Chuvash diaspora
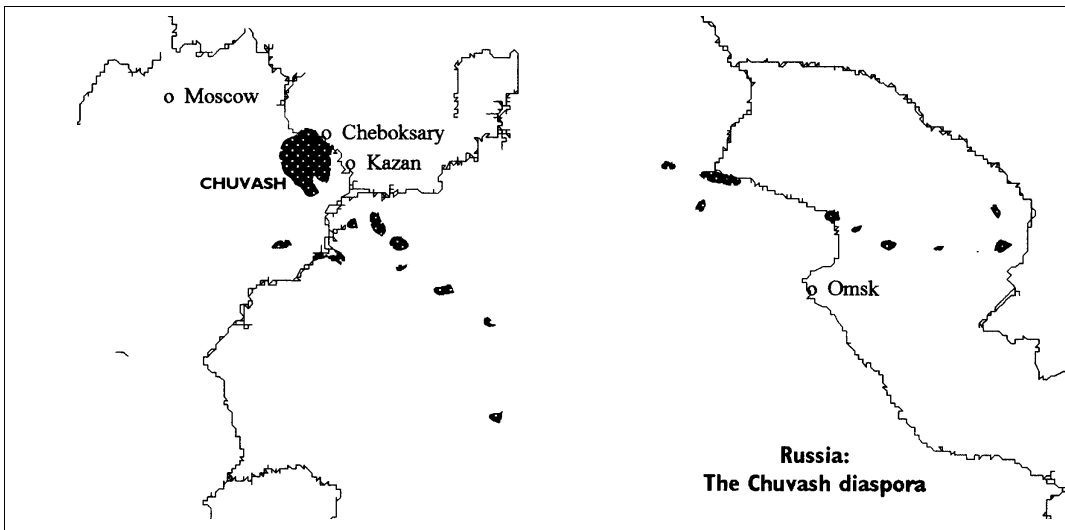
Making do

Akh mochi, kvartir yar,	Uncle, give me shelter,
Kvartir yar-ta khərne par.	Give me shelter and let me have the girl.
Khərə pultır tüshekle,	The girl will be my pillow,
tüshekki pultır syittila,	The pillow will be my blanket,
syitti pultır syüsyele,	The blanket will be my bedspread,
syüsyi pultır okala,	The bedspread will be my braid,
oki pultır syarsam.	The braid will be my knapsack.

Folk song sung by a Chuvash prisoner of war. From Robert Lach, *Gesänge russischer Kriegsgefangener.* Vol. 1 pt 4: Chuvash songs (Vienna, 1940) p. 126

In later medieval times the Mongol Khanate of Kazan' ruled the middle Volga. The Russian Emperor Ivan the Terrible took control in the mid 16th century. From then on, Russian administration and Russian immigration gradually linked the region more and more closely with the heartland of Russia. The process was not trouble-free. The 'Cossack' rebel, Stepan Razin, emerged from Chuvash country to challenge Russian rule in a series of bloody adventures in the 17th century.

Like the speakers of Mordvin, Chuvash speakers were converted to Orthodox Christianity by Russian missionaries. The Chuvash had not until now written their own language, but a form of the older Cyrillic alphabet was introduced by the missionaries about 1730. Active development of the literary language came only in the later 19th century, particularly after Y. Y. Yakovlev refined the alphabet in the 1870s. The script was revised again, based on modern Cyrillic, in 1938. For this alphabet see TURKIC LANGUAGES.

The language is used in schools and in the local press. It differs considerably from other Turkic languages, having long been influenced by neighbouring Uralic languages and by Russian. The first ten numerals are: *pər, ikə, vizhə, tavat, pilək, ult, zhits, sakar, takhar, vun*.

The Chuvash diaspora

Chuvash speakers form over two-thirds of the population of Chuvashia, in the middle Volga valley, with its capital at Cheboksary (Chuvash *Shupashkar*). But half the speakers of Chuvash live outside the republic, in a series of enclaves stretching eastwards across Russia and to Krasnoyarsk and beyond.

In Chuvashia itself two dialects are commonly recognised, *Anatri* or 'downstream' (in territory of the old Simbirsk province) and *Viryal* or 'upstream' (in the old Kazan province). Literary Chuvash is based on Anatri.

Medieval relatives of Chuvash were once spoken further down the Volga valley – in the extensive state of the Volga Bulgars – and far to the west, where Bulgar conquests reached the southern Balkans.
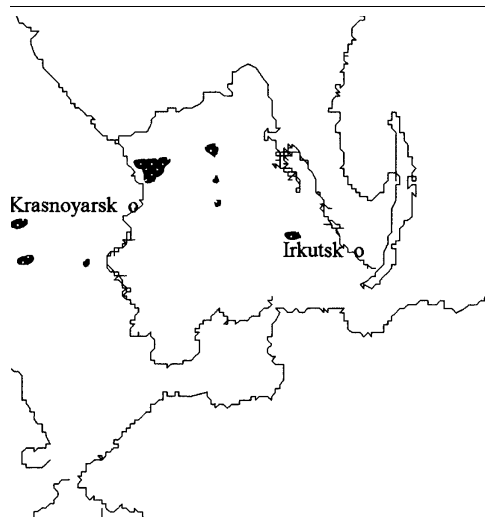
CIRCASSIAN

PERHAPS 1,500,000 SPEAKERS

Turkey, Russia and other countries

A group of dialects of the North West CAUCA-
SIAN LANGUAGES are often referred to together
as Circassian or Cherkess. In Russia they have
officially been regarded as two separate lan-
guages, Adyge and Kabardo-Cherkess.

There are nearly a million 'Circassians' in
Turkey, where many of their ancestors fled
to escape Christian rule when Russia conquered
this part of the Caucasus in 1864. Others came
to Turkey as slaves or 'migrant workers': in
some roles Circassians were much in demand,
young men as soldiers, young women as
domestic servants and concubines. King Hus-
sein of Jordan's personal guard is manned by
Circassians.

Most who call themselves Circassians in
Turkey nowadays are probably still speakers
of a dialect of Adyge, Cherkess or Kabardian –
but the language gets no encouragement in
Turkey and is now in steep decline. Smaller
numbers had settled in the regions that are
now Syria, Jordan, Israel, Serbia and Bosnia.
Circassian was taught in schools in Syria till the
1950s, and is still taught in a few schools in
Israel.

The speakers now to be found in Russia
number about 500,000: Adyge (West Circassian)
in the Adyge Republic, Kabardian (East Cir-
cassian) for those who shared the Kabardo-Balk-
ar ASSR but who have now voted to establish an
autonomous republic of Kabarda. Other speak-
ers of Kabardian live in the Cherkess Republic:
they are known as Cherkess. For a map see
ABKHAZ.

In the 19th century Circassian was occas-
ionally written in Cyrillic, more often in Arabic
script. Adyge and Kabardian were both written in
Latin script in the 1920s and in Cyrillic from the
mid 1930s.

Literary Adyge in Russia is based on the
Temirgoi dialect. Literary Kabardian is based
on the Baksan dialect.

Ubykh was once a quite separate language of
this family, spoken on the east coast of the
Black Sea by as many as 50,000 people.
Practically all of them fled in 1864: the lan-
guage is now extinct in the Caucasus. In
former Ubykh-speaking districts a distinct
dialect of Circassian is now to be heard. In
Turkey, Ubykh speakers turned to Turkish
and Circassian (in the latter they were al-
ready bilingual). The last speaker died in
1992. The three surviving dialects are called
by their own speakers *Adyge*, *Kabarday*,
Sherdjes.

Numerals in Abkhaz and Circassian

Abkhaz		Circassian
akè	1	zə
üba	2	t'ü
khp'a	3	śə
p'śba	4	ptə
khuba	5	tkh'u
fba	6	khə
bzhba	7	blə
ābà	8	jji
zba	9	bɣu
z'abà	10	pś'ə

Based on Adolf Dirr, *Einführung in das Studium
der kaukasischen Sprachen* (Leipzig: Asia
Major, 1928) p. 358

Coptic

CLASSICAL LANGUAGE OF EGYPT

After the conquests of Alexander (died 323 BC), Egypt was ruled by Greek-speaking monarchs, and ancient EGYPTIAN eventually ceased to be a language of government and prestige. Greek spread rapidly. It was to remain the lingua franca of Egypt and the eastern Mediterranean even under Roman rule.

Meanwhile, Egyptian remained the majority spoken language of Egypt. It was a new form of Egyptian, however: a language with an increasing number of Greek loanwords, and one that was no longer written in the hieroglyphic or demotic scripts, but in an alphabet based on Greek. This new form of Egyptian is called Coptic.

As the everyday speech of the Christians of Egypt, Coptic soon became a language of Bible translations and religious texts. The oldest Coptic texts are writings by early Christian saints and hermits, including St Antony of Egypt and St Pachomius. With the Arab conquest of Egypt, in the 7th century, Greek was soon forgotten and Coptic became all the more important as the special and unifying feature of the distinctive, eventually independent, 'Coptic Church'. Hymns of the 7th and 8th centuries demonstrate the power of Coptic religious poetry.

> In Greek the language and its speakers were *Aigyptioi*, 'Egyptian'. In Arabic this became *Qibṭ*. *Coptic* is a later Latinisation of the Arabic form.

Although religious texts in Bohairic Coptic are still used by the Christians of Egypt, they no longer understand the language. Arabic is now their spoken tongue. For modern scholars, meanwhile, Coptic has gained a new significance with the discovery of the Nag Hammadi library of sacred texts of the Gnostic religion. To linguists, Coptic is important as the only form of Egyptian in which the vowels were written: thus it helps in the reconstruction of the sound pattern of hieroglyphic texts.

Coptic dialects

Several dialects are recorded in early Coptic texts. As the language lost the official role that Egyptian had held, no single dialect was able to maintain its position as the standard.

The best-known dialect of Upper Egypt was *Sahidic* (Arabic *al-Ṣaʿīd* 'Upper Egypt'), essentially the dialect of Thebes, the southern capital. After the 5th century AD this was the only southern dialect still used in writing. Others known from earlier manuscripts are *Fayyūmic, Asyūṭic* and *Akhmīmic*.

The major Lower Egypt dialect was *Bohairic* (Arabic *al-Buḥairah*), reflecting the speech of Alexandria and Memphis. After the 11th century this became the standard dialect for all Coptic Christian texts.

Numerals in Coptic		
Masculine		**Feminine**
ογⲁ	ⲁ̄	ογⲉⲓ
ⲥⲛⲁγ	ⲃ̄	ⲥⲛ̄ⲧⲉ
ϣⲟⲙⲛ̄ⲧ	ⲅ̄	ϣⲟⲙⲧⲉ
ϥⲧⲟⲟγ	ⲇ̄	ϥⲧⲟⲉ
ϯⲟγ	ⲉ̄	ϯⲉ
ⲥⲟⲟγ	ⲋ̄	ⲥⲟⲉ
ⲥⲁϣϥ	ⲍ̄	ⲥⲁϣϥⲉ
ϣⲙⲟγⲛ	ⲏ̄	ϣⲙⲟγⲛⲉ
ⲯⲓⲥ	ⲑ̄	ⲯⲓⲧⲉ
ⲙⲏⲧ	ⲓ̄	ⲙⲏⲧⲉ

The Coptic alphabet

ⲁ ⲃ ⲅ ⲇ ⲉ ⲍ ⲏ ⲑ ⲓ ⲕ ⲗ ⲙ ⲛ ⲝ ⲟ ⲡ ⲣ ⲥ ⲧ ⲩ ⲫ ⲭ ⲯ ⲱ ⲳ ϥ ϩ ϫ ϭ †

a v gh dh e z ē th i k l m n x o p r s t w ph kh ps ō sh f ch h j g ti

---

Writing in Coptic

The alphabet is Greek – though it is usually printed in a different style, as shown in the box – with seven extra characters representing sounds unknown in Greek. The extra symbols are borrowed from the Demotic form of Egyptian script.

CREE

70,000 SPEAKERS

Canada

One of the Algonquian languages, Cree was spoken by an Amerindian people who occupied a vast territory in western Canada in the 17th and 18th centuries. In terms of population it remains Canada's largest indigenous minority language.

At their greatest reach Cree speakers ranged from Hudson's Bay westwards to Alberta, and prospered in the fur trade with the British and French. There is some Cree vocabulary (along with several other languages) to be found in James Isham's *Small account of the Indian language in Hudson's Bay*, written as long ago as 1743.

In the 19th century, somewhat reduced in numbers by smallpox epidemics and by war with Dakota and Blackfoot peoples, Cree speakers became of interest to missionaries. It was for Cree that the most important of the special scripts for North American Indian languages was first devised: it survives, though not all Cree speakers use it, and with modifications it also serves for Ojibwa and Inuit.

Some publications have appeared in Cree, both in the Cree syllabary and in the Latin alphabet (in various spellings). There was even a Cree newspaper, the *Native people*.

As a reminder of the importance of French trade contacts with American Indians, *Mitchif* or *French Cree* is the most long-lasting of all the mixed languages of Amerindian North America. It still has a few hundred speakers in North Dakota and Canada, who are 'Métis', descendants of French fur traders of the 17th and 18th centuries who married Amerindian women. Nowadays, speakers know neither French nor Cree.

There are four main dialect divisions of Cree but the two important ethnic divisions historically were the *Swampy Cree* of the forests of Manitoba and Ontario and the *Plains Cree* of Saskatchewan and Alberta. The latter were traditionally bison hunters, while the Swampy Cree hunted more generally: hare was a main constituent in their diet.

The Plains Cree were the more warlike of the two. Perhaps their dialect had already in the early 19th century become recognised by other Cree speakers as the most prestigious, the 'standard language'. For whatever reason, it was chosen by missionaries to be used for the Cree translation of the Bible, and this has had the effect of maintaining its prestige, so that when Swampy or Woods Cree speakers are asked for a Cree word they may well give the Plains Cree form rather than that of their own dialect.

The first ten numerals in Cree are: *peyak, nisho, nisto, newaw, niyalan, nikotwas, niswas, niyananew, shank, mita*.

The Cree syllabary

The Cree syllabary was invented in 1840 by James Evans, a Wesleyan missionary who had already worked with Ojibwa. It is partly inspired by English shorthand, partly by Evans' knowledge of Devanagari and other Indic scripts.

In an adapted form it is used for OJIBWA and for INUIT. The Ojibwa form is also used for 'Eastern Cree', the dialect of James Bay, and for Atikamek: for these Cree dialects see map at ALGONQUIAN LANGUAGES.

There is a special sign X for 'Christ'. Added characters for *r*- and *l*- syllables are found in the version of the script used by Roman Catholics.

The Cree syllabary

▽	▽·	∨	∪	٩	٦	٦	ᔆ	ᒪ	ᐸ
ā	wā	pā	tā	kā	chā	mā	nā	sā	yā
△	△·	∧	∩	ρ	ᑊ	Γ	σ	ᒡ	ᐱ
e	we	pe	te	ke	che	me	ne	se	ye
▷	▷·	>	ᗐ	ᑯ	ᒐ	ᒧ	ᑫ	ᖊ	ᐯ
o	wo	po	to	ko	cho	mo	no	so	yo
◁	◁·	<	⊂	ᑭ	ᒍ	ᒪ	ᑫ	ᓱ	ᐳ
u	wu	pu	tu	ku	chu	mu	nu	su	yu
◁	◁·	<	ᒼ	ᑲ	ᒡ	ᒪ	ᑲ	ᓴ	ᐲ
a	wa	pa	ta	ka	cha	ma	na	sa	ya

CRIOULO

PERHAPS 750,000 SPEAKERS OF CRIOULO AS A LINGUA FRANCA

Guiné, Cape Verde Islands, Senegal

A creole based on PORTUGUESE, Crioulo has been called by linguists *Kryôl* and *Upper Guinea Creole Portuguese* to distinguish it from the island creoles of the Bight of Benin.

Crioulo grew up as a pidgin used between Portuguese and Africans in the former Portuguese territories of Guiné Guinea-Bissau and Cape Verde. Nowadays in Guiné the creole is known to the great majority of inhabitants as a lingua franca, but it has also become the first language of many in towns.

The vocabulary of Crioulo is largely Portuguese but the grammar is closer to that of African languages of the hinterland: the influence of Mandekan is noticeable. Verb tenses are formed with prefixes and suffixes: *i bay* 'he went', *i ka bay* 'he didn't go', *i bay ba* 'he had gone', *i na bay* 'he is going'. The first ten numerals in Crioulo, still very similar to those of Portuguese, are: *un, dus, tris, kwatru, sinku, seis, seti, oytu, nobi, des*.

The majority of speakers, about 400,000, are in Guiné. In the Cape Verde Islands two dialects are recognised, *Sotavento* 'Leeward' and *Barlavento* 'Windward'. There are perhaps 50,000 speakers in the Basse-Casamance region of Senegal. There are still a few speakers of Crioulo in New Bedford, Massachusetts, to which a community of 'Bravas' emigrated in the 19th century.

CUSHITIC LANGUAGES

One of the major groups of AFROASIATIC LAN-GUAGES, Cushitic languages are spoken in Ethiopia, Somalia, Sudan, Kenya and Tanzania. According to one theory, this is the very region from which, many thousands of years ago, the earliest Afroasiatic languages began to spread. No Cushitic language was recorded in writing before recent times – yet Cushitic words have been recognised in ancient Egyptian references to southern peoples, a sign of the early presence of these languages in roughly their present location.

The language group is named after the ancient kingdom of *Kush* which once ruled the middle Nile valley (see NUBIAN LANGUAGES) – but the undeciphered 'Meroitic' language of this kingdom's inscriptions has no proved link with the Cushitic group.

Statistical work on their vocabularies suggests that the Cushitic languages began to differentiate – 'proto-Cushitic' began to divide into dialects – about seven thousand years ago. Beja, the northern Cushitic language, has probably never migrated very far: the rest perhaps spread generally southwards, and a notable migration must have led to the present isolation of Iraqw and the other Southern Cushitic languages in distant Tanzania.

Cushitic languages were almost certainly spoken, until less than two thousand years ago, in the parts of central Ethiopia where Semitic languages – such as Amharic – now predominate. The considerable differences between these 'Ethio-Semitic' languages and their close relative Arabic are to be explained by the influence of a Cushitic substrate, as linguists would say. In human terms, this means that an early form of Ethio-Semitic very rapidly became a language of communication for speakers whose mother tongue had been Cushitic.

Quara or *Felasha*, a dialect of western Agew, is the traditional language of the Falashi or 'Black Jews', a people of the central Ethiopian mountains remarkable for their adherence to a pre-Muhammadan religion closely resembling Judaism. The language is said to be now chiefly used in religious ritual: most Falashi speak Amharic or Tigrinya in daily life – or, since a recent mass migration to Israel, modern Hebrew.

Based on M. Lionel Bender, 'Introduction', Andrzei Zaborski, 'Cushitic overview' and Christopher Ehret, 'Cushitic prehistory' in *The non-Semitic languages of Ethiopia* ed. M. L. Bender (East Lansing: African Studies Center, Michigan State University, 1976) with other sources

Major Cushitic languages

The Cushitic languages of today may be divided into four subgroups. *Northern Cushitic* has just one member, BEJA, in north-eastern Sudan. *Central Cushitic* or *Agew* is a small group in northern Ethiopia: the language of the Falashi belongs to this group. *Southern Cushitic* consists of Iraqw and other little-known languages of Tanzania. Largest of the subgroups is *Eastern Cushitic*, divided in turn into *Highland* and *Lowland*. Major Lowland East Cushitic languages are AFAR, Konso, OROMO or Galla, Saho and SOMALI. Highland East Cushitic (including some of the languages once called *Sidama*) includes Gedeo, Hadiyya, Kembata and Sidamo.

In the absence of reliable linguistic censuses, estimates of numbers of speakers vary wildly, as indicated in the table.

AFAR	100,000–700,000	Eritrea, Djibouti, Ethiopia
BEJA	up to 1,000,000	Sudan, Eritrea
Bilin with Agw and Quara	200,000	Ethiopia, Eritrea
Darasa	250,000	Ethiopia
Gedeo	500,000	Ethiopia
Hadiyya	100,000–2,000,000	Ethiopia
Iraqw	111,000–338,000	Tanzania
Kambata (Timbaro and Alaba)	250,000–1,050,000	Ethiopia
Konso	60,000–150,000	Ethiopia
OROMO	7,500,000	Ethiopia, Kenya
Saho	120,000	Eritrea
Sidamo	650,000–1,400,000	Ethiopia
SOMALI	5,500,000	Somalia, Ethiopia

CZECH

12,000,000 SPEAKERS

Czech Republic

Czech is the official language of the Czech Republic, a mountain-ringed country of central Europe. It and its relative Slovak belong to the Western SLAVONIC LANGUAGES.

> The traditional name of the region where Czech is spoken is *Bohemia*, after the tribe *Boii* who possessed the territory in the 1st century AD, according to Roman sources. The dialects of *Moravia* to the south-east are also counted as Czech: this province occupies the valley of the *Morava*, a tributary of the Danube.
>
> Some think that Czech speakers' own name for themselves, *češi* (singular *čech*), derives from the word *četa* 'group', but this is uncertain. They call their language *česky*. The unusual English spelling, with its initial *Cz*-, comes via Polish: in other languages the word looks very different, French *Tchèque*, German *Tschechisch*. The term *Bohemian* was standard until the early 20th century.

Most scholars believe that the Boii were not Slavonic-speaking: but *when* Slavonic speakers first came to Bohemia is quite unknown. Some say they came westwards in the 5th century with Attila's Huns, and remained when Attila's invading force melted away.

A state of Bohemia can be traced from about 900 AD, when it freed itself from the short-lived and newly Christianised realm of Great Moravia. The roles were soon reversed: since that time Moravia has usually been ruled from Bohemia. The Kingdom was in general under German influence until the 17th century, when it became part of the Austrian Empire. Independence from Austria, and union (which lasted until very recently) with Slovakia, came at last in 1918.

Thus, for nearly a thousand years, German was the language of power in Czech-speaking lands. Yet Czech survived and flourished. Czech literature first appears in the 13th century. Prague University was founded under King Charles IV (1346–78), a patron of Czech culture. But the greatest figure of early Czech literature is the religious reformer Jan Hus (1369–1415: see box). The first Czech printed book, retelling the story of the Trojan War, appeared at Plzeň (German *Pilsen*, where the beer comes from) in 1468.

The decline of Czech literature under Austrian rule was stemmed by a national revival that began at the end of the 18th century. Folklore was a powerful influence on writers (and on composers such as Antonín Dvořák). Twentieth-century authors include the playwright, and now President, Václav Havel.

In the last hundred years, as many as 1,500,000 Czech speakers can be traced in the United States, not all of them still speakers of Czech.

Jan Hus (1369–1415)

Jan Hus, the religious reformer who gave his name to the Hussite movement, contributed important works to Czech literature – and also reformed Czech spelling. He identified the significant sounds (phonemes) of the language and prescribed one symbol per sound, adding diacritics to the Latin alphabet. His new letters are still used, some slightly modified: his *ṡ* and *ż* are now *š* and *ž*. Owing to his work, a page of Czech looks very different from a page of Polish, in which digraphs (double letters) such as *sz* and *rz* are used for similar sounds.

Numerals in West Slavonic languages

	Czech	Slovak	Polish	Upper Sorbian	Lower Sorbian
1	jeden	jeden	jeden	jedyn	jaden
2	dva	dva	dwa	dwaj	dwa
3	tři	tri	trzy	tři	tśo
4	čtyři	štyri	cztery	štyri	styŕo
5	pět	pät'	pięć	pjeć	pěš
6	šest	šest'	sześć	šěsć	šesć
7	sedm	sedem	siedem	sedm	sedym
8	osm	osem	osiem	wosm	wósym
9	devět	devät'	dziewięć	dźewjeć	źeẃeš
10	deset	desat'	dziesięć	dźesać	źaseš

In Czech and Sorbian, as in German, numbers such as '23' are formed on the pattern 'three-and-twenty' (German *dreiundzwanzig*, Czech *třiadvacet*, Upper Sorbian *tři a dwaceci*). In Slovak – as in English – such forms are thought old-fashioned or literary.

DANISH

5,500,000 SPEAKERS

Denmark, Greenland

Danish is a descendant of OLD NORSE. As Old Norse began to differentiate into dialects, the lowlands of southern Sweden and the crowded islands of Denmark allowed readier travel and communication than the backbone mountain chain of Scandinavia. So the 'East Norse' dialect of later Old Norse is the ancestor both of SWEDISH (see map there) and of Danish, a pair of languages that make up a single dialect continuum.

Danish has been the language of a powerful kingdom. In the 8th and 9th centuries Danish speakers ranged widely over England. At the Treaty of Wedmore, in 878, all England northeast of a line from Chester to the Thames was ceded to them. Although Norse speech did not survive for long in this 'Danelaw' which they settled and ruled, modern English has many loanwords, Norse, East Norse or Danish in character, as a result of the episode. *By-law* is a Norse loanword in English, meaning literally 'town law'.

The full union of England with Denmark, under Cnut (Canute) in the 11th century, was transitory. Meanwhile, the Viking conquests in Normandy must have retained their Norse speech for only a short time, though Norman French is still distinctive. From 1397 Denmark ruled Norway and Sweden: Sweden regained independence in 1523, but Norway remained under Danish political and cultural domination for four hundred years, and Norwegian still shows the signs of heavy Danish influence.

Danish, in turn, was influenced by the Low German of the Hansa cities, whose trade network criss-crossed the Baltic and the Danish islands.

Danish is distinguished from the other Scandinavian languages – from its close relative Swedish in particular – by its softened consonants. The stops *p t k*, between vowels, became *b d g* in Danish, and the *d* and *g* are pronounced as fricatives. For Swedish *kaka* (English *cake*) Danish has *kage*, pronounced with a light *gh*. Standard Danish also has a uvular *r*, like that of French, and many words end in a glottal stop. Together these phonetic characteristics give Danish a soft yet guttural sound pattern very different from all its neighbours.

The three principal dialects of Danish are Jutish (of the Jutland peninsula), Island Danish and the South Swedish or Dano-Swedish dialect group, which includes the island of Bornholm. This region of Sweden, the provinces of Skåne, Halland and Blekinge, was part of the Danish kingdom until 1658.

For historical reasons Danish has had an extensive reach as a literary language. Its range has gradually been shrinking since the 17th century. Until a separate Norwegian literary language was developed, from about 1830 onwards, Danish was the official language of Norway and Norwegians were prominent as Danish authors – Holberg was one. The best-known authors in the language are Søren Kierkegaard, the philosopher, and Hans Christian Andersen, whose folk tales were originally intended for adult reading though English versions are usually aimed at children.

Danish is still one of the official languages of the Faroes (see FAROESE) and of Greenland (see INUIT). It was the official language of Iceland until 1944.

Danish numerals

For the numerals from '1' to '10' see table at SWEDISH. Danish is unusual in the way that the tens from '50' to '90' are expressed. The recently devised 'new form' is used for writing cheques.

	Traditional forms	Literal meaning	New official form
10	ti	ten	ti
20	tyve	twenty	toti
30	tredive	thirty	treti
40	fyrretyve, fyrre	forty	firti
50	halvtredsindstyve, halvtreds	half the third twenty	femti
60	tresindstyve, tres	three twenties	seksti
70	halvfjerdsindstyve, halvfjerds	half the fourth twenty	syvti
80	firsindstyve, firs	four twenties	otti
90	halvfemsindstyve, halvfems	half the fifth twenty	niti
100	hundrede	hundred	hundrede

Dargwa

300,000 SPEAKERS

Russia

Dargwa or Dargin belongs to the North East CAUCASIAN LANGUAGES (see map there). It is the official and literary language for a group of peoples speaking related dialects, the most distinct of which are Kaitak or Qaidaq (with perhaps 15,000 speakers) and Kubachi. Standard Dargwa, however, is based on the dialect of Akusha.

The 2,500 inhabitants of the village of Kubachi were once famous in Russia, the Caucasus and the Middle East as goldsmiths and silversmiths. Their own name for themselves is *Ughbug*. Among Persians and Arabs they got the name *Zirekhgeran*, 'makers of chain mail'.

The Muslim speakers of Dargwa, Kaitak and Kubachi once used Arabic as their literary language. Dargwa became a written medium in the late 19th century, at first using Arabic script. As with so many minority languages of the Soviet Union, the Latin alphabet was introduced in 1928, to be abandoned in favour of Cyrillic script in 1938.

Lak is the language of about 120,000 people originating in the Kazi Kumukh valley in southern Dagestan. Related to Dargwa, Lak has separate official status as a literary language. Its speakers have moved repeatedly with the vagaries of Soviet ethnic policy. When the Chechens were deported en masse in 1944, many Lak moved from their high valleys to the newly empty town of Aukhov, renamed Novolak 'New Lak'. In 1992 a second Lak migration, to the Caspian coast near Makhachkala, allowed the Chechens to return.

'The Lak are among the most multilingual of all Dagestanis. They frequently know Kumyk and Russian as well as their native Lak language. Many are also fluent in Azeri, Avar and/or Dargwa' (R. Wixman, *The peoples of the USSR* (London: Macmillan, 1984) p. 122).

Numerals in Dargwa and Lak		
Dargwa		**Lak**
tsa	1	tsa-
k'wel	2	k'i-
ḥɛwal	3	shan-
aw'al	4	muq-
shwal	5	khyo-
urighal	6	rakhy-
werḥal	7	arul-
gaḥal	8	myay-
urch'imal	9	urch'-
wits'al	10	ats'-

DINKA

1,350,000 SPEAKERS

Sudan

One of the Nilotic group of NILO-SAHARAN LAN-GUAGES, Dinka is spoken in the upper Nile valley in southern Sudan.

Dinka speakers call themselves *Jieng*. The Arabic version of this is *Denkawi*, whence English *Dinka*.

A warlike and previously independent people, the Dinka were raided for slaves and exploited commercially by Egypt, supported by British adventurers, in the 19th century. The pattern thus established of exploitation from the north, punctuated by revolt, has continued.

Arabic and English have been used as lingua francas among the Dinka. Their closest linguistic relatives are their traditional enemies, the speakers of Nuer, whose territory is said to have expanded rapidly westwards, at the expense of Dinka speakers, in the 19th century.

Dinka and Nuer have a distinction of vowel register, contrasting 'hard' or 'clear' with 'breathy', comparable to that of south-east Asian languages such as KHMER.

The first ten numerals in Dinka are *tok, róu, dyak, 'nguan, wdyech, wdetem, wderóu, bêt, wde-nguan, wtyer*.
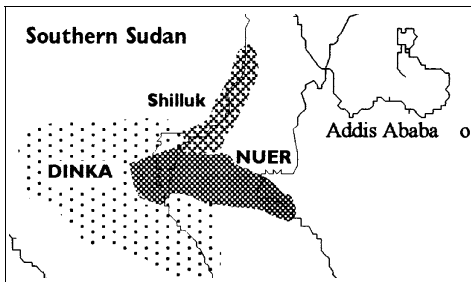
How Dinka and Nuer became rivals

'A legend relating the common ancestry states that in the past Dengdit, the great spirit of the Dinka, married a woman called *Alyet* in Dinka dialect and *Lit* in that of Nuer. Alyet gave birth to Akol who married Garung from whom were descended Deng and Nuer, brothers and ancestors of the Dinka and Nuer tribes respectively. When Garung died he left behind a cow and a calf, the former being bequeathed to Deng and the latter to Nuer. Deng stole the calf of Nuer and Nuer resorted to raiding Deng – just as today the Nuer raid Dinka cattle.'

Audrey Butt, *The Nilotes of the Anglo-Egyptian Sudan and Uganda* (London: International African Institute, 1952) p. 26

Dinka and Nuer on the map

There are four main dialects of Dinka: *Northern Dinka* or Padang; *Western Dinka* or Rek; Central or *South-western Dinka* or Agar; Eastern or *South-eastern Dinka* or Bor.

Nuer has 850,000 speakers on the Upper Nile and the Sobat river: their territory extends into Ethiopia. Their own name for themselves is *Naath*; *Naath cieng*, 'Homeland Nuer', for those west of the Nile; *Naath doar*, 'Bush Nuer', for those to the east. *Atuot* or *Thok cieng* is an isolated dialect spoken in the Lau valley.

Southern Sudan

Shilluk

Addis Ababa

NUER

DINKA

DIOLA

500,000 SPEAKERS

Senegal, Guiné, the Gambia

Diola or *Jola* has about 500,000 speakers on either side of the lower Casamance river, which reaches the Atlantic in southern Senegal. To the north the language extends, across a mainly rice-growing lowland zone, to the Gambian frontier; to the south it is also spoken at the mouth of the Cacheu river, in northern Guiné. It belongs to the Atlantic group of NIGER-CONGO LANGUAGES, and is one of the regional official languages of Senegal. Speakers call themselves *Kuyolāk*.

Diola was important in the early history of European exploration and exploitation of the West African coast. Two of its dialects can be traced in very early records. *Floup*, spoken at the mouth of the Cacheu, long an artery of Portuguese trade, is first recorded in a wordlist made for the French Royal Senegal Company in 1690. *Kasa* or Carabanne is the dialect of the Senegalese river towns on the Casamance, from Ziguinchor down to the Atlantic ports.

The major dialect in Senegal nowadays is the majority *Fony* or Fogny, spoken all the way from the Ziguinchor district to the Gambia border: it is the official standard, used on radio and television. For a table of numerals see WOLOF.

West Africa: Fulani, Diola, Serer and Wolof
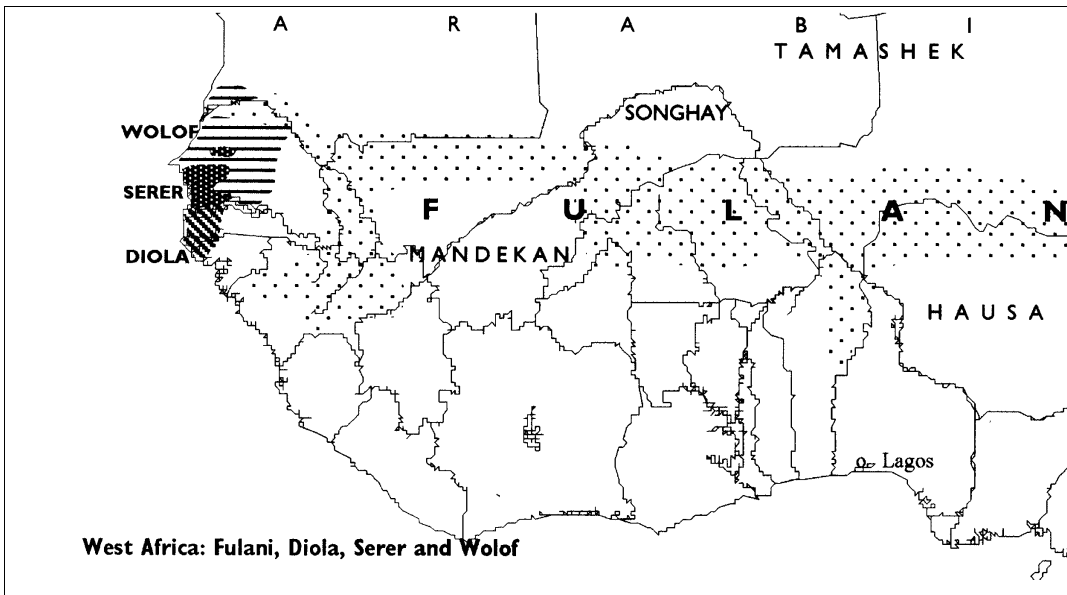
A sentence from J. David Sapir's *Grammar of Diola-Fogny* (Cambridge: Cambridge University Press, 1965) shows how loanwords can change so much in sound as to be scarcely recognisable. *Babay bukanak kariɛm pɔntitɛr ning Kuyɔlāk mankuriɛm ɛmānɔ,* 'There are people who eat *potatoes* the way the Diola eat rice.' 'Pɔntiter' is borrowed from French *pommes de terre*.

Fulani, Diola, Serer and Wolof

In spite of their scattering, the FULANI dialects remain in general mutually intelligible.

The dialect of the Futa Sénégalais is sometimes called *Poular*.

The dialect of the Futa Djallon is sometimes called *Fula*.

East of these a major Fulani centre extends from Macina to the Niger bend.

The major Nigerian dialects are those of Kano and the north, of Sokoto, and of Adamawa. There are two *Adamawa* dialect divisions, 'East Fulani' or *Foulbéré* or *Fulfulde Funaangere*, centred on Diamaré in Cameroun, and 'West Fulani' or *Fulfulde Hiirnaangere*.

Two other members of the Atlantic group are especially closely related to Fulani.

Serer, with 650,000 speakers, is one of the six official languages of Senegal. Its main dialects are the coastal *Serer Non* and *Serer Sin* of the Saloum river valley. A word-list in this language (as *Séraire*) was made for the French Royal Senegal Company as early as 1690.

WOLOF, with over two million speakers, is in practice Senegal's national language. Both of these are also spoken in the Gambia.

More distantly related are Diola, Balanta and Mandyak.

Diola or Jola has about 500,000 speakers. It also belongs to the Atlantic group, but is less closely allied to the other three. The major dialect of Senegal is *Fony* or Fogny, extending from the Ziguinchor district to the Gambia border. *Kasa* is the dialect of the river towns on the Casamance, including Ziguinchor itself. *Floup* is the main dialect of Guiné.

Balanta or Brassa has about 300,000 speakers, most of them in northern Guiné.

Mandyak (Portuguese *Manjaco*, French *Mandjaque*) has 150,000 speakers north-west of Bissau, the capital of Guiné.
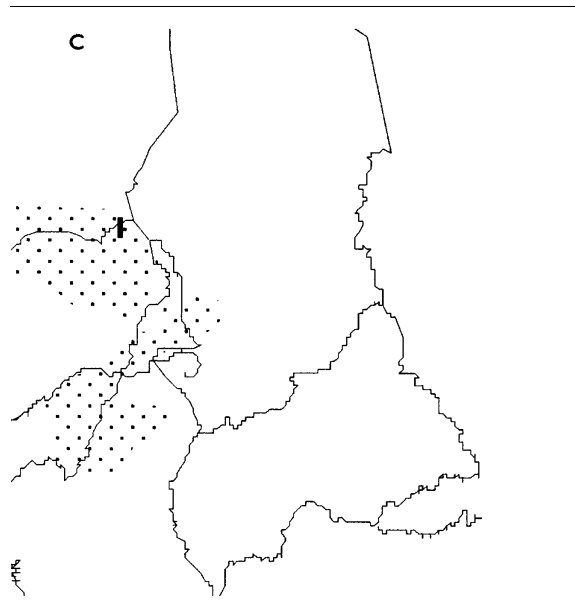
DIVEHI

100,000 SPEAKERS

Maldives

One of the INDO-ARYAN LANGUAGES, Divehi or Maldivian is the everyday language of the Maldives, now an independent state in the western Indian Ocean.

It has been suggested that the language was brought here by colonists from Sri Lanka, perhaps sixteen hundred years ago. At any rate, Divehi shows a closer relationship with SINHALA (see map there) than with the rest of the Indo-Aryan group: dialects of the southern islands show additional Sinhalese influence from later contacts. The Maldives were at first Buddhist, like Sri Lanka: Islam reached the islands in the 10th century and is still the majority religion.

The first twelve numerals are: *eke, de, tine, hatare, fahe, haye, hate, ashe, nuvaye, dihaye, ekolahe, dolahe*. These basic numerals resemble those of Sinhala, but larger numbers used to be built on a duodecimal base: '13' was *dolos-eke*, '14' was *dolos-de*, '24' was *fassehe*. A base of ten is now used.

The Tana script

Divehi shows its independence from Indian cultural influence partly in its script, called 'Tana' or 'Thaana'. This is quite different from Indic scripts in structure. Like Arabic script it is written from right to left, and its unknown inventor borrowed the shapes of the Arabic numerals and of some other Arabic characters. The script has been in use since about 1700: earlier Maldivian inscriptions, which date back to the 12th century, are in Indic-type scripts.

／ 𑀈 ᴍ ᴋ ʊ ʏ Ⅴ Λ 9 ＞ ⸱ ᴘ Ꙅ ɤ Ꙅ ᴡ ʏ ᴌ

h ś n r b ḷ k a v m f d t l g ñ s ḍ

DOGRI

1,300,000 SPEAKERS

India

One of the INDO-ARYAN LANGUAGES, and some-times regarded as the most aberrant of the dialects of PANJABI, Dogri is the language of the small former state of Jammu – which is now part of the resolutely Indian segment of the divided state of Kashmir.

Older Dogri texts can be found written in a distinctive script, related to the Gurmukhi script that is used for Panjabi. Nowadays Dogri is written in the Devanagari script familiar from Hindi. Dogri is rich in folk literature – from riddles to long tales and ballads – but there was little written literature before Indian independence in 1947. The Jammu radio station, opened in 1948, broad-casts in Dogri and encouraged local literature, as did the Cultural Academy founded in 1958.

The first ten numerals in Dogri are: *ik, do, trai, cār, pañj, che, sat, aṭh, nau, das.*

Hāṃ-re, kar-ke mahabbat mānue de rāh vic rahnde;
tāre gindī nūṃ raiṇ bihāwe.
Once having loved, she always longs for her man;
she passes the night counting the stars.

Couplet from a Dogri folk song

DONG

1,500,000 SPEAKERS

China

Dong is the most important language of the Kam-Sui group of KADAI LANGUAGES. The comparative linguist Fang Kuei Li discovered this language group, or at least brought it to the attention of European scholars for the first time, and demonstrated its distant relationship to the Tai languages.

Dong is the official Chinese name of the language. Its speakers call themselves *Nin-kam* and their language *Kam. Tung, Tung-chia* and *Tung-jen* are found in older sources. Chinese records sometimes confuse the Dong with the MIAO.

All the Kam-Sui languages are spoken by minority peoples of southern China. Dong is to be found further north-east than any other language of the Kadai family, where the borders of Hunan, Guizhou and Guangxi meet. Dong speakers are well known as musicians and singers – and as weavers of a specially fine cloth. The majority are wet rice farmers: MIAO speakers are their hill-farming neighbours.

'The Kam language is noteworthy for its extraordinarily large number of tone distinctions. Counting six pitch distinctions in "checked" syllables [ending in a consonant or glottal stop], most dialects of Kam are said to have fifteen different tones. This is surely close to a record.

(S. Robert Ramsey, *The languages of China* (Princeton: Princeton University Press, 1987) p. 244).

The first ten numerals in Sui are: *dau, gha, ham, hi, ngo, lyok, shet, pat, chu, sup.*

Songs at the drum tower

'Villages are located near water if possible. A unique structure is the so-called drum tower [*ta lei*]. These towers, of wood with tiled roofs, may reach 100 feet in height. The ground floor space serves as a combination village meeting hall and men's house. A wooden drum, suspended from the roof, is beaten in times of emergency and as a preliminary to village meetings. During the period between planting and harvesting, when farmwork is light, villagers gather at the drum tower, often inviting young people from neighbouring villages. On these occasions the two sexes may sing antiphonally throughout the night.'

Frank M. LeBar and others, *Ethnic groups of mainland southeast Asia* (New Haven: Human Relations Area Files Press, 1964) p. 231, based on work by Chen Kuo-chün and Inez de Beauclair

---

The Kam-Sui languages

The two important languages of this small group are Dong and Sui. *Dong* or Kam has about 1,500,000 speakers. *Sui* or Shui or Pa-shui has perhaps 200,000 speakers in Guizhou and Guangxi provinces – and a magical writing system all of its own. Speakers of Sui are said to have been settled in Guangdong province until they were forced to move in 1732.

CHINESE

SUI   DONG

YUE (CANTONESE)

Hong Kong

VIETNAMESE

Kam-Sui languages
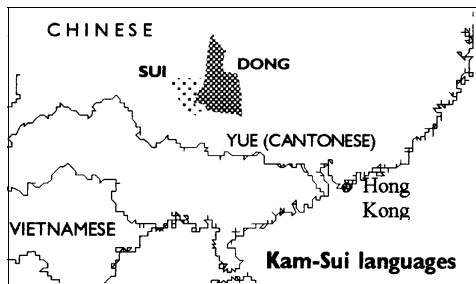
Dravidian languages

The Dravidian languages of southern India have a total of well over 170,000,000 speakers. They include four state languages (Malayalam, Telugu, Kannada and Tamil) the last of which is also an important language of Sri Lanka.

Drāviḍa is a Sanskrit name for the old Tamil kingdom whose capital was at Kancipuram near Madras. Already in the 8th century a Sanskrit writer had grouped the languages of the south as *Andhradrāviḍabhāṣā*, 'languages of the Telugu-Tamil group'. On this basis, *Dravidian* was suggested as a convenient label for the whole family by Robert Caldwell, Bishop of Madras, who published his *Comparative grammar of the Dravidian or South-Indian family of languages* in 1856.

They are unrelated to the Indo-Aryan languages to their north, which belong to the Indo-European family as do Greek, Latin and English. But there are tantalising similarities between the two groups (see box). Some period of shared development, or some episode leading to close mutual influence, must have produced this effect – at a date in prehistory which is now difficult to pin down.

There are minority Dravidian languages to be found far north of the line that marks the Dravidian/Indo-Aryan boundary, notably Brahui of western Pakistan. There is evidence that Dravidian once stretched further still: many linguists, ever since Caldwell's *Comparative grammar*, have believed that the family includes the mysterious, ancient Elamite language known from inscriptions and clay tablets from western Iran.

How early, then, did speakers of Indo-Aryan dialects meet speakers of prehistoric Dravidian? Or did other, now forgotten, languages transmit the same typical regional features to both groups? Is Burushaski a surviving specimen of

such a language? For even Vedic Sanskrit, spoken in the far north-west of India as long ago as 1000 BC, had already adopted the retroflex consonants so typical of Indian languages.

India has been a single cultural area for most of two millennia, and all the Dravidian-speaking region is heavily influenced by Sanskrit, the universal learned language of the subcontinent. Three of the major languages, Malayalam, Kannada and Telugu, have such a large proportion of Sanskrit loanwords that early Western linguists believed the whole family to be an aberrant group of Indo-Aryan languages, directly descended from Sanskrit, like those of the north.

The retroflex consonants

'Most of the languages of India have a set of retroflex, cerebral or domal consonants [formed by turning the tip of the tongue towards the roof of the mouth] in contrast with dentals. The retroflexes include stops *ṭ ḍ ṭh ḍh*, nasals *ṇ*, also in some languages sibilants *ṣ*, laterals *ḷ*, tremulants and even others. Indo-Aryan, Dravidian, Munda and even the far northern Burushaski form a practically solid bloc characterised by this phonological feature. In Dravidian it is a matter of the utmost certainty that retroflexes in contrast with dentals are proto-Dravidian in origin . . . In Southern Dravidian, moreover, several languages have three phonemic series in the front of the mouth – dental, alveolar, retroflex – a possibility hardly envisaged by the makers of the International Phonetic Association's alphabet.'

Murray B. Emeneau, *Language and linguistic area* (Stanford: Stanford University Press, 1980) pp. 110–11

English has alveolar and not dental consonants (in other words, the English *t d* are

formed by touching the alveolar ridge with the tongue, while in French and many other languages the comparable *t d* sounds are formed by touching the upper teeth with the tongue). To Indians, these English alveolar consonants sound more like the retroflex consonants *ṭ ḍ* than the dental consonants *t d*. So, in speaking English, most Indians are likely to pronounce *t d* as retroflex consonants – the characteristic feature of what the English call an 'Indian accent'.

Dravidian languages and others

Northern Dravidian languages include BRAHUI and KURUKH. The most important Central Dravidian language is GONDI: see the map there for information on others. Some include Telugu in Central Dravidian, but this is to overstate its differences from the other major languages of the Southern group.

The Southern Dravidian languages include KANNADA, MALAYALAM, TAMIL, TELUGU and TULU. Tamil is the only one with large numbers of speakers outside the Indian subcontinent – in Sri Lanka, Vietnam, Malaysia, Singapore, Zanzibar and even further afield.

Dravidian languages are not as well mapped as those of northern India. The great *Linguistic Survey of India* did not explore the states now called Tamilnadu, Andhra Pradesh and Karnataka: it missed some Dravidian languages even north of this.

The *Kurumbar*, nomadic shepherds of the mountains of southern India, have no separate language but speak dialects of those in their neighbourhood (see box at Kannada).

Several 'language isolates', apparently not belonging to any known language family, survive in India as reminders that the linguistic history of the subcontinent is more complicated than we can now reconstruct.

Burushaski, with 40,000 speakers in Kashmir valleys, is the most famous. The first ten numerals in Burushaski are: *hin, altan, iskin, walto, sundo, mashindo, thalo, altàmbo, huncho, tormo*.

The language of Elam

The great three-language inscription of King Darius at Behistun, in Iran, was the key that eventually allowed scholars to decipher all three languages. The most mysterious – at first with no recognised links to any other tongue – was Elamite, later found in a series of inscriptions and tablets dating from about 2200 to about 400 BC. They are the records of a once independent people, of what are now the provinces of Khuzestan and Fars in Iran, who were eventually incorporated in the great Persian Empire. Their language must have died out not long afterwards.

Bishop Caldwell, the 19th-century Dravidian scholar, was struck by the similarities between Elamite grammar and that of the DRAVIDIAN LANGUAGES. Scholars still pursue this insight – but without as yet reaching an agreed conclusion.

The cuneiform script is based on AKKADIAN, simplified, yet clearly not well adapted to Elamite sounds. Linguists have to work from scribes' spelling mistakes or variations to puzzle out the existence of such features as nasal vowels and final consonant clusters.

Duala

80,000 SPEAKERS

Cameroun

Duala, one of the BANTU LANGUAGES, is spoken in the once-famous trading city of Duala on the coast of Cameroun (for map see EWONDO).

European traders began to visit this coast in the late 15th century; the slave trade was established by 1500. These may have been the impulses that brought Duala speakers from an earlier inland territory to colonise the coast and found Duala around 1650. The city soon became a major exporter of slaves, ivory and aggrey, a blue coral-like rock much prized on the Gold Coast. Duala speakers made long trading journeys, inland and along the coast of Calabar, and their language became familiar to many other tribes as a lingua franca of trade.

Once fabulously wealthy, Duala crumbled, from about 1840, as the external market for slaves died. When the Germans invaded Cameroun in the late 19th century they themselves began to develop the inland trade of the region, displacing Duala speakers. At the same time, however, Duala was taken up by the Basle mission, which built on its existing position as a second language of trade to make it the vehicle of education in the districts of Victoria, Kumba and Mamfe – well beyond the area where Duala was a first language. Thus Duala became the first African language of Cameroun to be used for literature, and it provided the standard for official spellings of place names. The Bible was translated into Duala in 1862–72; the first grammar, by Alfred Saker, had appeared in 1855.

However, even in its heartland, Duala was in competition with another lingua franca – the Cameroun variety of pidgin English now usually known as KAMTOK. The German rulers, though they did not approve of Kamtok, found it easier to learn than Duala, and in practice they used it regularly. It became the language of plantations and of the market place in coastal Cameroun, while Duala became once more the local language of the city of Duala and its neighbourhood.

Duala has been influenced by neighbouring EFIK. Loanwords include Efik *makara*, Duala *mukala*, 'white man'. The first ten numerals are: *èwɔ, bebǎ, belalò, benêi, betanù, mùtoba, sàmba, lɔ̀ɱbì, dîbùa, dôm*.

Dutch

20,000,000 SPEAKERS

Netherlands, Belgium

One of the GERMANIC LANGUAGES, Dutch is in origin the western extremity of the dialect spectrum of Low GERMAN. It is the national language of the Netherlands. It is also one of the two national languages of Belgium, and there is a small Dutch-speaking minority in northern France near Dunkerque.

Nederland, or in full *Koninkrijk der Nederlande*, 'Kingdom of the Low Countries', is the official name of the country; *Holland* is sometimes used in English as shorthand for this. In French it is called *Pays-Bas*, a direct translation of the native name. *Belgium*, French *Belgique*, Flemish *België*, is a revival of a tribal and provincial name from the time of the Roman Empire – but the *Belgae* of that period spoke a Celtic language.

The native name of the language is *Duits*, which is in origin the same as *Deutsch* (see GERMAN). As a national language, however, Dutch is called *Nederlands. Vlaams* (*Flemish*) used to be the standard name in Belgium.

Most of the Netherlands, and almost all of Flemish-speaking Belgium, belong to the Lower Franconian dialect division of Low German. Standard Dutch originates as a variety of this – the language of the cities of Holland and Flanders, notably Amsterdam, Utrecht and Antwerp. As they passed through the hands of various European dynasties in medieval times, these centres of world trade became more prosperous and more powerful. Their independence, largely established by William of Nassau, Prince of Orange, in the 16th century, was recognised at the Treaty of Westfalia in 1648.

Belgium seceded in 1830 and became an officially bilingual state – one in which the language boundary is not to be ignored and linguistic rivalry sometimes borders on civil war. Brussels, its capital, forms an enclave where French is now the majority language in a region which is historically Flemish-speaking.

Belgium once ruled Zaire, Rwanda and Burundi, but French rather than Flemish served as the administrative language there. Dutch, however, was once the official language of a colonial empire and trade network that has included Indonesia, Sri Lanka, parts of South Africa, a varying number of possessions on the Indian and African coast, Suriname and some Caribbean islands. Names such as *Nassau* (capital of the Bahamas), *New Amsterdam* (the early name of New York) and *New Zealand* are reminders of the historical extent of Dutch exploration and trade.

Several local and mixed forms of Dutch grew up in the Dutch East Indies (Indonesia): all are now on the way to extinction, if they are not already extinct. *Petjo* was spoken by 'Indos', descendants of Dutch men and Asian women, in Batavia (Jakarta). *Javindo*, the mixed Dutch-Malay language of Semarang, was also known as *Krom-Hollands* 'Bad Dutch'. But AFRIKAANS, an offshoot of Dutch, now one of the major languages of an independent South Africa, stands as the greatest linguistic monument to Dutch expansion, trade and settlement overseas.

Linguists speak traditionally of *Old Franconian* but of *Middle* and *Modern Dutch*. The oldest, 9th-century, manuscript sources of Old Franconian are translations inserted in the Latin text of the *Laws of the Salic Franks*: there are also translations

of the Psalms. There is some 12th- and 13th-century poetry in Middle Dutch. The full Bible translation of 1619–37, called *Staten-Bijbel*, is a landmark of Modern Dutch literature.

Nederlands and Vlaams

The boundary between the Netherlands and Belgium is of recent origin and has nothing to do with major dialect divisions. However, differences do exist between the two national standards of the Netherlands and of Belgium. Amsterdam and the Hague are in the *Hollands* dialect zone, while Antwerp and Brussels are in *Brabant*. Manuals are published for Flemish speakers, urging them to avoid 'Brabantisms' or regionalisms. For 'sandpaper' they are told to use the Standard Dutch *schuurpapier*, not the Brabantism *zandpapier*.

Low German, Dutch and Frisian on the map

FRISIAN has significant dialect divisions. *West Frisian* is spoken in the Netherlands (some linguists distinguish *West Frisian* of the coast and the island of Terschelling from *Central Frisian*, spoken inland). *East Frisian* is spoken in a small moorland enclave in Saterland (the language of the East Frisian islands is Low German). *North Frisian* is spoken on the islands of Sylt, Amrum, Föhr and Helgoland (*Island Frisian*) and on the neighbouring Schleswig coast.

Low GERMAN consists of a series of dialects extending from east to west along the Baltic coast and its hinterland. *Niederpreußisch* and *Ostpommersch* are on the way to extinction as the German-speaking populations of the Königsberg (Kaliningrad) enclave and Poland were expelled at the end of the Second World War. Most now live in Germany. At the other end of the range, *Nordsächsisch* or *Niedersächsisch*, 'Lower Saxon', extends into the Netherlands, but the three most westerly dialects, *Limburgisch*, *Brabantisch* and *Holländisch*, grouped together as *Niederfränkisch* 'Lower Franconian', are spoken almost wholly in the Netherlands and Belgium and form the basis of modern Dutch.

Dutch as a world language

Throughout the former Dutch empire Dutch was an administrative language and was spoken by expatriates and mixed populations. In particular the local forms of Dutch in Indonesia were once varied and important, but they are now almost forgotten.

The most lasting linguistic legacy of Dutch expansion is certainly AFRIKAANS, the modified or creolised Dutch of South Africa.

New Amsterdam Dutch was spoken in parts of New York and New Jersey until about 1900. The dialects of Dutch settlers and of their black slaves were said to be noticeably different. Although the English took New Amsterdam (New York) in 1664, Dutch remained the official school language till 1773.
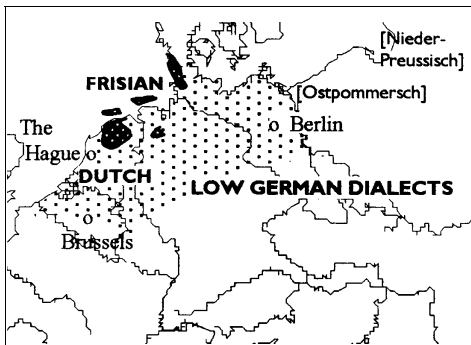
Numerals in Dutch, Afrikaans and Frisian

	Dutch	Afrikaans	Frisian
1	een	een	ien
2	twee	twee	twa
3	drie	drie	trije
4	vier	vier	fjouwer
5	vijf	vyf	fiif
6	zes	ses	seis
7	zeven	sewe	saun
8	acht	agt	acht
9	negen	nege	njuggen
10	tien	tien	tsien

**Dutch as a
world language**

[New Amsterdam Dutch]

DUTCH

[Negerhollands]

Berbice, Skepi

[Indonesian forms of Dutch]

AFRIKAANS
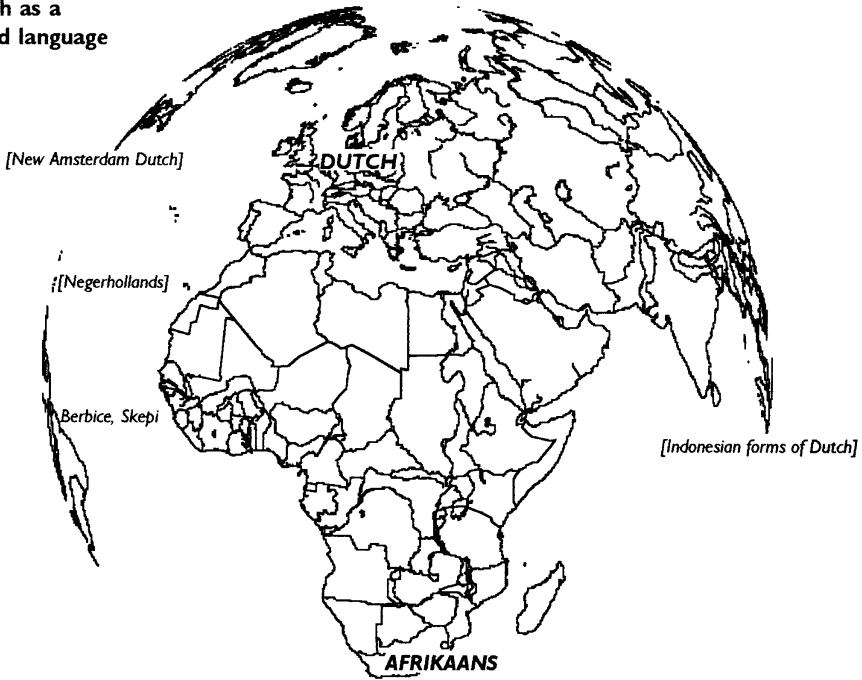
Negerhollands was the Creole Dutch of the Virgin Islands and Puerto Rico, now United States territories. Some publications appeared in Negerhollands in the 19th century, including a New Testament translation in 1818. It is probably now extinct.

In Guyana two Dutch creoles, *Berbice* and *Skepi*, have been discovered recently, both spoken by small communities in inland river valleys. They, also, are nearing extinction. But Dutch has had a considerable influence on PAPIAMENTO, the creole language of Curacao, and on Sranan, the lingua franca of Suriname (see ENGLISH CREOLES AND PIDGINS).

DZONGKHA

PERHAPS 500,000 SPEAKERS AS A SECOND LANGUAGE

Bhutan

As Mongolia, Tibet and other strongholds have succumbed to invasion and annexation, Bhutan is the one surviving independent state that continues the political and religious traditions of northern Buddhism. It became a unified state in the 17th century and a hereditary monarchy early in the 20th century.

Two of the three major regional languages of Bhutan (see map) belong to the 'Bodish' or Tibetan group: in other words, they are descended from the earliest reconstructable form of Tibetan (and are thus SINO-TIBETAN LANGUAGES). Pronunciation has changed so far, however, that they are mutually unintelligible with one another and with modern central Tibetan. After long and relatively undisturbed development in a country where inter-valley communication is not easy, numerous local dialects have grown up.

Classical Tibetan was, and still is, the language of religious education and culture across the whole of Bhutan. With unification, however, the need for an everyday lingua franca must have increased. The origin of Dzongkha can thus be traced to the 17th century, when it began to develop as a second language of army, administration and trade on the basis of Ngalong, the language of Thimphu.

Dzongkha (sometimes called *Bhutanese*) is, in written Tibetan, *rDzong-kha* 'language of the fortress'. It is now taught in all Bhutanese schools, as is English; a written form of it has been developed, a new use of the existing Tibetan script, which is most often seen in its local cursive form *rgyug-yig*.

For a table of numerals see TIBETAN.

Information from Michael Aris,
Bhutan: the early history of a Himalayan kingdom (Warminster: Aris & Phillips, 1979)

Languages of Bhutan

Ngalong is the language of the six valleys of western Bhutan, including Thimphu, the capital. Dzongkha originates as a lingua franca form of Ngalong.

Bumthang or Kebumtamp is spoken in central Bhutan; its territory also extends beyond that of Tsangla into the Indian state of Arunachal Pradesh, where it is sometimes known as 'northern Mon-pa'. Close to Bumthang is the language of the *Dag-pa* or Dap or Adap, a pastoralist community on the eastern edge of Bhutan. Earlier studies of this language called it 'Takpa' and 'Dwags', but it is quite different from the dialect of the *Dwags* province of southern Tibet.

Tsangla or Sharchagpakha or Sharchop, still a Sino-Tibetan language but more radically different from Tibetan than the other two, is the largest of the three regional languages, with perhaps 600,000 speakers in eastern Bhutan. There is a local assumption that Tsangla is the 'oldest' or 'original' language of the country.

NEPALI is spoken across a belt of south-western Bhutan.

Languages of Bhutan
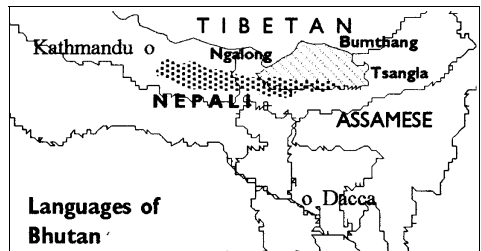
Edo

1,350,000 SPEAKERS

Nigeria

One of the Benue-Congo group of NIGER-CON-GO LANGUAGES, Edo was spoken in the once powerful and long-lasting Kingdom of Benin. It is now the major language of Bendel State, Nigeria.

Edo is the old local name of Benin City, capital of the kingdom. The language has also been called *Bini* and *Benin* by Europeans.

Oral tradition traces the historical kingdom back fifteen generations before the first European contact, the Portuguese visit to Benin in 1485. There are Edo legends of an older, mythical dynasty of sky kings, *ogiso* (see box) whose adventures linked them with personified animals and plants and pitted them against the animal trickster, the tortoise, *egwi*.

Benin grew rich in trade with the Portuguese, Dutch and British, exporting slaves, leopard skins and Guinea pepper. Its borders fluctuated and are now difficult to reconstruct – but Lagos is said to have been a Benin foundation and Onitsha was once under Benin suzerainty. IGBO titles of nobility are of Edo origin. Through the slave trade Edo exerted an influence on European creole languages, notably the Portuguese Creole of São Tomé.

The British conquered Benin in 1897, but re-established a titular kingdom. Though not one of the national languages of Nigeria, Edo is today a major regional language and is used in education and the media.

The Edo and the Yoruba

'According to Edo mythology, the Benin kingdom was founded by the youngest of the children of *Osanobua*, the high god. With his senior brothers, who included the first kings of Ife and other Yoruba kingdoms and the first king of the 'Europeans', he was sent to live in the world (*agbõ*). Each was allowed to take something with him. Some chose wealth, material or magical skills or implements but, on the instructions of a bird, the youngest chose a snail shell. When they arrived in the world they found it covered with water. The youngest son was told by the bird to upturn the snail shell and when he did so sand fell from it and spread out to form the land. So the first *Oba* of Benin became the owner of the land and his senior brothers had to come to him and barter their possessions in return for a place to settle. Hence, though he was the youngest son, he became the wealthiest and most powerful ruler'.

R. E. Bradbury, *The Benin kingdom* (London: International African Institute, 1957) p. 19

Edoid languages

Edo, with 1,000,000 speakers, is spoken across the central territory of the old kingdom.

Esan or Ishan, with 200,000 speakers, is locally an official language, used in primary schools and on television. It is said to originate with emigrants from Benin who established their own domains along the lower Niger.

Yekhee, Etsako or 'Kukuruku', with 150,000

speakers, is said to mark the position of a sub-ordinate kingdom. Esan and Yekhee may be regarded as dialects of Edo.

South West Edoid languages are *Urhobo* and *Isoko*. Together they have 650,000 speakers. They are not close enough to be intelligible to Edo speakers; the same applies to the *North West Edoid* and *Delta Edoid* (including *Engenni*) languages, which are spoken by much smaller communities.

Nigeria: Edoid languages
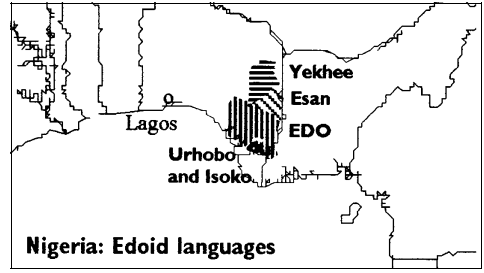
EFIK AND IBIBIO

2,750,000 SPEAKERS

Nigeria

One of the NIGER-CONGO LANGUAGES, this is a language of the Cross river shores and a lingua franca of Cross River State in Nigeria – a role that Nigerian Pidgin English may now begin to take over. Efik is used in education up to university level.

Efik is properly the name for the language spoken as a mother tongue in Calabar, Creek and along Cross river as far as Itu. *Ibibio* and *Annang* are a series of dialects partly mutually intelligible with Efik.

Ibibio traditions claim that they have been settled in their present location for many centuries, and this is likely to be true. Ibibio speakers were once numerous in the Atlantic slave trade. Ibibio speakers were reluctant colonials, continuing to resist the British until 1929 and after.

Missionaries began to work and teach in Efik at Calabar in 1846. As a vehicle of trade, Efik was already known along the routes that Calabar traders used. It now became a language of religion and culture, used in church, in education and in print. So although Efik has a relatively small number of speakers, it is nowadays ac-

West Africa: Efik and Ibibio
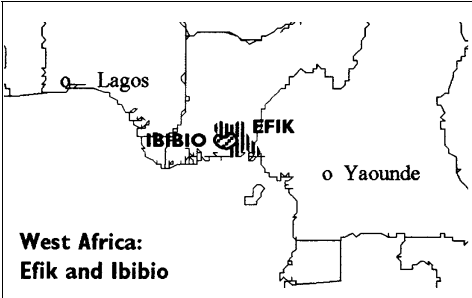
cepted as the standard language among Ibibio speakers generally, and continues to be used in the press and media.

Efik has four tones. An initial nasal, *m* or *n*, functions as a vowel and can take a tone. The first ten numerals in Ibibio are: *kèt, ìba, ìtá, ìnã, ìtyõ, ityôket, ìtyâba, ìtyâitá, ùsúkkèt, dwòp*.

Ibibio dialects on the map

Efik is accepted as the standard throughout Ibibio country. *Andoni* of the Atlantic coast and the western *Annang* are the most aberrant of the dialects.

EGYPTIAN

EXTINCT LANGUAGE OF EGYPT

Egyptian is one of the AFROASIATIC LANGUAGES, and the second oldest of all written languages. In its later form, COPTIC, Egyptian survived as a spoken language until a few hundred years ago: some Coptic texts are still used in the Christian churches of Egypt.

In its unmistakable hieroglyphic script, Egyptian was already being recorded in stone inscriptions before 3000 BC. It was the language of government and of the all-pervading state religion throughout the long life of the Egyptian kingdom, which eventually succumbed to Persian and then to Macedonian conquest in the 4th century BC. The fate of the dead was an abiding preoccupation of Egyptian culture – a preoccupation that resulted not only in the pyramids, but also in the lengthy historical, ritual and magical texts that make up the majority of surviving Egyptian literature.

The classical form of the language is Middle Egyptian, reflecting spoken Egyptian of about 2200 BC but formalised and frozen in official use for two thousand years.

The last important text in Middle Egyptian is the Rosetta Stone. Erected in 196 BC, this is inscribed in Middle Egyptian (hieroglyphic script), in Demotic and in Greek – the three written languages of prestige in Greek-ruled Egypt. It was crucial to the decipherment of the Egyptian scripts. In 1822 Jean-François Champollion identified the names in the Hieroglyphic and Demotic versions, showing that some signs were alphabetic while others were classifiers or represented whole words. The work of decipherment, following Champollion, was largely completed in the course of the 19th century.

The first ten numerals in Egyptian are – as written – *w, snw, ḫmt, fdw, diw, sisw, sfḫ, ḫmn, psd, mj*. For the vowels as recorded in a later form of the language, see table at COPTIC.

Periods of written Egyptian
Old Egyptian: 3100 to 2200 BC
Middle Egyptian: 2200 to 200 BC or later
Late Egyptian: 1600 to 700 BC
Demotic: 700 BC to AD 400
COPTIC: AD 200 to 1400 or later

The Hieroglyphic script: the main alphabetic signs

ʾ i y ʿ w b p f m n r h ḥ ḫ ẖ s š ḳ k g t ṯ d ḏ

ENGLISH

350,000,000 SPEAKERS

United States of America, United Kingdom, Canada, Australia, South Africa, India, Ireland, Malaysia, Singapore, Sri Lanka and many other countries

One of the GERMANIC LANGUAGES, English is rather different from the other members of the group (Dutch, German, Danish and the rest) – different in its early history and unique in its status in the 20th-century world.

Although Chinese has more mother-tongue speakers, no language rivals English in the extent to which it is used across the world. There are probably 800,000,000 people who speak English either as a first or second language. All available evidence tells us that this number will go on growing.

The history of English

English and Frisian must both have grown out of the unrecorded lingua franca of the North Sea seaways in the 4th and 5th centuries AD, when Germanic mercenaries fought for and against the Roman Empire, eventually overwhelming it. In Italy, southern France and Spain their migrations had come overland. To the north, in northern France and Britain, the lands of the empire were approached by sea.

Out of this ethnic and linguistic melting-pot, Frisian was to emerge in the mainland ports: but Frisian is now nothing more than a regional language in districts of the Netherlands and north-western Germany. English owes its origin to the migrants ('Angles, Saxons and Jutes') who crossed the sea and settled in the Roman provinces of Britain, particularly after the Roman army had withdrawn, from 449 onwards if the traditional dating is correct.

By about 600, Anglo-Saxon kingdoms covered most of what is now England, though not Cornwall or the Lake District. The Latin of the lowlands had died away (there are no direct traces of it in English loanwords), as would the Celtic language that had once been spoken there. Wessex (*West Seaxe*, the 'West Saxons') became the most powerful of the new kingdoms, the only one able to withstand the Viking invasions that followed in the 9th century. It was in Wessex that a written language first flourished, at the court of Alfred (849–99). Alfred himself was a capable poet and a translator from Latin.

The dialects of the kingdoms of Mercia, Northumbria and Kent, rather different from 'West Saxon', are known from a few literary texts and documents.

Anglo-Saxon poetry

Lytle hwile leaf beoð grene,	A little while leaves are green,
ðonne hie eft fealewiað, feallað on eorðan	Then they go yellow, fall to earth
and forweorniað, weorðað to duste . . .	and die, turn to dust.
	Solomon and Saturn

Old English verse is not rhymed: each line is alliterated on two or three stressed syllables. Similar patterns are found in the oldest German and Old Norse poetry.

England (medieval Latin *Anglia*; French *Angleterre*, 'land of the Angles') comes from the most obscure of the three tribal names associated with the Germanic migration. It is the land that the Angles and Saxons conquered for themselves, out of the larger territory that the Romans had called *Britannia*, 'Britain'. *English* is its language; but its culture before the Norman conquest is traditionally called *Anglo-Saxon*, and the same name used to be used for the language at that period. Linguists now prefer to emphasise the continuity between it and later periods and to call it *Old English*.

The greatest Old English literature consists of epics (*Beowulf*) and shorter poems in alliterative verse. Their makers were Christians, but Christianity was new to Anglo-Saxon England and had not yet overlaid the ethos of its traditional heroic poetry.

The Norman Conquest in 1066 drew a line under Anglo-Saxon literature and so, as far as we are concerned, under the Old English language as well. For the next two centuries the language of the court was French, and very little English was written – or at least very little survives. When 'Middle English' is again recorded, it had become a different language, heavily influenced by Norman French (for the first century of the conquest) and by Parisian French (from the mid 12th century onwards). The language is full of French loanwords – as it still is. In Old English there were several ways of forming the plural of nouns: in later English there is practically only one, *-s*, the same as in medieval French.

The Middle English period may be said to begin in 1066 and end with the introduction of printing in English in 1475. Alliterative verse was still written, such as the anonymous 14th-century masterpiece *Sir Gawain and the Green Knight*, but fashion turned to rhymed verse and to poetry and prose based on, or simply translated from, French. The most original writers – Geoffrey Chaucer (died 1400) in verse, the 15th-century 'knight prisoner' Sir Thomas Malory in prose – transformed these foreign influences into characteristic English forms. In the 12th and 13th centuries some of the greatest writers used not English but French (its English form is known as Anglo-Norman). Others wrote in Latin, the language of historical prose and of 'serious' and scholarly writing in many subjects.

The history of modern English has three important themes. The first is the extension of the language into new subject areas. Three crucial dates are the first printed English translation of the Bible, in 1525, the 'Authorised Version' of the English Bible in 1611, and the revised *Book of Common Prayer* in 1662. The latter two texts were in everyday use in Anglican churches until the 1970s: they influenced the speaking and writing of English for over 300 years. As in religion, so in science and scholarship, English came to take the place of Latin in the 16th century and after. Vast numbers of loanwords have been added to the language to make this possible.

The second theme is the spread of English to many parts of the world. It is the native language of English-speaking colonies, which have become independent and powerful states. Spread by British trade and influence, it is the second language of many other states that continue to need English because it is an international language and because, within their arbitrary ex-colonial frontiers, they have no better choice as national language. Worldwide, English is by far the most popular second language. It is the universal language of diplomacy and science (roles once played by French and Latin). It is the language that people usually try first when talking to 'foreigners'.

The third theme is the growth in England of a new and more widely accepted standard language, the language of London and its elite, nowadays identified as 'Standard British English', 'the King's (or Queen's) English', 'BBC English' and 'Received Pronunciation'. This has gone hand in hand with the spread of education and literacy, the extension of printing and publishing, and recently the influence of radio and television. All these influences have tended to freeze pronunciation, spelling, and spoken and written style.

But 'Standard British English' is now only one focus out of many.

Modern English, as it has gradually developed over the last five or six centuries, has some striking differences from earlier forms of the language. Final unstressed -*e*, a very common feature, was dropped; most former long vowels underwent a sound shift, the 'English Vowel Shift', and became diphthongs. Some examples: *dame, mole, pure* and *five*, and hundreds of words like them,

used to be pronounced as two syllables. They used to have a simple long first vowel (*a* as in *father, o* as in French *pôt, u* like the *oo* in *boot, i* like the *ee* in *been*). Now they are one-syllable words whose vowel sound has turned into a diphthong. But, when the change in pronunciation took place, English spelling did not change to reflect it: we do not write *deym, meul, pyur, fayv*.

Modern English noun forms

girl	singular base form
girl's	singular possessive
girls	plural base form
girls'	plural possessive

The last three of these have the same pronunciation. The distinction between them is a spelling rule that is taught by schoolteachers, with varying success: misplaced apostrophes have long been a common feature of public notices and of everyday writing. Copy editors and sub-editors maintain the rule in published books and newspapers.

Over-use of the possessive is a feature of unskilled writing in English by non-native speakers (*dog's food for dog food; *the paint's colour for the colour of the paint).

In marketing in France, English words are fashionable. The apostrophe is a reliable marker of English, and is favoured in hybrid forms such as *jean's* 'jeans', *pin's* 'lapel badge'.

Modern English verb forms

'Weak' verbs	'Strong' verbs	
play	take	1. base form
plays	takes	2. 3rd person singular, present
playing	taking	3. present participle
played	took	4. past
played	taken	5. past participle

Forms 2 and 3 can practically always be predicted from form 1. Forms 4 and 5 are identical, and predictable, in the so-called 'weak' verbs (the majority); they sometimes differ, and always have to be learnt individually, in the so-called 'strong' verbs. Young children sometimes get these wrong (*taked). The past tenses of strong verbs tend to vary from dialect to dialect (*brung* for *brought*; *blowed* for *blew* and *blown*). There are even a few differences among the standard varieties of English (US *dove, gotten*; British *dived, got*).

English has developed far from a proto-Indo-European structure, in which both nouns and verbs had many possible forms defining their function in the sentence. In English most nouns only have two forms in speech, four in writing; verbs have four or five forms (see box).

Influences on English

The oldest external influences on English come from the language that was spoken in England

when English first began to develop – its Celtic speech, a variety of early Welsh. Many river names came from Celtic, like *Avon* from Welsh *afon* 'water'. Some other place names are Celtic too: the town names typically derive from the Latin version of an even older Celtic name, which had been adopted by the Romans at the time of their conquest, like *Londinium*, modern *London*. In the form in which they were learnt by the Anglo-Saxons, some of these included the Latin word *castra* 'fort', which comes

through as modern *chester*, *-caster*: thus *Exeter* comes from the Latin (originally Celtic) town name *Isca* plus the Latin label *castra*, all passed on to the Anglo-Saxon invaders by no doubt Celtic-speaking inhabitants.

Few other English words are of early Celtic origin. Numerous examples have been proposed (*lad*, *lass*, *trudge*, *whap*, *mattock*, *drill*, *bran*, *bodkin*) but some have been disproved and most others are doubtful. A group of words including *bogey*, *boggle* and *bugbear* certainly has Welsh relatives. Other likely cases are *brisk*, *hog*. Additional Celtic words can be found in the western and northern dialects of English: sometimes, like *combe* (Welsh *cwm*) and *glen* (Gaelic *gleinn*), these eventually become known by speakers of standard English.

After the Anglo-Saxon settlements came the Viking raids and invasions of eastern England. It is because of these that English has a strong element of Norse words: Old Norse was a Germanic language, like Old English, but its forms were already quite distinct. Sometimes the Norse loanwords form 'doublets', when the same Germanic word already existed in Anglo-Saxon with a slightly different sense, as with Anglo-Saxon *shirt* and Norse *skirt*. Northern and eastern dialects have more Norse loanwords than standard English: Yorkshire *lake* 'play'; *gate* 'street' in place names.

Next came the single most powerful influence on English, that of French. The long term linguistic effect of the Norman Conquest was the incorporation of hundreds, even thousands of French loanwords in Middle and Modern English. Many of these are semantic doublets of existing Germanic words, with essentially the same meaning but with a cultured or elevated nuance: *clean* from Germanic, *pure* from French; *song* from Germanic, *chant* from French; *bed* from Germanic, *couch* from French.

In recent centuries English has drawn new words from most of the languages of Europe but, perhaps even more important, from the languages of other parts of the world: *chocolate* from Nahuatl; *jaguar* from Tupi; *curry* from Tamil or Kannada; *springbok* from Afrikaans.

Scientific terms: where from?

In the Western Roman Empire, Latin was the everyday language but Greek was the language of science and technology. If you spoke or wrote on scientific subjects in Latin, you would use Greek technical terms: *arithmetica, astronomia*.

In medieval western Europe, Greek was no longer widely known. Latin survived as the spoken and written language of scholarship. As new progress came in philosophy and science, in later medieval times, the Greek loanwords of older Latin were revived and new ones were wanted. Greek became the primary source for loanwords in medieval and modern Latin.

After the Renaissance, with the spread of literacy and printing in the modern languages of Europe, Latin gradually lost its role as the language of science. Science was now taught and written about in English, French, Italian, Spanish and other languages. New words were needed in these languages to cope with new subject areas; naturally the loanwords came from Latin – and naturally many of them were based on Greek – because these were the words that had already been used, by the same writers or by their forerunners, in scientific Latin.

New words are still invented all the time in science and technology. Most of them are still put together on the basis of Latin and Greek roots: *astrophysics, biomedicine*. This is the way it has always been done, and scientists are just as conservative as other people.

In German and Russian the tendency has been not to make up new words on Latin roots but to use the language's own resources. This can be seen in the language of psychoanalysis, a theory and technique developed by the Viennese Sigmund Freud (1856–1939). In his own writings Freud used ordinary German words in special senses for many of his new technical terms: *das Ich, das Es*. Freud's English translator, Ernest Jones, found it best not to use parallel English expressions (**the I, *the It*) and not to borrow the German words directly (**the Ich, *the Es*) but to use new *Latin* loanwords that matched Freud's terms. This is why English-speaking psychoanalysts talk of *the Ego* and *the Id*.
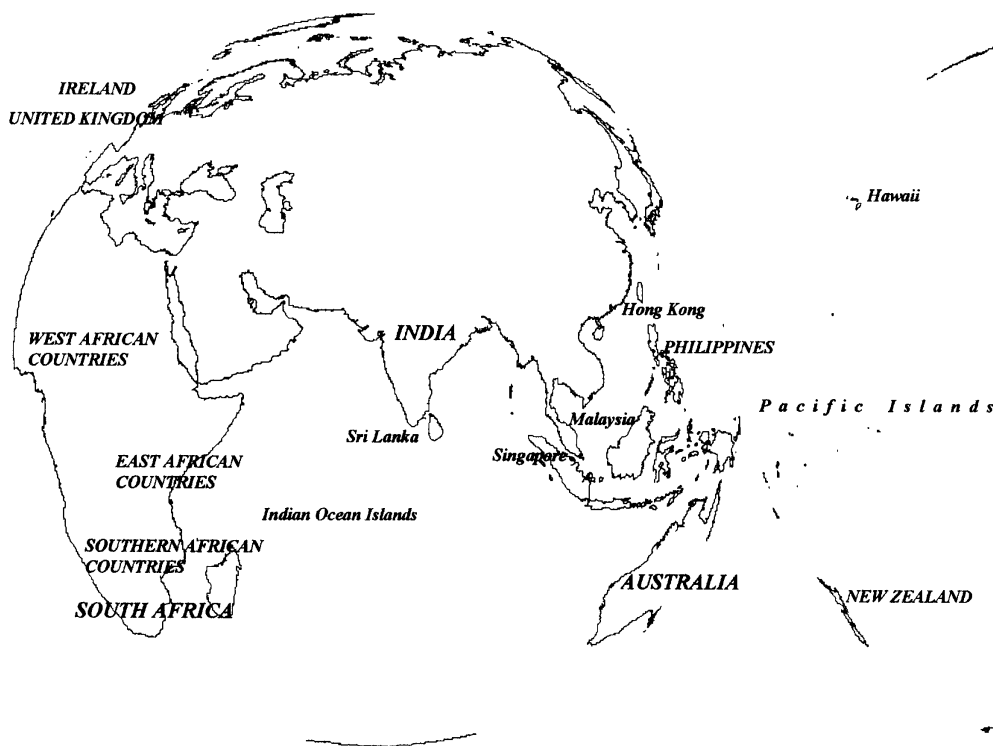
'Englishes' across the world

Neither American nor British English is noticeably more 'modern'. Each preserves features of older English which the other has lost. The regional dialects of North America and Britain are full of older words and forms that have disappeared from the standard varieties of English.

The Americas

Speakers of many languages crossed the Atlantic, voluntarily or involuntarily, to trade and settle in what are now the United States and Canada: English, French, Basque, Spanish, Dutch, German, Wolof, Akan, Yoruba, pidgin English and pidginised versions of several others. English arrived with John Cabot's expedition in 1497. From all the languages used there, English had emerged by the 18th century as North America's lingua franca.

Its history has been the merging of the languages of individuals and communities into a larger unity. But the unity of North American English has its own subdivisions: regional variants on an educated standard, regional dialects, city colloquials, an easily identifiable 'Canadian English', a range of forms grouped as 'Black English'. There has been plenty of room for these modern American dialects to retain fea-

been entirely shaped by migration. Nowadays, with powerful centralised media influences, the trend is strengthening, though speakers of highly distinct varieties – jargons of black city youth, 'hillbilly' dialects – may find the use of an almost private language advantageous and attractive.

English is a major language of communication in the central and eastern Caribbean, in a range of forms, from the regional standard (which has its slight variations from Standard American or British English) to local dialects of the island creoles.

Guyana is the furthest point south for the English language in the Americas (except for English-speaking minorities in Argentina and other South American countries). The furthest west is marked by two detached states of the US, Alaska and Hawaii: in the latter, standard English coexists with an English creole.

The British Isles

Standard British English differs from the American standard in its pronunciation (which Americans often describe as 'clipped'), in its written style, and in its vocabulary: *van* for *truck; tights* for *pantiehose; jam* for *jelly*.

The local dialects of England have been developing without interruption for 1,500 years – while continually influenced by the standard language and by one another. It is impossible to draw boundaries around them: every small locality has its own peculiarities of accent and vocabulary. On the larger scale, however, it is easy to recognise West Country, West Midland, East Midland, East Anglia, Cheshire, Lancashire, Yorkshire and North East dialects: cities such as London (*Cockney*), Bristol, Birmingham, Liverpool (*Scouse*) and Newcastle (*Geordie*) have their own colloquials. The country dialects are in a slow decline: increased population mobility, schools and broadcasting all work against them. The city dialects, along with the varieties of 'British Black English', are still vigorous. The Black English of British cities has close historical links with Jamaican and other Caribbean creoles – but its musical and poetic traditions have developed new and original features.

CANADA

UNITED STATES

Bermuda

Bahamas

Belize

Caribbean Islands

Guyana

Suriname

Argentina

English: the world language

tures of the original languages and dialects of the migrants. Black English, in particular, is partly shaped by the creoles that were the everyday speech in slave communities of the American South and the Caribbean.

On the margins of North American English, numerous other languages survive in local use: French, Spanish, German, Yiddish, English creoles such as Gullah, the French creole of Louisiana, Amerind languages. All these can still provide new loanwords in regional American English dialects.

The trend is towards standardisation rather than fragmentation. Perhaps it has always been so in these two countries whose demography has

The English of Wales has been developing for some hundreds of years. Two main dialects can be recognised: the southern is closer to English dialects of the West Midlands and the West Country; the northern is closer to north-west England and shows continuing strong influence from Welsh. Both have a characteristic sentence intonation comparable to that of Welsh.

English in Ireland dates from the English settlements of the 16th century and after. True Irish English is somewhat different from the conventional 'brogue' of fiction and theatre. The standard language of broadcasting and government incorporates Irish loanwords such as *garda* 'police', *Dáil* 'Parliament'. The Northern Irish variety is different again. The two religious communities of Northern Ireland have lived side by side for four hundred years, yet they are distinguished by certain features of accent.

The North Country score

Yan, *tan*, *tethera*, *methera*, *pimp*, *teezar*, *leezar*, *cattera*, *horna*, *dik*, *yandik*, *tandik*, *tetherdik*, *metherdik*, *bumpit*, *yan-a-bumpit*, *tan-a-bumpit*, *tethera-bumpit*, *methera-bumpit*, *jigot*. There are many versions of this counting jargon, used until quite recently in the north of England for special purposes – to count sheep; to count stitches; and in skipping games. My mother came from Lancashire, and this is the version that she knew.

Clearly it has something to do with the Welsh numerals. But it is hard to know whether it was remembered in the hills through a millennium and more, ever since Cadwallader fell and the Welsh were driven out of northern England; or whether it was learnt in some later contact between Wales and the North Country. The really striking feature is the way of saying 'sixteen' to 'nineteen'. The compound words mean 'one-and-fifteen', 'two-and-fifteen' and so on. This is exactly what is done in modern Welsh – but not, it seems, in very early Welsh.

The special history of Scots

The Anglo-Saxon settlements, from which the English language originated, actually extended north of the present border between England and Scotland. Gradually, both before and after the Norman conquest of England, English speakers went on spreading into the Lowlands of Scotland. They became a majority, though no one knows when this milestone was reached. Their form of speech eventually became the official tongue of the Scottish court and government.

In later medieval and early Renaissance times this language, Scots, was used not only in documents of all kinds but also in a great literature, from the poetry of Barbour (14th century) and Dunbar (c. 1320–95) to the prose of *The Kingis Quair* (whose author is usually thought to be James I of Scotland, 1394–1437).

When James VI of Scotland succeeded his cousin Elizabeth and became James I of England, the two courts became one. The political and cultural centre moved from Edinburgh to London. The Authorised Version of the Bible (1611), which was in the English of the South, was to have an enormous influence on literature all over Britain.

As an official and literary language, Scots quickly lost ground, though still used not only by Robert Burns (1759–96) but by other major writers such as the 19th-century novelist John Galt and the 20th-century poet Hugh Macdiarmid.

In Scotland now there is a continuum of speech forms. At one extreme is broad Scots, the descendant of the language of the medieval kingdom, strikingly different from southern English – and varying considerably across the geographical spread of Scotland. At the other is standard English spoken with a 'Scottish accent' and a few distinctive words such as *outwith*.

Scots, which in past times was the language of an independent government and a rich culture, is discouraged in schools and has no official status.

Africa and Asia

A highly distinctive standard is that of South Africa – the only one in Africa in which English

serves as a mother tongue. The vowel sounds of South African English are immediately recognisable to English speakers from other parts of the world. The standard English of South Africa shows the influence of Afrikaans, Zulu, Xhosa and other local languages.

In the other African countries of the Commonwealth, in India, Sri Lanka, Malaysia and Singapore, English is not generally a mother tongue but a national auxiliary language used to varying degrees in education, government and business.

Malaysian and Singaporean varieties range from a standard very close to British English (as used in publishing) to forms of speech that would be barely intelligible to an English-speaking visitor. Spoken Malaysian English often omits the verb *be/is: He very stupid*. Only one verb form is used: *He call out*, a natural development since many final consonants are dropped. Spoken sentences are often completed by an expressive particle, *la* (a loan from Min Chinese) or *man*.

Within Sri Lankan English, regional differences can be recognised between the north-east and the south: there are also differences between Sinhala and Tamil speakers.

Indian English

In the 17th century, Englishmen who went to India needed to learn Portuguese, or, better, the Indo-Portuguese pidgin which was then the lingua franca of trade. Persian, as the language of local diplomacy and the elite, was equally important.

English gradually replaced Portuguese, and English itself became a ruling language in India, in the course of the 18th century. From the beginning it required a special vocabulary for hundreds of details of Indian administration and the Indian way of life which were wholly different from those of England. These loanwords came from Portuguese, Persian, Arabic, Bengali, Tamil, 'Hindustani' (Hindi and Urdu) and from many other languages of southern Asia. Ever since that time, Indian English has been distinguished less by its accent than by its voca-

bulary: *lathi* 'baton'; *out of station* 'away on business'; *tiffin* 'lunch'.

The Indian English accent is also distinctive: along with nearly all Indian languages it has the retroflex *t, d, n* sounds pronounced with the tip of the tongue turned upwards. These form a handy substitute for the alveolar *t, d, n* sounds of other varieties of English, in which the tongue points to the alveolar ridge behind the upper teeth.

> In standard Indian English high numbers are counted in a special way. *Millions* are not used. Instead, *one lakh* is 'one hundred thousand' and *one crore* is 'ten million'. Numerals are punctuated to match. 1,27,55,380 is read in Indian English as *One crore, twenty-seven lakhs, fifty-five thousand, three hundred and eighty*: in British English it is *Twelve million, seven hundred and fifty-five thousand, three hundred and eighty*. American English omits the two *ands*.

'With the establishment of the first universities in 1857 English for all practical purposes became an Indian language' (S. Mathai). By the time British rule in India had flourished and declined, to end with independence in 1947, the English language was so solidly entrenched in education and in communications among speakers of the various indigenous languages that it was impossible to do without it. Its constitutional position has varied – but English remains, in practice, an essential lingua franca of India.

Australasia

New Zealand English is perhaps not easily distinguished from that of Australia by non-Antipodeans, though in fact the two regional forms have their differences. New Zealand English has naturally borrowed from Maori: *pakeha* 'white man', literally 'flea-bringer'. It has its own favoured phrases, such as *fire in the fern* 'wildfire'. Forenames drawn from Maori, such as *Ngaio*, have been fashionable at times.

Australian English has a history of almost two hundred years, and already has its own

regional dialects. It has many loanwords from aboriginal Australian languages (and *Kylie* is a name borrowed from an aboriginal source), but it is best known for its distinctive colloquial-isms at various levels of fashion and status: *sanga* 'sandwich', *the big spit* 'vomit', *sheila* 'woman'.

ENGLISH CREOLES AND PIDGINS

MANY MILLIONS OF SPEAKERS

This is a survey of the offspring of ENGLISH: the languages that have grown up in various parts of the world, in the last few centuries, wherever English, spoken in trade and conquest, has met other languages and mixed with them.

Why pidgins and creoles?

Pidgin languages are likely to develop wherever groups of speakers of two different languages need to communicate – and there is not the time or incentive for in-depth language learning. Typically this happens in the course of travel, trade, migration and conquest.

Pidgins tend to have limited vocabularies: enough for the circumstances in which they are used, buying and selling, giving and receiving orders, but no more.

Creoles are the offspring of pidgin languages. When pidgins for some reason become so essential to a community that they begin to be the first, and the only, language of at least some speakers, they are called creoles. Unlike most pidgins, creoles become capable of carrying the full range of information that members of any human community need to exchange: they are potentially as varied and as complex as any other language.

Pidgin languages are common: creoles less so. Not all pidgins ever become so essential within a community as to be the first language that children learn. But when migration throws speakers of different languages together permanently, then the conditions exist that make it likely that a creole will develop.

The Atlantic slave trade that began in the 16th century, and the 19th-century Pacific labour trade, both created these conditions. In many third world countries today, countries in which numerous regional and local languages are spoken, the powerful pull of city life and of industrial wages brings speakers of different languages together in a similar way. There, too, in great cities and busy towns, on marketplaces and at bus stations, pidgins are needed: in mixed-language communities they may develop into creoles.

There have been many guesses at the origin of the term *Pidgin*. The word is first found in N. Berncastle's *Voyage to China* (1850), and it is often said to derive from a Chinese pronunciation of the English word 'business' – or of the Portuguese word *ocupação*, which means the same. Others say that it comes from a Chinese phrase *pú-ts'ĭn* 'paying up'; or from the Portuguese expression *pequeno Portugues*, 'little Portuguese', which was used as a name for the Portuguese pidgin spoken in the coastal cities of Angola; or from the same Portuguese word *pequeno* in its Sranan form *pəcī*. There are other, less likely, suggestions, including a derivation from Hebrew *pidjom* 'ransom'.

The word *Creole*, French *créole*, Spanish *criollo*, originates in Portuguese. The form found there, *crioulo*, meant 'home-bred' (Portuguese *criar* 'breed' from Latin *creare*, 'create'). Creole was viewed, in the period of Portuguese conquests, as the language spoken by slaves who had been brought up in the household rather than captured as adults.

Pidgin English is often called *Broken English*. Creoles are sometimes known, especially in the Caribbean, by the French name *Patois* 'peasant dialect'.

The tongues of European trade

The English-speaking traders who explored the Atlantic and Indian Ocean routes from about 1600 onwards found pidgin PORTUGUESE and SPANISH spoken. As the English prospered, their Amerindian, African and Asian trading partners and employees soon realised that they could make themselves better understood by using English words. Thus the oldest recorded English pidgins of the Atlantic and Indian Oceans are not really new inventions: they consist, in essence, of the Portuguese West African pidgin, 're-lexified' with English words, but still retaining Portuguese features – and still showing, in simplified form, a structure reminiscent of the NIGER-CONGO LANGUAGES of West Africa, notably WOLOF, one of the first with which the Portuguese had come into contact.

These first English pidgins, carried around the world by seamen, migrant workers and slaves, are the linguistic ancestors of several major languages of today – and of a host of less-known creoles and pidgins of small, isolated, half-forgotten communities in every continent.

Languages of the Atlantic

In and around the Caribbean, the destination of so many African slaves in the 17th and 18th centuries, there grew up a family of English creole languages, each of which has had its own history, though all are closely related. None of them is an official language, but the creoles of Jamaica (see separate entry), several Antillean islands, the Bahamas, Belize, Guyana and Suriname are indispensable languages of national communication and cohesion. In most countries where English is official, speakers nowadays learn to adopt a range of spoken styles from 'standard English', or a local standard, all the way to 'pure creole'.

First there are JAMAICAN CREOLE and the languages of the English-speaking Leeward Islands, Antigua, St Kitts, Nevis, Montserrat, Anguilla, Barbuda, and of the Virgin Islands. Barbadian or Bajan Creole played a pivotal role in the development of these creoles in the 17th century, though it is now scarcely spoken on Barbados itself. Related English creoles are also spoken in the Dutch Windward Islands, and in Tobago, St Vincent and Grenada. Trinidad Creole, though linked with Barbadian, was also influenced by BHOJPURI (better known locally as Hindi) and by KRIO, because of later migration from India and Sierra Leone. Echoes of all these Caribbean creoles can be heard in the Black English of modern Britain.

To the south-east, Guyanese Creole, like that of Trinidad, has been influenced by 19th- and 20th-century labour migration as well as by the earlier slave trade. Again, strong recent links exist with the Krio of Sierra Leone and with Bhojpuri.

To the west, in the former British Honduras, Belize Creole is the mother tongue for most of the urban population of Belize City, while country people, whose first languages are Spanish or Mayan languages or Black Carib, learn the creole as their second language. An offshoot, Miskito Coast Creole, has long been spoken in harbour towns and villages along the north coast of Nicaragua. There is a Panama variety of Caribbean Creole, with as many as 100,000 speakers, in Panama City, Colón and Bocas del Toro.

To the north, Bahamas Creole is sharply distinct from Caribbean varieties – for good historical reasons. The Bahamas were long ago united administratively with the British colony of Carolina; later, after the American Revolution, Loyalists from the southern states, and their slaves, settled in the Bahamas in large numbers. This explains why Bahamas Creole has a close relative in *Gullah* or Geechee, still spoken in the Sea Islands and along the southeastern United States coastline from Florida to North Carolina. Otherwise, the 18th- and 19th-century creole of slave communities in the southern United States does not survive directly – but it is not forgotten. Modern Black English of the United States retains some features of this language, as does Liberian Creole.

Church Creole in Suriname

The Moravian Missionaries started about 1780 to use creole in church. They translated the Bible, compiled a hymnbook, published a monthly paper and countless pamphlets, tracts and edifying stories.

Foreign missionaries have not always been gifted language learners. Their pronunciation of creole was not always correct. They were not corrected by their congregations, but imitated. So, by institutionalised mispronunciation, they created a creole variety which was then imitated by others as more fashionable and given superior status, because it was used on solemn occasions by people belonging to the former upper caste.

Church Creole	English	Common Creole (Sranan)
pikin	small child	pikiế
belə	belly	bere
tem	time	tế
helpi	help	yepi
rœstə	rest	lostu

After Jan Voorhoeve in *Pidginization and creolization of languages* ed. Dell Hymes
(Cambridge: Cambridge University Press, 1971) pp. 310–13

The lingua franca of Suriname

Sranan (sometimes called Taki-Taki or Negro-English) is the lingua franca of Suriname, the mother tongue of most city and coast dwellers, known as a second language to many who live upriver.

Sranan is important in the study of Creole history because DUTCH, not English, is the standard language of Suriname. Sranan is not continually subject to the normalising influence of a closely related standard, as Jamaican Creole is. Thus to the creole specialist Sranan seems 'conservative', even 'archaic', in comparison with its Caribbean relatives.

The English ruled Suriname for only twenty-six years, from 1651 to 1667. Most of the English planters left soon afterwards, and took many of their slaves with them. Why has an English creole remained the normal, everyday language of the country ever since? For the next two hundred years, Dutch slave-owning society maintained a caste system charged with symbolism. Slaves were not allowed to learn Dutch (or to wear shoes, or to become Christians). The English creole, known to existing slaves, learnt by new arrivals from Africa, was thus necessary to everyone.

At emancipation in 1863 the same symbolism determined policy. Since all were free, all must now use Dutch. Creole (Sranan) was outlawed in education: even the Moravian missionaries, who alone had used it in their mission schools, had to turn to Dutch. The result was a bilingual society in which almost everyone learnt to switch at will from Dutch to Sranan and back. In the 1960s there was a move to raise the status of Sranan, and non-religious literature appeared in it for the first time.

The Bush Negro languages

Sranan is not the only creole of Suriname. Already in the 17th century runaway slaves had set up tribal societies in the river valleys of the interior. The first, probably, were the *Saramaccan* and *Matuari*. Having broken away from the pidgin-creole continuum so early, these two languages are now quite distinct from their relatives. They have tones, like many West African languages; their vocabulary seems almost evenly split between Portuguese and English words, but there is also a strong KONGO element. These are important clues to the nature of the 17th-century Atlantic pidgins.

Historical facts are scarce, but linguistic evidence suggests that the Ndjuka, Paramaccan and Boni (or Aluku) tribes were founded by runaways of a slightly later period, perhaps the early 18th century, when English was closer to replacing Portuguese on Atlantic trade routes. Of these three languages *Ndjuka* (also called Aukaans) is the best known. It has a syllabic writing system related to those of some West African languages – compelling evidence that some of the escaped slaves who formed this tribe had been literate in an African language before their enslavement.

Song and story in Saramaccan

In Saramaccan storytelling, often performed at funerals, songs are interspersed with the prose narrative. Many are at least partly in an obscure poetic language quite different from that of everyday life. Here, in the course of storytelling, a woman character is introduced to the audience:

Agangaai, i sá kíi m. Mhmm.	Agangaai, you could kill me –
Agangaai, i sá kíi m. Mhmm.	Agangaai, you could kill me –
Di hánse fa a du m te. Mhmm.	Your beauty does something to me,
A du m te mooi ta yáa sónu. Mhmm.	It's enough to make the sun come up,
Agangaai, i sá kíi m. Mhmm.	Agangaai, you could kill me!

Richard Price, Sally Price, *Two evenings in Saramaka* (Chicago: University of Chicago Press, 1991) p. 87

African English creoles and pidgins

Liberian English has well over a million speakers. It has also been called Merico, Americo, Brokes, Waterside English and Settler English. Spoken especially in Monrovia and other coastal towns, it was implanted on the creation of this 'homeland' for freed American slaves; it naturally derives from the creole of United States slaves of the 19th century, though it shows signs of strong influence from the local languages of Liberia.

Liberian English is thus distinct from the KRIO of Sierra Leone and from the other English creoles and pidgins of West Africa. These form a sequence along the Guinea coast: they include the old-fashioned 'Krio' used by the Kru fishermen (originating in eastern Liberia) in their work and trade along the West African coast. They also include the West African Pidgin English or Nigerian Pidgin English that is a major lingua franca of Nigeria.

As elsewhere, creole languages are more likely to achieve stability and permanence if no longer in competition with the standard language from which they derive. An example is Fernando Po Creole or 'Porto Talk', which derives from Freetown Krio and West Indian Creole English, brought by immigrants who came to Fernando Po in 1830. It is now conservative, even archaic, and easily distinguished from its neighbours, since Spanish, not English, is the standard language of Fernando Po. A better known example is KAMTOK, now a language of communication in much of Cameroun.

Indian Ocean and Pacific Ocean pidgins

Until recently, the various kinds of pidgin English of the Indian and Pacific Oceans had not developed into stable creoles. There have been many of them: the so-called *Babu English* of northern British India, supposed typical of Indian clerks writing English; the Madras Pidgin or *Butler-English* of the south, English-based but much influenced by Dravidian languages; the famous *China Coast Pidgin*, now perhaps spoken only in Taiwan and on Nauru, surviving till recently in Hong Kong and among old people in Shanghai, once so essential that Chinese tradesmen and servants in some cities could not avoid learning it if they were to deal effectively with non-Chinese, and that manuals were compiled to teach it to 19th-century Englishwomen.

Far more recent, but already lost in history, are the Bamboo English of the Korean War, the Vietnam Pidgin of the Vietnamese War, and the so-called Japanese Pidgin of the 19th century and of the Second World War.

Of more lasting importance than any of these is *Beach-la-mar*, the English pidgin of the 19th-century western Pacific, with its modern descendants: Melanesian Pidgin English, TOK PISIN, BI-SLAMA and Solomons Pidgin or *Pijin*. Another offshoot was the older Australian Pidgin and the *Kriol* to which some Australian speakers are moving. Hawaiian English, with its 500,000 speakers, is usually considered a creole: it shows the influence of Chinese, Japanese, Hawaiian, Portuguese and various Philippine languages. The language of Pitcairn Island, with its offshoot on Norfolk Island, is an English-based creole with Tahitian elements – the inhabitants are descendants of the mutineers of HMS *Bounty*.

Many of the English creoles and pidgins are on record in travellers' memoirs, old and new. But these records depend on authors' varying linguistic skills. Some, clearly, heard incorrectly, or tried to reconstruct conversations heard long ago and did so inaccurately. Few if any had a clear understanding of the structure of the language they were attempting to write. Naturally, novelists' 'Pidgin English' is to be taken with a pinch of salt. The same goes for novel and film versions of American Indian Pidgin English, a language first recorded in 1641 and now extinct.

ESKIMO-ALEUT LANGUAGES

From Greenland across the Arctic edge of America to eastern Siberia, speakers of Eskimo-Aleut languages have established cultures perfectly adapted to a harsh environment. The languages have no close relatives elsewhere. The family includes the eastern Eskimo language INUIT, the language of western Alaska and St Lawrence Island known as Yupik (18,000 speakers), and a much more distant connection, Aleut.

Eskimo, by way of French *Esquimaux*, is a loanword from an Athabaskan language of North America. The word meant 'stranger'.

The Arctic: Eskimo-Aleut languages

Chaplino

Northern Alaska dialects

I N U

Y U P I K

Central Yupik

Western Canada d

ALEUT

Pacific Yupik
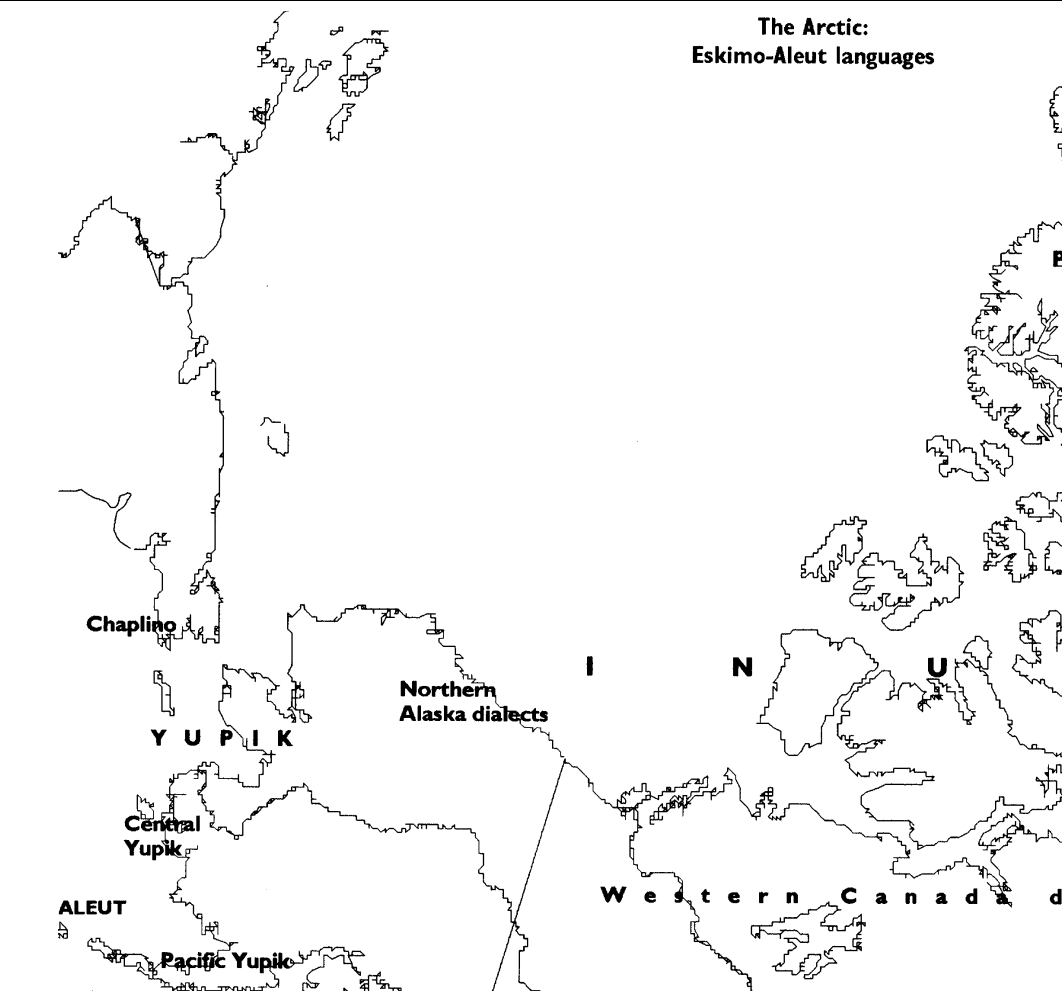
Aleut is now spoken by only 500 or fewer people in the Aleutian and Pribilof Islands, Alaska (especially at Atka), and perhaps by a few in the Commander Islands, Siberia. In the 19th century, when Alaska was still Russian territory and Aleut speakers were much more numerous, some were literate. Aleut was written in the Cyrillic alphabet with extra characters. This tradition is dead, and the language itself dying, both in Siberia and in Alaska.

Yupik is a group of Eskimo dialects of western Alaska and the eastern tip of Siberia. In Siberia and in St Lawrence Island, Alaska, the majority dialect is Chaplino or Ungazik (1,500 speakers). Schoolbooks, grammars and dictionaries used to be published in Siberia in Chaplino: at first in a phonetic script; after 1936, for political reasons, in Cyrillic. A few texts in Roman script have been issued in St Lawrence Island. On the Alaskan mainland the majority dialect is Central Yupik, which has a tradition of religious and educational publications. Pacific Yupik (sometimes confusingly called Aleut) is spoken by 2,000 or fewer speakers especially at English Bay, Alaska.

In the 18th and 19th centuries Eskimos developed pidgin versions of their languages to help speakers of English, Russian, Danish and NA-DENÉ LANGUAGES to understand them. More recently research on Eskimo-Aleut languages has taken place in Russia, the United States and Canada, and all three governments have presided over a rapid decline in the daily use of these inconveniently international languages of tiny minorities. In Greenland, INUIT has a fuller place in official and everyday life. Even there, the difficult accommodation between Eskimo and 'Western' ways of life results in a gradual reduction in the number of Inuit speakers.
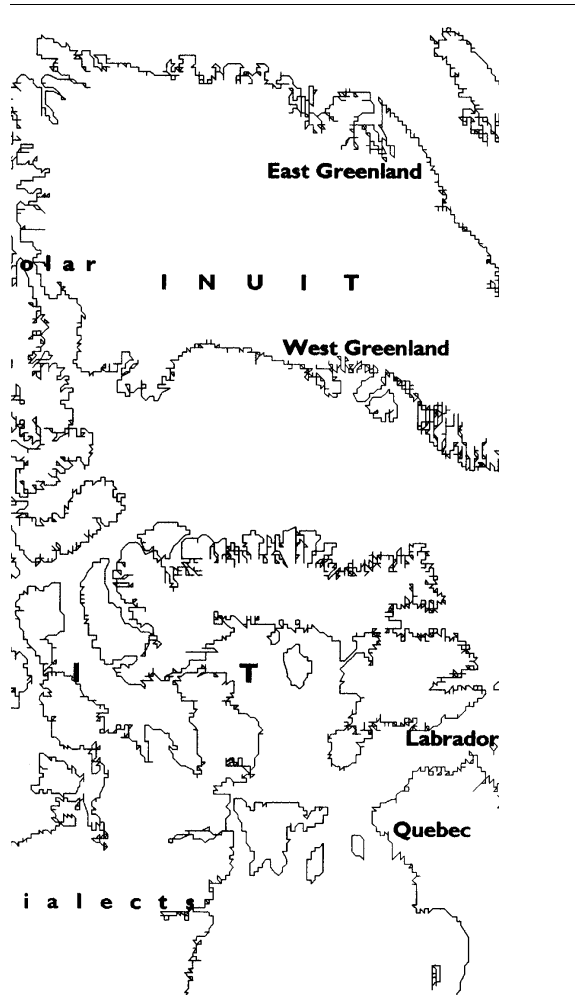
Estonian

1,100,000 SPEAKERS

Estonia

With Finnish, Estonian is one of the twin Balto-Finnic languages divided by the Gulf of Finland. They belong to the wider family of URALIC LANGUAGES.

The Estonians, *Eesti*, may possibly be the *Aestii* of the Latin writer Tacitus (1st century AD) and are certainly the *Eistneskr* of the Norwegian poet Þjóðólfr's *Ynglingatal* (11th century). The meaning of the word is unknown. Traditionally, Estonians called themselves *maa rahvas*, 'people of the country'.

Estonia was apparently warlike and independent at the end of the first millennium AD. After Danish raids in the 12th century, the country was conquered by German knights in the 13th. Estonia was subject to German landowners – and generally under Swedish or Russian government – till the late 19th century. Thus, while Finnish was influenced by Swedish, Estonian was for many centuries under strong German influence. Estonian bulges with German loanwords – including many from the Low German of the Baltic trading ports (see box at FINNISH).

There are a few medieval records of the language, such as the Estonian personal names in the *Chronicle* of Henry of Livonia. The first known printed book was the bilingual German-Estonian catechism, by S. Wanradt and J. Koell, which appeared (in far-off Wittenberg) in 1535. No complete copy survives. An Estonian grammar was printed in German, for the use of priests, in 1637.

As a serf population, few Estonians received education before the 19th century. The emergence of Estonian as a literary language can be dated by the publication of a language magazine,

Beiträge zur genauern Kenntniss der ehstnischen Sprache, 'Studies towards better knowledge of the Estonian language' edited 1813–32 by the clergyman Heinrich Rosenplänter – and by the *Ehstnische Litterarische Gesellschaft*, 'Estonian Literary Society' which was founded in Kuressaare in 1817. A sign of the continuing domination of German, even under Russian rule, is that both these institutions had German names. It was only around 1840 that Estonians themselves began to predominate as authors of works in and about their language.

Estonian literature now flourished, notably with F.R. Kreutzwald's *Kalevipoeg* (1857–61), a literary epic emulating the Finnish *Kalevala*. It is based on Estonian prose legends and on the metres of oral lyric poetry: these resources were needed because Kreutzwald found that unfortunately there were no Estonian oral epics to draw on!

Estonia was independent from 1918 to 1940, and has been independent again since the breakup of the Soviet Union. In this century literary Estonian has been almost rebuilt, many loanwords eliminated, and others replaced by dialect words or Finnish loans. The language is now emerging once more from a period of heavy Russian influence.

Standard Estonian is based on the midland dialect of North Estonian, spoken in Tallinn and its hinterland (see map at FINNISH). It has four (some linguists say three) possible lengths for both vowel and consonant sounds. Length variations often change the meaning completely and are difficult for foreign learners to distinguish. The four lengths are not fully marked in normal spelling, but are specially written in these

examples: *sadaa* 'hundred', *saaadaa* 'send!', *tahab saaaada* 'want to get'; *kanu* 'of hens', *selle kannu* 'this jug's', *seda kannnu* 'of this jug'. Nouns have 14 cases, and verbs change form to show person, tense, mood and the active/passive opposition.

Examples based on Aimo Turunen, 'The Balto-Finnic languages' in *The Uralic languages: description, history and foreign influences* ed. Denis Sinor (Leiden: Brill, 1988) pp. 58–83

Numerals in Estonian, Finnish and SAMI

	Estonian	Finnish	Sami
1	üks	yksi	åkta
2	kaks	kaksi	guokte
3	kolm	kolme	gålbma
4	neli	neljä	njællje
5	viis	viisi	vitta
6	kuus	kuusi	gutta
7	seitse	seitsemän	čieža
8	kaheksa	kahdeksan	gávci
9	üheksa	yhdeksän	åvci
10	kümme	kymmenen	lågi

The name of Estonia's capital, *Tallinn*, is a reminder of the country's history of foreign domination. Originally it was *Taani linn*, 'Danes' town'.

South Estonian

'In the 17th to 19th centuries South and North Estonian emerged as distinct languages, and there is some printed literature in South Estonian. The New Testament appeared in South Estonian in 1686, in North Estonian only in 1715. Even nowadays, if south Estonian youngsters return from study with a standard (northern) accent, they may be greeted with the rebuke *Kas ma' su tuuperäst kuuli saadi, et sa' mul sääl joba rääkmä nakkat?* ''Do you think I sent you to school for you to start *talking*?'' '

From Alo Raun, Andres Saareste,
Introduction to Estonian linguistics
(Wiesbaden: Harrassowitz, 1965) p. 82

ETHIOPIC

Classical language of Eritrea and Ethiopia

One of the SEMITIC LANGUAGES, Ethiopic is a relative of Arabic and Hebrew, but most closely related to the South Arabian languages and to the modern South Semitic languages of Ethiopia, notably AMHARIC. Ethiopic is the traditional language of the Ethiopian Christian church.

> Ethiopic is often called *Ge'ez*, the local term for a letter of the Ethiopic script – and thus for the classical written language of the country.

South Semitic languages have gradually spread southwards across Ethiopia from the Red Sea coastal region. Ethiopic is recorded in stone inscriptions from the 3rd century onwards, linked with the kingdom of Axum in what is now Eritrea. Ever since the conversion of King Ezana in the 4th century, Axum and Ethiopia have been a stronghold of Christianity, though long isolated from the rest of Christendom by the Islamic conquests to the north.

Ethiopic literature began with translations of the Bible and of Christian texts of all kinds: early translations were made from Greek, but many later ones came from Arabic versions, and in many cases the intermediate texts do not survive. So Ethiopic preserves a record of early Arabic Christianity, which has long since been overlaid by Islam. Original writings in Ethiopic include important histories, such as the legendary *Kebra Negast*, 'Glory of the Kings', which tells of the Queen of Sheba's visit to Solomon, and of their child, Menelik, mythical founder of Axum.

As a spoken language Ethiopic gave way to later local standards which developed into modern TIGRINYA, Tigre and Amharic. As an official language, and especially as the language of church worship, Ethiopic has lasted far longer. The classic period of Ethiopic literature was the 13th to 17th centuries. *Weddase Mariam* 'Praise of Mary', a 14th-century poetic collection, accompanies the Psalms in the Ethiopic canon. The law code *Fetha Negast* 'Justice of the Kings' dates to the 17th century.

Ethiopic was still the language of literature, religion and official documents in the 19th century, but Amharic has now invaded all these fields.

The Ethiopic alphabet

The distinctive Ethiopic script was developed from an alphabet used for an early South Arabian language. In this script, which is written from left to right, the original consonant signs are modified in a fixed pattern to indicate one of seven following vowels. Thus, unlike the other writing systems used for Semitic languages, that of Ethiopic represents the sounds of the language fairly fully. The 'first order', shown in the box, is of consonants followed by the vowel ä.

Ethiopic script: the 'first order'

ሀ ለ ሐ መ ሠ ረ ሰ ቀ በ ተ ኀ ነ አ ከ ወ ዐ ዘ ኸ የ ደ ጠ ጸ ፀ ፈ ፐ

h l ḥ m š r s k' b t ḫ n 'ʾ k w ä z y d g t' p' c z f p

Ethiopic script: 'h' with the seven vowels

ሀ ሁ ሂ ሃ ሄ ህ ሆ

hä hu hi ha he hǝ ho

Ewe and fon

4,000,000 SPEAKERS

Togo, Benin, Ghana

Ewe and Fon are the best-known names for one of the Kwa group of NIGER-CONGO LANGUAGES. Linguists call it Gbe. Dialects of this language are spoken in south-eastern Ghana – east of Lake Volta – and across southern Togo and Benin to south-western Nigeria.

Tado and Notse in Togo, sites with impressive ancient earthworks, are said to be the ancestral homes of Gbe-speaking peoples, from which they spread southwards and to east and west. They had reached the Atlantic coast, and begun to come into contact with European traders, over five hundred years ago. A Gbe wordlist (the language is called *Mina*) was put together for Portuguese traders in 1480.

Gbe speakers later had no overarching political systems. Their one hundred and twenty small chieftainships were in British, German and French possession at the end of the 19th century.

The form of Gbe spoken on the Togo coast, at Anécho and Lomé, was perhaps already somewhat distinct from those of the hinterland because its population had been built up by 18th-century migrants from southern Ghana. Gbe, in the form of the Gẽ-Anglo dialects of the so-called Slave Coast, became a lingua franca of the inland trading routes as they were developed by the German administration of Togo in the late 19th century. Along the railway line inland from Lomé, Ewe was the usual means of communication, spreading to become the trade language of the major towns Atakpamé and Sokodé. Its expansion westwards towards the Volta valley was promoted in German times by both missionaries and administrators, who actively discouraged trading contacts with the Twi (AKAN) speakers of the British-dominated Gold Coast.

The modern standards

The Anglo dialect of coastal Togo is now the standard form not only for Togo itself but also for the Ewe-speaking districts of Ghana, where – spoken with a local accent – it is used in broadcasting, in churches and in schools. In Togo, too, there is an Ewe press in the Anglo dialect. At home, local dialects continue to be used.

The most widespread dialect of Benin, the one that has become an official and literary language there, is Fõ. Within this Gũ is the most used sub-dialect, heard at the capital, Porto Novo, and along the coast into western Nigeria.

Typical of Gbe phonetics are the labiovelar sounds *kp*, *gb*, *ngm* – the unfamiliarity of these sounds explains the wide variation in European attempts to spell Gbe words. The language has four tones and is largely monosyllabic: only compounds and foreign loanwords have more than one syllable. Unusually for African languages, tone in Gbe is 'lexical' only – it distinguishes between words of different meanings, not between different forms of the same word. The suffix *–wo* marks the plural of nouns and is also the third person plural pronoun 'they'. In Gbe, *ètsò* means both 'yesterday' and 'tomorrow' – but the context makes the meaning clear.

Gbe is the local term for 'language'. It was adopted by linguists in 1980 as an overall term because neither 'Ewe' nor 'Fon' was acceptable to all speakers of the language. *Ewe (Ephe* in German writings, *Eve* in French) is the traditional name for the standard form which is closest to *Anglo* and *Gẽ*: its proper local name is *Evegbe*, while the Gẽ dialect is

properly called *Gẽgbe* and the Fon dialect *Fõgbe*.

'The Ewe are called *Bubutubi* by the Anglo, *Benigbe* by the Avatime, *Bayikpe* by the Santrokofi, *Manyigbe* by the Akpafu, *Bowli* or *Ayigbe* by the Gã and Dangme. In the west they are called *Hua* by the Akan. The central and eastern Ewe, especially those near the coast, are called *Popo* by the Yoruba and *Ima* by the Kposo. Other names found in older literature are *Eibe*, *Krepi* and *Krepe*' (Madeline Manoukian, *The Ewe-speaking people of Togoland and the Gold Coast* (London: International African Institute, 1952) p. 10).

The earliest extended text, and the first printed book, in any language of West Africa is a bilingual Spanish–Gbe catechism printed in Madrid in 1658: *Doctrina christiana . . . en nuestra idioma Español y en la lengua Arda*, 'Christian teaching in our Spanish language and in the Arda tongue'. *Arda* is the name of a now-forgotten Gbe-speaking kingdom that then ruled the coast hereabouts.

This milestone in African language history 'was lost for nearly three centuries, partly because the surviving copy was catalogued under "languages of America" . . .' (P.E.H. Hair in *Language and history in Africa* ed. D. Dalby (London: Cass, 1970) p. 53. There is indeed an obscure and now extinct Amerindian language of Peru with the same name *Arda*.

The Gbe dialects

Ewe (or *Vhe*) is the usual name for the western dialect group, spoken in Ghana and Togo. The most important of these, the one from which a written form was developed, is *Anglo* (or Anlo or Awuna), spoken along the coast between the mouth of the Volta and Lomé. *Gẽ* (or Gen or Guin or Gain or Mina) is the name for the Togo group of dialects. *Adja* is the dialect spoken from Natja eastwards. There are about 1,200,000 speakers in Togo and 1,500,000 in Ghana.

Further east, *Fon* (or Fõ or Dahoméen) is the dialect group from which the standard language used in Benin originated. Local dialects include *Gũ* (or Goun or Egun), spoken in the immediate neighbourhood of Porto Novo, capital of Benin, and the north-eastern dialect *Mahi*. There are 1,000,000 speakers of Fon as a mother tongue; at least half a million more speak it as a second language.

As can be seen from the many alternative spellings, the sounds of this essentially monosyllabic, four-tone language, with its complex initial consonants and its nasal vowels, are difficult at first for western European language speakers to catch.

Numerals in Ewe and Fon		
Ewe		**Fon**
ḍékà	1	ḍòkpó
èvè	2	we
ètɔ̀n	3	àtɔ̀n
ènĕ	4	ɛ̀nɛ̀
àtón	5	àtɔ́n
àdé	6	àyìzɛ́n
àdré	7	tɛnwɛ
ènyí	8	tântɔn
ènyíḍé, àsièkè	9	tɛnnɛ
èwó	10	wǒ

In Fon, '3' and '5' differ only by the tone of the second syllable.

M. Malherbe, *Les langues de l'humanité* (Paris: Laffont, 1995) pp. 829, 855 and other sources

West Africa: Gbe
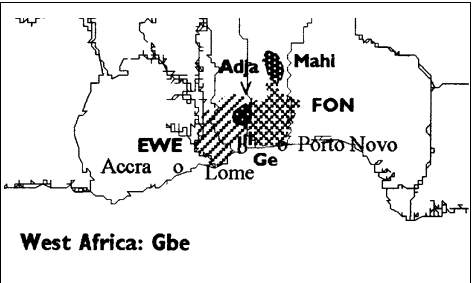
Ewondo, Bulu and Fang

1,350,000 SPEAKERS

Cameroun, Equatorial Guinea

Ewondo and its relatives are BANTU LANGUAGES spoken in the valleys of western Cameroun, Gabon and Equatorial Guinea. Speakers of early western Bantu languages settled here well over three thousand years ago. This particular group of dialects probably spread to its present inland locations in the 14th and 15th centuries. With colonial political frontiers, several closely related languages emerged as local standards. Speakers of any one can quite easily understand the others.

Together the languages are sometimes called the 'Sanaga-Ntem group' after the rivers that define their territory to north and south. But *Yaoundé* (German spelling *Jaunde*) was the name that the Germans learnt from coast-dwellers for the inland region where they planned to establish their new trading post. Yaoundé has remained the official name of what has now become the capital of independent Cameroun. It is really a 'foreign' version of the name that the people of the district give to themselves and their language, which is *Ewondo*.

Fang is locally known as *Fangwe* or *Mpangwe*. This last form gave rise to the older French name – utterly different at first sight – *Pahouin*.

Bulu, the language of Ntem, Dja and Lobo divisions of South Province, Cameroun, was increasingly important in the early 20th century, when Cameroun was a German colony. The American Presbyterian missionaries in southern Cameroun published a Bible translation and textbooks in the dialect, and used it in their schools: speakers of other local languages found that Bulu was the route to education. Some resisted this development – the Ngumba, for example, who left the Presbyterian church because the missionaries would not use their language. However, some of those whose languages were closest to Bulu, such as the section of the Fang that live in Cameroun, accepted Bulu from its use in school and church as their standard language. For these it is now a mother tongue.

Bulu as a second language has given way before the spread of French and Ewondo.

Ewondo or Yaoundé is the language of the neighbourhood of the capital of Cameroun, Yaoundé, which began its history as the trading post from which the Germans advanced into inland Cameroun. The Ewondo were at hand to be employed as labourers, porters and soldiers. They and their language spread with the growth of the colony: when French replaced German as the language of government in 1918, Ewondo was already familiar to many as a second language throughout southern Cameroun, though not on the coast, where DUALA and KAMTOK were the lingua francas.

A pidgin language with its roots in Bulu and Ewondo has rapidly developed to permit easy communication among migrant workers, rail-waymen and truck drivers in Cameroun. This is known as *Ewondo Populaire*, *Bulu beDiliva* ('Bulu of the drivers') or – by linguists with a liking for classification – *Pidgin A70*, because

A70 is one way of designating the Ewondo group of Bantu languages. This pidgin began before 1930, when the construction of the railway inland to Yaoundé attracted migrant labour, and it is now much used in market-places and on transport routes throughout the country. It can also be heard in Gabon and Congo (Brazzaville). Ewondo Populaire is still a pidgin or jargon of men's employment, little used by women, children or old people.

Fang, like Bulu and Ewondo, spread in the 19th century, when, with coastward migrations, it became the language of the Gabon and Ogooué estuaries at the time when the Spanish and French interests in the region were first being asserted. Fang had clearly followed the local trade routes that had developed between coast and hinterland, in a region where European colonisation did not immediately follow European trade. Fang was the language of Catholic missions, in contrast with the Presbyterian use of Bulu.

So Fang, earlier spoken in inland southern Cameroun and northern Gabon, is now also the major language of Mbini, the mainland territory of Equatorial Guinea. On independence in 1969 it was abortively named the sole official language of the whole country, temporarily supplanting Spanish, much to the dissatisfaction of inhabitants of the island of Annobon, where Spanish is the language of culture and Fang was little known.

Ewondo, Duala and some relatives: the map

Ewondo, of the Yaoundé region, has 575,000 speakers and is one of the national languages of Cameroun.

Bulu is the major language of South Province, Cameroun, with about 175,000 speakers. *Beti* and *Eton* belong to the same dialect continuum: they are languages of central Cameroun. Ewondo is the language of education and administration here.

Fang and *Ntumu*, with 525,000 speakers, spread towards the coast in the 19th century where Fang is the principal language of Mbini (formerly Rio Muni or Spanish Guinea, a Spanish possession from 1778 to 1969). For speakers in Gabon and Cameroun Ewondo serves as the standard language.

DUALA, a distant relative of the Ewondo-Bulu-Fang group, is the language of the old trading city of Duala, once important regionally but now restricted to Duala and its neighbourhood. It has 80,000 speakers.

The island of Bioko (Fernando Póo), the other half of Equatorial Guinea, has a language of its own, *Bubi* (once spelt *Booby*: 20,000 speakers), representing an extremely early Bantu migration. The British naval base of Clarence, occupied 1827–44, brought rapid migration of West African peoples: as a result an English creole is the main language of coastal regions. On independence from Spain in 1969 there was an unpopular attempt to impose Fang as sole official language.

	Ewondo	Bulu	Fang	Bubi
		Numerals in Ewondo, Bulu, Fang and Bubi		
1	fóg	fok	fokh; -boré	-de
2	be	-bae	-bè	-ba
3	lá	-lal	-lal	-cha
4	nyie	nyin	-né	-ñe
5	tan	tan	-tan	-chio
6	saman	saman	-samé	lade
7	zamgbál	zangbwal	nžañ gwal	la ba
8	moom	mwôm	oñwam	la cha
9	ebul	ebul	ébul	la am
10	awóm	awôm	awôm	lēño

Information from Jeanne d'Arc Lacoin and others in
M. Malherbe, *Les langues de l'humanité* (Paris: Laffont, 1995)

Noun classes in Bulu

As in other Bantu languages, nouns in Ewondo, Bulu and Fang belong to classes marked by prefixes:

	Singular			Plural	
Class		**Example**	**Class**		**Example**
I	m-ongo	child	VII	b-ongo	children
V	əm-vu	dog		bə-mvu	dogs
	ən-jo	scissors	IX	mə-njo	scissors
III	ə-ci	egg		mə-ci	eggs
II	n-tomba	sheep	VIII	min-tomba	sheep
IV	e-mvang	calabash	X	bi-mbang	calabashes
VI	o-non	bird	XI	a-non	birds

From Yashutoshi Yukawa, 'A tonological study of Bulu verbs' in
Studies in Cameroonian and Zairean languages (Tokyo: ILCAA, 1992) pp. 67–93

West Africa:
Ewondo, Bulu and Fang
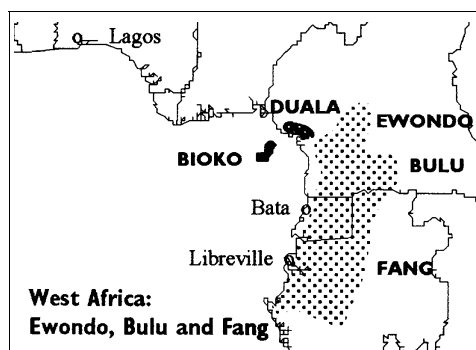
Faroese

50,000 SPEAKERS

Faroe Islands

Except for a few communities of Celtic monks, Vikings from south-western Norway were the first inhabitants of the Faroe Islands (see map at NORWEGIAN), which they colonised in the 9th and 10th centuries. Faroese is thus a direct descendant of OLD NORSE, and it does still contain a few Celtic words. Like Icelandic, it has undergone very little influence from any other language – except Danish, since the Faroes were long treated as a Danish colony and are still linked politically to Denmark.

The name of the Faroes, *Føroyar*, means 'sheep islands'.

Although there were only about 9,000 Faroese in the 1850s, when the movement for national and linguistic self-assertion began in the islands, it has been far more successful than many similar movements in Europe. It is now the language of the press, of a very small but thriving publishing industry, and of local government and education. Danish is taught as a second language.

For a table of numerals see SWEDISH.

FIJIAN LANGUAGES

350,000 SPEAKERS

Fiji

Fijian is the national language of the Fiji island group. Because of massive Indian migration to the islands in the 19th century under British rule, Fijian is spoken by less than half the population. It is a Polynesian language, belonging to the Oceanic branch of AUSTRONESIAN LANGUAGES.

There is such great dialect variation in Fiji, which was traditionally divided into small chiefdoms, that Fijian is usually considered to be at least two languages. The development of a single literary language is due to early 19th-century missionaries. They arrived by way of Tonga and first learnt their Fijian in the Lau islands, close to Tonga. However, as they prepared to work on the larger islands, they found it best to start from Bau, the most powerful chiefdom at the time. The missionary language, Old High Fijian, used in church and in the Fijian Bible translation, was thus *Bauan* with some residual *Lauan* features. Its style was also noticeably influenced by the missionaries' mother tongue, English.

The long term result is that Bauan became the prestige dialect all over Fiji, and it remains so to this day. It is used in literature and the media, in religion, and very largely in schools. By now most if not all Fijians are able to understand Bauan, even if they cannot speak it. It is also the basis of the Colloquial Fijian that is the everyday speech of town-dwellers.

Tongans ruled much of Fiji in the 19th century: British rule followed in 1874 and lasted until independence in 1970. Fijian has borrowed extensively from both Tongan and English. Western Fijian and other country dialects are now gaining loanwords from Standard Fijian.

The local 'Hindustani' language of mainly BHOJPURI origin, spoken by half the population, has also supplied some loanwords.

> Pidgin Fijian has grown up as the language of markets and village centres, used for daily contact among speakers of Fijian, 'Fijian Hindustani' and Chinese. It is highly simplified: where Bauan has 135 distinct forms of pronouns, Pidgin Fijian has only six.

The first nine numerals in Fijian are: *dua, rua, tolu, vā, lima, ono, vitu, walu, dhiwa*. '10' is *tini* in Eastern Fijian, *chini* in Western Fijian.

Based on Paul Geraghty, 'Eastern Fijian' in *Comparative Austronesian dictionary* ed. Darrell P. Tryon (Berlin: Mouton De Gruyter, 1995–) pt 1 pp. 919–23 and other sources

Languages of Fiji

Eastern Fijian is a group of dialects spoken in Vanua Levu, the eastern half of Viti Levu, and most of the smaller islands. *Bauan*, the dialect of a small island off Viti Levu, is the basis of standard written Fijian.

Western Fijian (50,000 speakers) is a group of dialects of western Viti Levu. Western and Eastern Fijian are not mutually intelligible.

The other major language of Fiji is *Fijian Hindustani* (see BHOJPURI), spoken by the descendants of labourers imported by the British rulers of Fiji in the 19th century. Many still work in the sugar-cane plantations. Most also speak Fijian or Pidgin Fijian. *Chinese, Rotuman* and *English* are also spoken in Fiji by small communities.

Rotuman, related to Fijian, has about 8,500 speakers on the nearby island of Rotuma.

Western Pacific

Rotuman

Vanua Levu

FIJIAN

Bau

Viti Levu

Lau Group
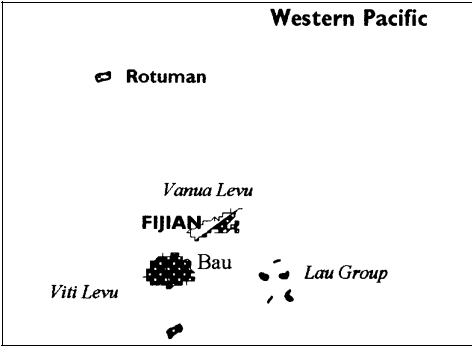
FINNISH

5,000,000 SPEAKERS

Finland, Sweden, Russia

Finnish belongs to the URALIC LANGUAGES and, with Swedish, is one of the two national languages of Finland. It has two close relatives, spoken by much smaller numbers in north-west Russia: Karelian and Veps. Otherwise, the nearest relative of Finnish is Estonian. These two languages, spoken in neighbouring countries on either side of the Gulf of Finland, are almost similar enough to be mutually comprehensible (for a table of numerals see ESTONIAN).

Finn was originally the name for the SAMI or Lapps. The country that was seen as their homeland came to be called *Finnland* by Scandinavians. Its settled inhabitants, in distinction from the nomadic Sami, were then logically named *Finnlendingar* 'Finlanders' in Old Norse – whence the usual modern name *Finnish* for their language. In Finnish itself the country and language are called *Suomi*. This derives from older Swedish *somi* 'mass', a reminder that Swedes once classed Finlanders disparagingly as 'the masses, the plebs'.

In the first millennium BC speakers of proto-Finnic were already settled around the southern Baltic shores – and certainly extended much further east into Russian lands, as surviving Finnish place names show.

Karelian, which may be described as a sequence of eastern Finnish dialects, is actually recorded earlier than Finnish itself (see box). The differences between present-day Finnish and Karelian come largely from their contrasting history. Finnish has been influenced by Sweden, by western Europe and by Protestantism; Karelian by Russian and by the Russian Orthodox Church.

Finnish written literature did not begin to develop till the Reformation, with Michael Agricola's Finnish translation of the New Testament in 1548. As Bishop of Åbo (Finnish *Turku*), Agricola based his written Finnish on the south-western dialect, relatively close to Estonian. A complete Finnish Bible appeared in 1642 at Stockholm.

Meanwhile, long a part of the Kingdom of Sweden, Finland began to lose its linguistic autonomy in the 17th century, under King Gustavus Adolphus. Centralisation led to the dominance of Swedish in public life and culture. Even after the Russian conquest of Finland in 1808, Swedish speakers remained in local control.

Then the nationalist movement began to gain strength. Its most famous exponent was certainly Elias Lönnrot (1802–84). He was Professor of Finnish at Helsingfors (Helsinki) University: much more important, he collected folk songs from Karelia and remoulded them into the most successful 'epic' of the whole Romantic period, the *Kalevala*, a poetic retelling of Finnish mythology. Later written Finnish has gradually incorporated more features of the eastern (Karelian) dialects: the *Kalevala* is essentially in northern Karelian.

Finland asserted its independence at the time of the Russian Revolution in 1917, but its eastern boundary, especially after Russian gains in 1944, left most speakers of Karelian dialects inside the Soviet Union. Finnish speakers in Sweden face discrimination.

Finnish and related languages

Finnish, Estonian and their relatives form the surviving western end of a dialect continuum that once probably stretched, unbroken, across European Russia as far as MARI and MORDVIN territory – an area now long since occupied by Russian speakers. Though crossed frequently by travellers, the Gulf of Finland eventually effected the separation between early Finnish (with Karelian and Veps) and early Estonian (with Livonian and Vote).

Estonian dialects are strongly marked, from the conservatism of the south (separated from the rest by moorland and forest) to the Finnish-like character of the north-east, where only two vowel and consonant lengths are distinguished.

There are two dying languages related to Estonian. In western Latvia lived the *Livonians*: fewer than 200 people now speak this language. In Ingria, to the east, the *Vote* language is now spoken by about ten old people.

Finnish is the majority language of Finland. A very distinct dialect, Tornedal Finnish, has about 75,000 speakers, mainly inhabitants of the Torne valley in Sweden. There are also tens of thousands of Finnish migrants and settlers in central Sweden.

Karelian has about 120,000 speakers. Religious and other texts have been printed in Karelian for two hundred years, but the Soviet attempt to establish a Karelian literary language independent of Finnish was a failure. Russian and Finnish are the usual written languages of Karelia, a self-governing republic within Russia which has a strong Russian majority.

Karelian has several major dialect divisions. To the south, *Olonets* and *Ludian* (or Lydian) are sometimes considered separate languages. So is *Ingrian*, the Karelian dialect spoken until 1944 in 'Ingermanland' to the south of St Petersburg. That was when the Ingrians (*Iz-horas*), numbering 12,000, were deported to labour camps in the Kola Peninsula and Kazakhstan. As a result, Ingrian now has only a few hundred speakers. Even further south lay a community of Karelian speakers whose ances-

tors migrated from the western part of Karelia in 1617, when their homeland was ceded by Russia to Sweden, and settled near Tver, not far from Moscow.

Curiously, this event was repeated in 1944, when a tranche of western Karelia was ceded by Finland to the Soviet Union. This time, all Karelian and Olonets speakers in the district moved en masse to Finland.

Veps is spoken by fewer than 8,000 people living around Lake Onega. In the 1930s there was some publishing of textbooks in Veps, using the Latin alphabet, but nowadays children in these districts are taught in Russian.

Loanwords in Finnish and Estonian

Proto-Finnic, ancestor of Finnish and Estonian, was in contact with Baltic and Germanic languages in the first millennium BC. Baltic loanwords from this period include Finnish *meri* 'sea', *herne* 'pea', *lohi* 'salmon', *tuhat* 'thousand'. Germanic loanwords include *raha* 'money' (originally 'squirrel skin'), *leipä* 'bread' (cf. English *loaf*), *tunkio* 'dung heap' (cf. English *dung*).

Finnish and related languages

Around 500 AD speakers of Slavonic languages, spreading north-eastwards, were in touch with early Finnish and Estonian for the first time. Some later Slavonic loanwords demonstrate the introduction of Christianity to Finland by missionaries from Russia. They include Finnish *pappi* 'priest', *raamattu* 'book' – and Estonian *roosk* 'whip', *turg* 'market'.

Swedish and German domination gave rise to a large number of recent loanwords in Estonian, such as *värdjas* 'bastard', *piibel* 'Bible', *pööbel* 'mob'.

Kalevala and its children

Kalevala is the name of the legendary country where the mythological tales unfold. Lönnrot's long poem is full of brief magic charms, spells and lyric songs – sometimes disrupting the flow of the story – but perhaps its most valuable component, for they are wholly authentic records of the oral poetry of Karelia and Finland.

Ohrasta oluen synty,	The origin of beer is barley,
humalasta julkijuomen,	Of the high drink the hop plant,
vaikk' ei tuo ve'että synny,	Though it is not made without water
eikä tuimatta tuletta.	And a good hot fire.
Humala, Remusen poika,	Hop, son of Remunen,
piennä maahan pistettihin,	Was put into the ground when small,
kyynä maahan kynettihin,	Was ploughed into the ground like a snake,
viholaisna viskottihin . . .	Was tossed away like a nettle . . .

'The origin of beer': *Kalevala* 20, lines 139–46.

The success of *Kalevala* found echoes in English literature. Several translators tried successively to recreate its simple but haunting rhythm and insistent parallelism. Then the same rhythm was used in a new romantic 'epic' that achieved vast sales – Longfellow's *Hiawatha*. Longfellow's verse form and style were inspired by a German translation of *Kalevala* by his friend Ferdinand Freiligrath, though the story of Hiawatha is based on folk literature of the Ojibwa as retold by H. R. Schoolcraft in *Algic researches* (1839).

FORMOSAN LANGUAGES

PERHAPS 200,000 SPEAKERS

Taiwan

Three separate groups of AUSTRONESIAN LAN-GUAGES are spoken in the inland regions of Taiwan. The long separate linguistic development of each group makes them coordinate with the huge Malayo-Polynesian group to which belong all the remaining thousand Austronesian languages, with their 270,000,000 speakers, extending from Malagasy to New Zealand and Easter Island. These fascinating 'aboriginal' languages have apparently been spoken in Taiwan for at least six thousand years, and it must be from here that the first major Austronesian migration embarked on its southward route to the Batan islands and the Philippines.

> The island was called *Formosa*, 'beautiful', by the first Portuguese mariners who reached it, in 1498. Now outdated as a place name, this word still serves to identify the language group.

Chinese speakers began to settle in the lowlands of Taiwan in the early 17th century. But the fate of the Formosan languages was sealed when Taiwan became the haven of the ousted Nationalist government of China. This brought an alien, top-heavy national bureaucracy to Taipei, previously a rather isolated provincial capital. It also brought settlers from the mainland, rapid commercial development and a large international presence. Chinese, already the language of the elite, is spoken by an ever-increasing majority: it is the language that brings success and prosperity in the capital, Taipei, which lies all too near the remaining districts of 'aboriginal' speech. There were once twenty-five Formosan languages, but several have already become extinct.

The first studies of Formosan languages were made during a brief period of Dutch rule in the mid 17th century. The island belonged to Japan from 1895 to 1945 and some Japanese research took place then. More recently, research on these languages has been supported by Taipei's Academia Sinica. Politically, they have no status.

They are utterly different from Chinese and very far removed from the minority languages of mainland southern China. The typical word in the Formosan languages has more than one syllable, and there are no tones. Word stress falls on the last syllable in Atayal, on the last-but-one in Paiwan.

Formosan languages
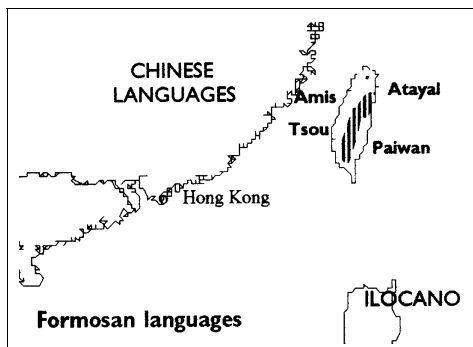
The Formosan languages

Amis (136,000 speakers, in the plains along the railway from Hualien to Taitung) and *Paiwan* (53,000, in the southern and south-eastern mountains) are classified in the Paiwanic group.

Atayal or Tayal (41,000 speakers, in the northern mountains) is the best-known member of the Atayalic group.

The Tsouic language group is closer to extinction. *Tsou* itself has about 4,000 speakers on the western slopes of Mount Yu in south central Taiwan.

Numerals in Atayal and Paiwan		
Atayal		**Paiwan**
qun	1	ita
rusha'	2	ḍusa
tū'	3	cəḷu
shupat	4	səpac
tima'	5	ḷima
matū'	6	unəm
pitu'	7	picu
pat	8	aḷu
qishu'	9	siva
pugh	10	ta-puḷuq

Comparative Austronesian dictionary ed. Darrell P. Tryon, (Berlin: Mouton De Gruyter, 1995)

FRENCH

70,000,000 SPEAKERS

France, Belgium, Canada, Switzerland, Italy, United States and many other countries

French is the most northerly of the ROMANCE LANGUAGES that descend from Latin, the language of the Roman Empire. Historically it is the language of northern France: it became France's national language, and spread to many other parts of the world with French conquest and trade.

> In Roman times the part of Europe in which French is now spoken was called *Gallia*, 'Gaul'. When the empire succumbed to 'barbarian' conquerors most of Gaul became the kingdom of the *Franci*, 'Franks'; they soon forgot their Germanic language, but the name of their kingdom survived as *France*. Its national language is therefore *Français* or *French*.

The Celtic-speaking inhabitants of Gaul were among the first non-Italians to take a full part in the culture of the Roman Empire. Not surprisingly, there are Celtic loanwords in Latin and in all the Romance languages. But there are more of them in French: they include *chemise* 'shirt'. The language of the Franks, which was close to Old High German, also contributed loanwords to French, including *fauteuil* 'armchair', originally **faldistôl* 'folding stool', an item of army equipment; *rôtir* 'roast', a cooking method more familiar to the invaders than to the Roman Empire; *houx* 'holly'; *marais* 'marsh'.

For reasons which have not been fully worked out – but must be linked with the speech patterns of Celtic and Frankish speakers – early French underwent a process of more rapid change than did the other Romance languages of the period. The final result is that French now looks and sounds far more different from Latin than do other modern Romance languages such as Spanish, Italian and Romanian.

As in the rest of western Europe, Latin remained in use as a written language while everyday speech continued to develop. The earliest clear evidence of a new language in the making in northern France is in the Strasbourg Oaths of 842 (see box at ROMANCE LANGUAGES), because the words of the agreement were faithfully recorded, in French and German, in Nithard's Latin history.

There are a few documents and religious texts in French of the 10th and 11th centuries, but the first real flowering of French literature is in epics, the first and greatest being the *Chanson de Roland* 'Song of Roland' of around 1200. They were recorded in manuscript form for oral recitation. From this beginning, French poetry soon became more varied and more consciously literary.

Although the language of Paris and of the neighbouring royal monastery of Saint-Denis was already influential, medieval French texts have varied dialect links. This is natural since Paris was not the only major centre of French cultural life. After the Norman conquest in 1066, London was another: for nearly two centuries after that date not English but the Anglo-Norman variety of French was the usual language of literature in England (alongside Latin). The oldest and best manuscript of the *Chanson de Roland* is Anglo-Norman.

As the connections between England and France grew more distant, Anglo-Norman – instead of developing into a new modern Romance language – regressed to a jargon of lawyers and courtiers. Its descendant, 'Law French',

can still be found in fossilised phrases in modern English legal terminology. But English, now revived as a language of culture and literature, had taken in a mass of loanwords from French, involving most aspects of everyday life, often providing near-synonyms to Germanic words: thus while English still uses Germanic terms such as *ox*, *sheep*, *pig* for the domesticated animals, it uses the French loanwords *beef*, *mutton*, *pork* (modern French *boeuf* 'ox', *mouton* 'sheep', *porc* 'pig') for their meat.

Meanwhile Paris was asserting its position at the centre of French culture. The central role of French, the French of Paris, followed from this. Two landmarks are the foundation of the University of Paris, chartered in 1231; the spread of printing, at the end of the 15th century; and the *Ordonnance de Villers-Cotteret*, 1539, which ruled that legal proceedings in France must be *en langaige maternel françois*, 'in the French mother tongue'. In practice, this asserted the uniquely privileged status of French not only against Latin but also against OCCITAN, BRETON, BASQUE and the local dialects or *patois* of French.

Yet French does borrow from its regional languages: *bijou* 'jewel' is a Breton loanword, while *bouillabaisse* 'fish soup' is one of many food words borrowed from Occitan dialects.

By the 16th century, French was the language of an astonishingly rich literature – and writings in French were read, admired, translated and imitated across all of western Europe. Among the greatest of older classics had been the poetic *Romance of the Rose* (adapted in English by Chaucer), the Arthurian romance sequence *Lancelot* (the main source for Malory's English *Morte Darthure*) and the vivid chronicles of the Hundred Years War written – in French that was influenced by his native Picard dialect – by Jean Froissart. The 16th century was a period of exciting and varied experiment, and also of much linguistic borrowing from Latin and from Italian.

A reaction followed, often identified with the influence of François de Malherbe (1555–1628). Written French became a rule-bound language, with an artificially restricted vocabulary. In spite of the Enlightenment (French was the language

of the great *Encyclopédie* of Diderot and others), in spite of the French Revolution and all that has followed, in many ways written French is still rule-bound. Spelling and usage are overseen by the Académie Française, a self-elected college of eminent authors and intellectuals, under government patronage. Standard French differs rather widely from most people's everyday speech. Traditional French verse, which some still write, demands a special pronunciation (see box). For all this, French remains the language of a very rich and flourishing literary culture, in some ways the most vital in Europe.

Counting in French

Un**e**, deux, trois,	One, two, three,
J'irai dans les bois –	I'll go to the woods –
Quatr**e**, cinq, six,	Four, five, six,
Cueillir desceris**es** –	To pick cherries –
Sept, huit, neuf,	Seven, eight, nine,
Dans un panier neuf:	In a new basket:
Dix, onz**e**, douz**e**,	Ten, eleven, twelve,
Ell**es** seront	
tout**es** roug**es**!	They will all be red!

This traditional children's song has a regular 3-syllable/5-syllable rhythm – only if one does not count the 'mute' *es*. Thus it breaks the basic rule of written French poetry. In reciting literary French poems, the final *es* must be sounded.

Europe: French

FRANCE, BELGIUM,
SWITZERLAND

Lebanon

NORTH AND
WEST AFRICAN
COUNTRIES

Indochina

CENTRAL AFRICAN
COUNTRIES

New Caledonia,
New Hebrides

Indian Ocean islands

MADAGASCAR
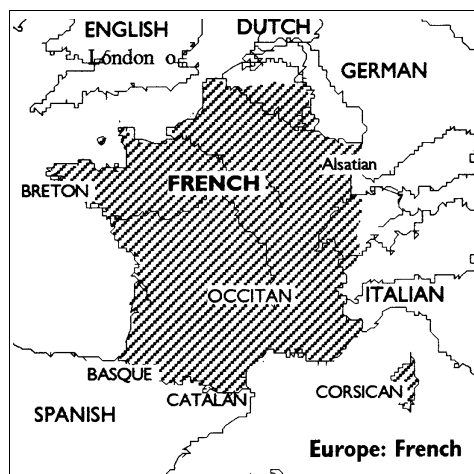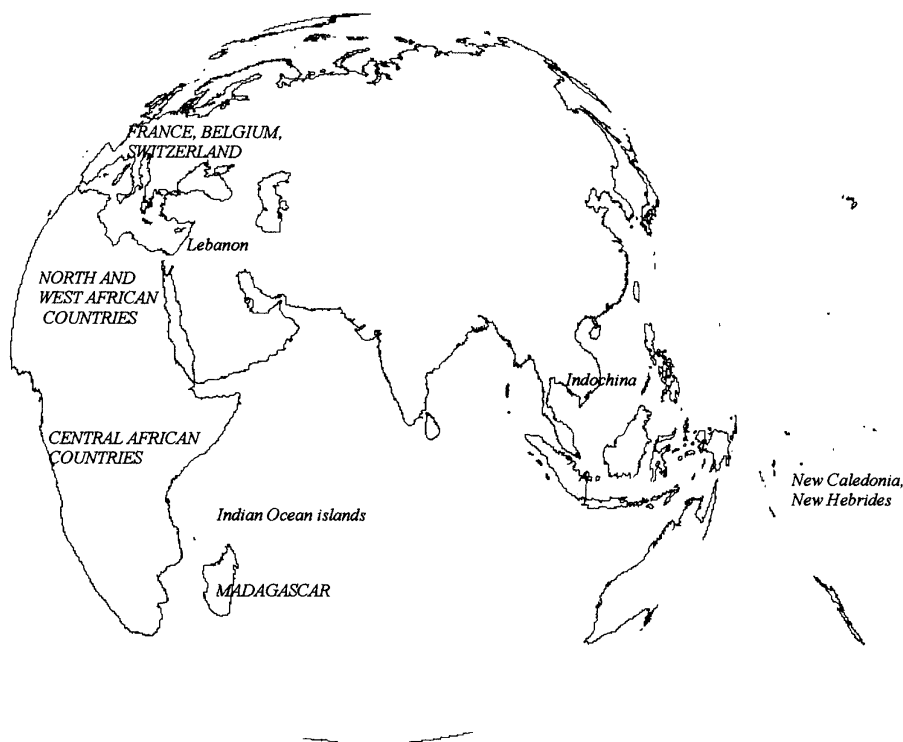
La Francophonie: **French as a world language**

Unchallenged as the national language of France, French is also one of the major languages of the world. Until recently its status as the worldwide language of diplomacy was unrivalled: in practice this role has now been taken over by English. For most children of the former French Empire it is the national language, used in school from an early age; for many other children across the world it is the first foreign language that they learn, but, once more, this role is more and more taken by English.

The cultural role of French as a foreign language has a long history. William the Conqueror found that Edward the Confessor's court, in London in 1051, spoke French. Franco-Italian,

the literary French of northern Italy, goes back at least to the 13th century. In this distinctive form of French were written not only a major variant of the *Chanson de Roland* but also the record of Marco Polo's travels (known as *Il Milione*).

Much later the Venetian adventurer Giacomo Casanova (1725–98) chose French as the language for his memoirs. This was far from unusual: in the 17th and 18th centuries French was the first language of culture in many European courts. Frederick the Great of Prussia, tutored by Voltaire, wrote a history of German literature in French. French was the favoured language of the 18th- and 19th-century Russian court; in the 19th and early 20th centuries French was cultivated

QUEBEC,
Acadia, other parts of Canada

Louisiana

HAITI,
other Caribbean islands

Guyane

French Pacific Territories

original borrowing took place. *Eventuellement* must be translated not 'eventually' but 'possibly'; *sensible* not 'sensible' but 'sensitive'; *librairie* not 'library' but 'bookshop'. Even recent English loanwords emerge as false friends: *slip* is not 'slip' but 'panties, knickers'; *parking* is 'parking lot, car park'.

French spelling differs almost as wildly from the real pronunciation of the language as does that of English. This is a legacy of the printers and scholars of the 16th century, who – instead of adapting spelling to changing speech habits, as earlier writers had done – tried to embody the origins and history of words in their spellings.

French and its dialects

France still has over a million native inhabitants whose mother tongue is not French: the significant minority languages are Basque, Breton and the Alsatian dialect of German. In this context no figure can be given for the Occitan-speaking minority, because Occitan is popularly felt to be a French patois and very few parents will now encourage their children to speak it.

Across the borders of France, French is also one of the official languages of Belgium, Luxembourg and Switzerland. French is particularly vigorous as a language of culture and literature in southern Belgium. It is the everyday language of the Belgian capital, Brussels, which is now also the de facto capital of the European Union. French is naturally one of the working languages of the Union.

The Channel Islands, once part of the Duchy of Normandy, became English possessions around 1200. Yet *Les Îles Normandes* (as they are known in France) were still largely French-speaking at the beginning of the 20th century.

After eight hundred years of political separation, the French of Jersey and Guernsey (now spoken only by older people in country districts) is rather different from the dialects of modern Normandy, and still more different from standard French.

Each island has its own constitution. Guernsey

enthusiastically by courts and intellectuals in the new states of eastern Europe. Practically all European languages have numerous French loanwords for items of culture and fashion.

The difficulty for English speakers in learning French comes at the beginning, with the pronunciation and the spelling. French vocabulary seems familiar to English learners because, although the languages are only distantly related, English has borrowed a great number of words from French and both languages have borrowed freely from Latin. But there are 'false friends', words that look the same but mean something different, usually because their meaning has changed in one or both languages after the

has its *Douzaines*, 'parish councils', and Sark has its *Greffier* or Registrar. The local Protestant tradition is signalled by the word for 'church', *église* in mainland French, *temple* in the Islands.

French outside Europe

With the explorations and conquests of the last few centuries French has spread far across the world. The ugly but necessary word *Francophonie* designates the world community of French-speaking countries.

French is one of the two official languages of Canada. The French-speaking community there, centred on Quebec, feels (and is) threatened by the attractions of the English-speaking culture of North America: the province of Quebec legislates incessantly to give the advantage to French. Parisian French is official, but there is a local colloquial standard and also local dialects, which show special affinity with the Normandy dialect, since the majority of migrants to Canada came from Normandy and northern France.

There is also an old-established French minority in *Acadie*, a district of New Brunswick. Many Acadians migrated to the French colony of Louisiana in the late 18th century, and they now form the main French-speaking minority in the United States. Their dialect is *Cajun*. Also in Louisiana a FRENCH CREOLE is spoken, a result of migration from the Creole-speaking islands of the Caribbean.

Elsewhere French is the standard language of French colonies and conquests, both those which remain under French rule (known as the *DOM-TOM*, *Départements et Territoires d'Outre-Mer*) and those which have become independent. Naturally, local colloquials become established, influenced by the various local languages of these countries: for example, the French of Guinea has a sound pattern clearly related to that of WOLOF.

The 'Servants' French' of Vietnam

Tây Bôi, 'boys' French', was a pidgin spoken in Vietnam under French rule from about 1860. French words were cut down to fit Vietnamese sound patterns: the coin *piastre* became *bi-at*, while *ordre* 'order' became *ot*. (One sees why the French used to call the Vietnamese *mangeurs de syllabes*, 'syllable-eaters'.) Vietnamese classifiers accompanied French nouns, *trai coco* 'coconut fruit'. Numbers were reorganized: eleven was *dit-ong*, derived from French *dix-un* 'ten-one'. The French departed in 1947 and *Tây Bôi* was already almost forgotten by 1960, when an American army pidgin was beginning to take its place.

After John E. Reinecke in *Pidginization and creolization of languages* ed. Dell Hymes (Cambridge: Cambridge University Press, 1971) pp. 47–56

Pidgin forms of French developed in contacts between Europeans and Indians in Quebec and Louisiana in the 17th and 18th centuries: two survivals are *Mitchif* (see CREE) and *Métis French* of the Canadian West, both of which became the language of mixed French-Amerindian communities. In Africa, *Petit Mauresque* 'Little Moorish' was a pidginised French that seems to have emerged from the old Lingua Franca (see box at ROMANCE LANGUAGES) eventually to merge into the local colloquial French. *Petit-Nègre* 'Little Negro' was a similar temporary development in West Africa, especially Ivory Coast. Both were most used in and around the French Army. More important than these pidgins are the FRENCH CREOLES dealt with in the next article, for these serve as mother tongues for many millions of people and some of them have reached official status.

FRENCH CREOLES

SEVERAL MILLION SPEAKERS

AS FRENCH was used across the world in trade and empire, a group of creole languages developed from it, mainly used by mixed communities arising from intermarriage and slavery.

For the origin of the name *Creole* see ENGLISH CREOLES AND PIDGINS. In the Caribbean, everyday Creole may be called *kreòl rèk* 'crinkly Creole' or *gwo kreòl* 'fat Creole'. Elevated Creole, closer to French, may be called *kreòl swa* 'silky Creole' or *kreòl fen* 'thin Creole'.

While some creoles struggle for any form of recognition, others have become official languages, like SEYCHELLOIS. HAITIEN is so universally used in Haiti that its status is no longer threatened by French. The varying status of creoles has much to do with local history and attitudes.

In Martinique and Guadeloupe, where schooling and administration are in French, self-conscious attempts to make the Creole a literary language have failed. Yet Creole does appear in everyday writing more and more – in advertisements, in several very popular comic strips – and Creole is the language of local popular music. In Dominica (French 1632–1732, 1778–83, British at other periods, now independent), English is the formal language. There is a strong movement to give some recognition to the Creole, with an annual *Jounen Kwéyòl* (Creole Day).

The first ten numerals in the Creole of Guadeloupe are: *an, dé, twa, kat, sen, sis, sèt, uit, nèf, dis*.

French creoles:
the Caribbean and Indian Ocean groups

French creoles are classified by linguists into *New World Creoles* and *Isle de France Creoles* – the latter group named after an insignificant island in the Indian Ocean.

Major New World Creoles include HAITIEN, French Guyanese (50,000 speakers) and the well-known *Louisiana Creole* (40,000 speakers), sometimes called Gumbo. The gift of Louisiana Creole to world children's literature is Brer Rabbit. These traditional tales were told in the 19th century and appeared in print, in Creole, in local newspapers from about 1875 onwards.

The subgroup of Lesser Antillean Creoles (*Antillais* or *Kwéyòl* or *Patwa*) has at least 1,000,000 speakers. It includes the creoles of Martinique and Guadeloupe, which are French *départements*, and those of St Lucia, Dominica, Désirade, Marie Galante, St Martin, Les Saintes and St Barthélemy. San Miguel Creole, now spoken in Panama, originates from St Lucia and belongs to this subgroup.

Isle de France Creoles include Réunion Creole or *Réunionnais* (once called Bourbonnais:

Indian Ocean:
French creoles

SEYCHELLOIS

Swahili

Malagasy

MAURICIEN
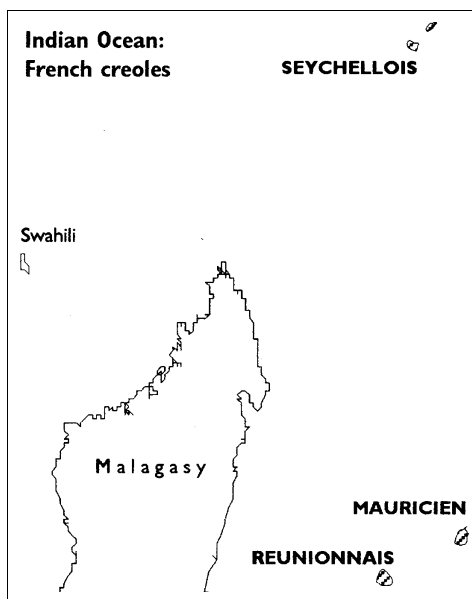
REUNIONNAIS

550,000 speakers), *Mauricien* (600,000 speakers) and SEYCHELLOIS. Mauricien and Réunionnais are also used as trade languages on the Malagasy coast and in the Comoro Islands. Réunion is a French département; Mauritius and the Seychelles, formerly under British rule, are now independent states.

Brer Rabbit: smartened up?

Frère Lapin stories are still told in Louisiana Creole, but the language now shows heavy English influence:

> Lapẽ mõjé tu lafer. Lapẽ smart!
> Rabbit eats the whole lot. Rabbit's smart!

Creole in Mauritian politics

The linguist Peter Stein arrived in Mauritius on 3 February 1975, the day on which MBC broadcast its first-ever programme in Creole: a political debate on the American takeover of Diego Garcia. He switched on his tape recorder just in time. Anerood Jugnauth, party leader of the Mouvement Militant Mauricien, was speaking:

. . . Sa mem ase pu fer nu fremir: Amerikẽ ule fer en baz militer lor Dyego Garsya; zot ule ãpil ban bom atomik laba.

That itself is enough to make us tremble: the Americans want to make a military base on Diego Garcia; they want to fetch in some atomic bombs down there.

After Peter Stein in Les créoles français entre l'oral et l'écrit ed. Ralph Ludwig (Tübingen: Narr, 1989) pp. 217–18

Jugnauth's name – and his effective oratory – are a reminder of the quadrilingual nature of Mauritian society. It is an Indian name, deriving from the Hindu god Jagannātha, but the spelling is 19th-century English: compare the English word *juggernaut*, which has exactly the same origin.

The Caribbean: major areas of **FRENCH,** ENGLISH and *SPANISH* speech
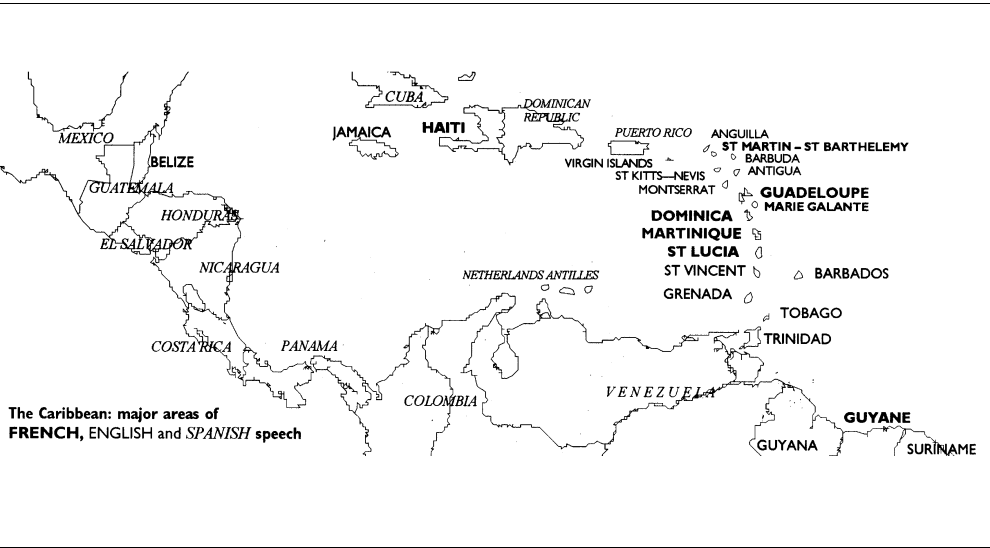
FRISIAN

750,000 SPEAKERS

Netherlands, Germany

One of the GERMANIC LANGUAGES, Frisian is a regional language of the province of Friesland in the Netherlands, spoken by about 5 per cent of the whole population of the country. It is also spoken by scattered and shrinking communities along the north German coast (see map and table of numerals at DUTCH).

> The Germanic tribe *Frisii* are recorded by a 1st-century Latin source as already settled north-east of the mouth of the Rhine. The language is now known to its speakers as *Frysk*.

It is no coincidence that Frisian is the closest relative of English: it was from this coast that the 'Angles, Saxons and Jutes' of traditional history set out in the 5th century. Frisian was evidently the language of the coast dwellers and seamen in those times, and easily became the lingua franca of the miscellaneous population that crossed from here to take army service – and eventually to settle – in Britain.

This is reconstruction, but it is supported by the fact that for three centuries after the migration Frisian certainly was a crucial language of the North Sea and Baltic seaborne trade. The Frisians' commercial centre was Dorestad, near Utrecht. Their power was destroyed by the Franks and the Vikings. Frisia remained a fairly autonomous but gradually shrinking political unit, extending from northern Holland to the mouth of the Weser, until the early 16th century, since when the western half of it has been more and more closely linked with the Netherlands.

'Old Frisian' was regularly used administratively until about 1500: it was then supplanted for most purposes by Dutch and German. But West Frisian, the dialect of Friesland, has been a literary language since the 17th-century writings of Gysbert Japicx, and has a small but flourishing publishing industry today.

Most speakers of West Frisian nowadays are bilingual in Dutch. True Frisian is a language of the small towns and villages of Friesland: in Leeuwarden and other large towns *Town Frisian*, a mixture of Frisian and Dutch, is the everyday speech.

The Frisian dialects of Germany are unlikely to survive long. All are heavily influenced by the surrounding Low German dialects and by Standard (High) German. North Frisian also shows the influence of Danish and English, a reminiscence of the English occupation of Helgoland, which contributed such friendly terms as *blakhol* 'prison cell' and *laisen* 'boat licence'.

> *Bread, butter and green cheese | Is good English and good Friese*, says the proverbial rhyme, a reminder that Frisian is closer to English than any other language of Europe. Nowadays, though spelling sometimes looks different, pronunciation is still close: Frisian *ikker*, English *acre*; Frisian *sliepe*, English *sleep*.

FULANI

PERHAPS 15,000,000 SPEAKERS IN THE COUNTRIES OF WEST AFRICA

Historically the Fulani have been a migrant, pastoral people. With their migrations the Fulani language, one of the Atlantic group of NIGER-CONGO LANGUAGES, has spread widely across the Western Sudan, the inland plains of West Africa. It is now a national language in Guinea, Niger and Mali and important regionally in several other countries (see map at DIOLA).

Ful, Fula, Fulani are English forms, based on the Hausa name for the speakers of this language, *Filani*. In Kanuri they are called *Felata*, in Mõõre the term is *Sìlmíìgá*. *Peul* and *Toucouleur* are the usual French names.

Speakers in Senegal call their language *Pulaar*, in Guinea *Pulle*, and in areas further east *Fulfulde*. They call themselves *Pullo* (singular), *Fulɓe* (plural).

Some Fulani words are noted by Arabic authors who had travelled in the Western Sudan in the 12th to 14th centuries. European records of the language date from the 17th century.

Fulani was (and is) the language of a pastoral people who gradually spread eastwards over these centuries across inland West Africa, from a starting point in lower Senegal. By the 16th century they were at Macina and the middle Niger.

At the beginning of the 19th century the Fulani Osman dan Fodio founded an empire among the Islamic peoples of what is now northern Nigeria. His follower Adama, who came from the plateaus to the east, was sent by Osman back to his own lands to stir up the Fulani of the region, to spread Islam and to extend the boundaries of the empire. The Adamawa plateau of eastern Nigeria and northern Cameroun is now named after him.

Adama was so successful that, in the area of the emirate that he established, Fulani are still the dominant group. They encouraged the growth of Islam among the peoples of the plateau, and their language became more and more widely used as lingua franca among them. There is also a pidgin version of it, *Kambariire*. This name comes from the Hausa *kambari* 'a Muslim living among non-believers'.

Thus Fulani speakers are now to be found all over West Africa, from the Gambia to Cameroun. They are not in a majority anywhere, but countries in which there are significant minorities include Senegal, the Gambia, Guinea, Mauritania, Mali, Sierra Leone, Liberia, Ivory Coast, Ghana, Burkina Faso, Niger, Benin, Togo, Nigeria, Chad, Sudan and Cameroun.

A literary form of Fulani, used in the oral recitation of texts, has special grammatical features and can serve as a 'secret language'. There are also pidginised forms of Fulani spoken by those who use it as a trade language.

Linguists once doubted the relationships of Fulani, partly because of the distinctive appearance of typical Fulani speakers. Meinhof, in the 19th century, classed it as a Hamitic (AFRO-ASIATIC) language. Its link with the languages now classed as the Atlantic group within Niger-Congo was first stated by the French linguist Faidherbe in 1875 – yet it is possible to read, even in books published in the 1990s, that Fulani 'originated in Egypt'.

Printed literature in Fulani was slow to develop, though missionaries produced Bible translations and Christian books in the dialects of Adamawa, Macina and Futa Djallon.

A remarkable feature of Fulani is initial consonant alternation. Related words have varying initial consonants, in a complex but regular pattern, depending on their grammatical status: thus *jeso* 'face', *gese* 'faces', *ngesa* 'big face'. Noun classes, as in other Niger-Congo languages of

West Africa, are marked by suffixes, which are different for singular and plural. The two factors combine to make plural forms very different from singular forms for many nouns: *gorko* 'husband', *wor be* 'husbands'; *linngu* 'fish', *liddi* 'fishes'.

Fulani has borrowed from Arabic and, more recently, from French and English: *nuuru* 'light' from Arabic *nur; galaas* 'ice' from French *glace; teebur* 'table' from English. For a table of numerals see WOLOF.

The Fulani alphabet

a e i o u b ɓ mb c d ɗ nd f g ng h j nj k l m n ŋ ny p q r s t w x y ŷ z

---

Writing in Fulani

Fulani is most often written in a variant of the Latin alphabet. But the usual alphabetical order is quite specific to Fulani: the vowels come first, and nasalised consonants follow immediately after corresponding non-nasalised consonants. ɓ and ɗ are 'explosives'; ŋ is a velar nasal, like English *ng* in *singing*, and it is sometimes written *ng*. Long vowels are written doubled.

This script derives from the 'Africa alphabet' once championed by the International African Institute in London, now taken up by Unesco. The problem with it has always been that special equipment or software is needed to type or print it. None of the many fonts used in this book contains the curled y which is the last-but-one letter of this Fulani alphabet.

GÃ

1,000,000 SPEAKERS

Ghana

Gã and Dangme are two closely related NIGER-CONGO LANGUAGES spoken in the south-eastern corner of Ghana. *Gã*, variant of *Nkrã*, is the name of the city now better known in its Europeanised form *Accra*.

According to local tradition Gã speakers brought their language southwards to the coastal towns of what is now eastern Ghana in a series of migrations down the Volta valley in the 16th and 17th centuries. The towns were traditionally small independent units: the Gã state was a British colonial creation.

A grammar of *Acraisk* or Gã, by C. Protten, was published in Danish in 1764, reminder of former Danish interests on the West African coast. Gã came to greater prominence when German missionaries began to work on the Danish Gold Coast in 1826. From then onwards, under Danish, British and independent governments, Gã has been a language of education and religion. Its special importance nowadays comes from the fact that it is the local language of Accra, the capital, and of a compact district immediately surrounding the city. For the majority of inhabitants of this district it is either a first or a second language, necessary at Accra markets.

Dangme or Adangme is a group of dialects closely related to Gã and spoken to the east and north-east. Until a few years ago, Gã was the standard language used in education and written contexts in both areas – and so most people there are bilingual in Gã – but a standard written form of Dangme has now been established.

Gã has two tones, Dangme three. Gã has borrowed heavily from Akan and will no doubt continue to do so: thus its vocabulary, in such details as the names of foodstuffs, is quite distinct from that of the Dangme dialects, which show less external influence.

The first ten numerals are: *eko, enyõ, etẽ, e'fe, enũo, ekpa, kpawo, kponyõ, nehũ, nyongmã*.

Gã and Dangme on the map

The *Gã* towns are Accra, Osu, Labadi, Teshi, Nungwa and Temma. Accra itself is in origin a federation of seven quarters, three of which once formed Jamestown or 'English Accra', the other four 'Dutch Accra'. There is a distinct pidginised form of Gã used by non-native speakers as a lingua franca in the streets and markets of the city.

The six dialects of *Dangme*, as usually counted, correspond to traditional political units. The inland dialects are Shai, Krobo and Osudoku. The coastal ones, from west to east, are Kpone (where speakers now use Gã or are bilingual in it), Prampram (close to Gã), Ningo and Ada (the last strongly influenced by EWE).

West Africa: Gã and Dangme
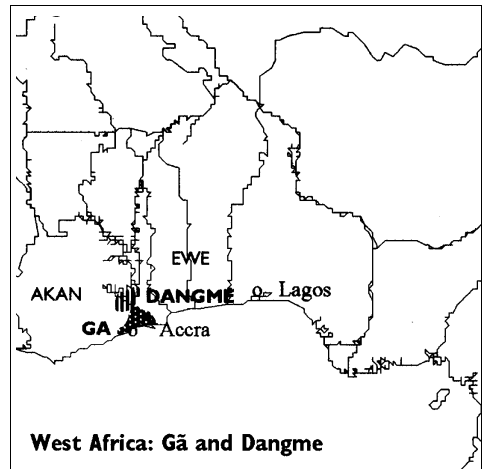
GAELIC

80,000 SPEAKERS

Scotland

Scottish Gaelic, one of the *q*-CELTIC LANGUAGES (see map there) and historically an offshoot of Irish, spread into western Scotland from the 4th century onwards. There was heavy migration across the narrow straits that separate Ulster from Galloway, and settlement spreading from south-western Scotland over land that had formerly belonged to the Picts and the Vikings. The even heavier influx of English speakers, which began a few centuries later, has in turn almost overwhelmed Gaelic, which is now commonly spoken only in some of the western islands and the most isolated parts of the western Highlands. It was never the language of the whole of Scotland.

> Gaelic is known to its own speakers as *Gàidhlig; Hielans*, the 'language of the Highlands', is a term that has been used in Scots.

Celtic Christianity reached Iona with Columba's arrival from Derry in 563 and spread across Scotland and beyond. The Gaelic-speaking realm of the west gradually absorbed the old Pictish kingdom, the British kingdom of Strathclyde (which spoke a variant of Welsh), and part of Northumbria, by the 11th century, but Gaelic was never widely spoken over this large territory. Norman French was adopted, as in England, as the court language, while the Scots variety of English gradually spread north-westwards in everyday speech.

By the 17th century Gaelic had retreated to the Highlands and the Hebrides, where it had meanwhile absorbed the remaining Scandinavian-speaking territory: Norse rule in the Hebrides had ended in 1266. Independent Gaelic culture was gradually undermined by central government: *bards*, the itinerant poets and story-tellers, were outlawed. The voluntary Gaelic schools of the early 19th century were replaced by English-speaking state schools after 1872.

In the first centuries of publishing, Irish was regarded as the literary standard for Gaelic: the Liturgy of 1567 and the Bible of 1690 are both essentially in Irish. Gaelic literacy is at its strongest now in northern Skye, Lewis, Harris and North Uist, where a Calvinist religious tradition encourages home worship and Bible study. The language is also spoken in Tiree, Mull, the remainder of Skye and parts of the western Highlands. These districts are the *Gàidhealtachd*, corresponding to the *Gaeltacht* of western Ireland. There is now some bilingual primary education in these areas. Practically all speakers of Gaelic are bilingual in English.

Comhairle nan Eilean, the Western Isles Island Area Council, created in 1975, has a bilingual policy, using both Gaelic and English in its activities and on public signs. There is broadcasting in Gaelic, but there has never been a generally accepted standard spoken dialect, and listeners tend to dislike hearing dialects that are not their own.

There are small communities of Gaelic speakers in Glasgow, Edinburgh, London and in the Canadian province of Nova Scotia. It was in Cape Breton Island, Nova Scotia, that the first all-Gaelic newspaper was published: *Mac Talla* 'The Echo' ran from 1892 to 1904.

> Little is known of the language of the *Picti* 'painted people' who were the Romans' main opponents across the Imperial boundary that was marked by Hadrian's Wall. Some scholars believe that this early language of Scotland

will have belonged to the Celtic family, though not especially close to the Gaelic of today.

Scottish Gaelic has many Scandinavian loan-words: *ób* 'creek', *mód* 'court', *úidh* 'ford'. Scottish names such as Lamont, MacCorquodale and MacLeod (*Mac Laomuinn, Mac Corcadail* and *Mac Leóid* in Gaelic) are ultimately Norse. As many as half of the place names of the western Highlands and the Hebrides are Gaelic versions of originally Norse names.

Numerals in Manx and Scottish Gaelic		
	Manx	**Scottish Gaelic**
1	unnane	aon
2	daa, ghaa	dà
3	tree	trì
4	kiare	ceithir
5	queig	céig
6	shey	sia
7	shiaght	seachd
8	hoght	ochd
9	nuy	naoi
10	jeih	deich

GALICIAN

3,000,000 SPEAKERS

Spain

One of the ROMANCE LANGUAGES, Galician is a twin language with Portuguese – but their later histories have been very different. It is spoken in the north-west corner of Spain, and until recently has had no official recognition as a minority language.

Galicia (in Roman times the territory of the *Gallaeci*) was the western extremity of that mountainous strip of northern Spain which never came under Muslim rule. Instead, in early medieval times, Galicia was the realm of the Suebi, one of the Germanic peoples who had invaded and parcelled the western Roman Empire. Galician *laverca* 'lark' may possibly be a loanword from Suebic.

In the 11th century Galicia was divided, along the River Minho, from a new province of *Portugal* to the south. Both were fiefs of the kingdom of Castile. But in 1143 Portugal became an independent kingdom.

The medieval language of the two provinces is often called Galician-Portuguese: no significant dialect difference existed between them. Since then, Portuguese, as the language of an independent country and a world empire, with a

Europe: Portuguese and Galician
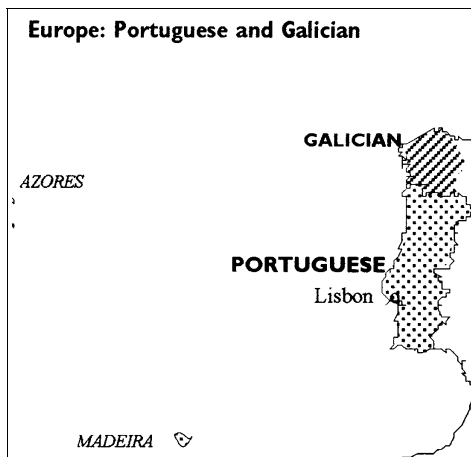
flourishing literature, has gradually grown apart from Galician, which remains of local interest only. The two are still much closer to each other than Galician is to Spanish.

For a table of numerals see CATALAN.

Galician and Portuguese are divided by the political boundary which follows the River Minho.

Ganda

3,750,000 SPEAKERS OF GANDA, SOGA AND GWERE

Uganda

One of the BANTU LANGUAGES, Ganda was the ruling language of the Kingdom of Buganda. Ganda, counted together with its eastern relatives Soga and Gwere, is spoken by a quarter of the population of Uganda.

'Readers who are unfamiliar with Bantu languages should note that these languages use systematic alternations of prefixes. The *Baganda* are the people who speak *Luganda*; they live in *Buganda*; and a single member of the tribe is a *Muganda*' (Peter Ladefoged and others, *Language in Uganda* (London: Oxford University Press, 1972) p. 17). In most cases the official English name of Bantu languages now omits the prefix: thus *Ganda*, not *luGanda*.

Early European explorers, approaching East Africa from the Swahili-speaking coast and guided by Swahili travellers, often used Swahili forms of proper names. The Swahili name for the kingdom of the baGanda was *uGanda* – hence the modern official name for the country. When the explorer J. H. Speke reached the outflow of the Nile on the northern shore of Lake Victoria, in 1862, he was told to call the place *uSoga*.

Ganda and related languages are now to be heard close to the point from which Eastern Bantu languages began to disperse, well over two thousand years ago. At that period there appears to have been some interaction near here between speakers of early Bantu, Nilo-Saharan and Afroasiatic languages, resulting among other things in the spread of iron-working and of banana cultivation among Eastern and other Bantu-speaking peoples.

However, the older history of the kingdoms north of Lake Victoria and of their languages is not known. They do not show strong influence from the other language families of eastern Africa: the old theory that the kingdoms are 'Hamitic conquest' states is unfounded. There are *baHima* cattle-farmers among the speakers of Ganda, but – by contrast with NKORE – no evidence that they were ever dominant or even respected.

In the late 18th century, from Egypt to the north and from the East African coast, Arabic and Swahili-speaking traders began to exert influence in the powerful kingdom of Buganda (whose language was Ganda) and its eastern neighbours, about twenty much smaller kingdoms now forming the district of Busoga. There came to be a royal monopoly on trade with the Arabs. They were followed, from the mid 19th century, by English-speaking travellers, traders and conquerors. Captain F. D. Lugard, of the British East Africa Company, dominated Kampala in 1890–2 (and wrote *The rise of our East African Empire*, 1893).

Buganda was the first of the states of modern Uganda to attract Arabic and Swahili trade, and eventually the first to submit to British government suzerainty in 1894. As a result, Kampala was already a regional centre of influence, and schools, roads and colonial administration naturally developed here first. Ganda speakers made ideal 'advisers' and civil servants in Busoga and other neighbouring states when they in their turn were taken over and developed. Thus Ganda gradually became the major language, a kind of lingua franca, in the whole of Uganda.

There was no conscious policy in this. Indeed, the British tried in the 1920s to make SWAHILI the official language of Uganda. But Ganda had too strong a hold. It was and is widely used in religion

and in education from primary level onwards: Soga, by contrast, is scarcely used in these contexts.

Ganda is also, alongside English, the principal language of the media. The first grammar of Ganda appeared in 1882, and it was soon followed by a growing range of publications in the language. At first this was missionary-inspired, but much traditional literature and history has appeared in print. For many decades Catholics and Anglicans kept to their own slightly different spelling rules.

Ganda is easily learnt by most Ugandans because their own languages are closely related to it: they include the speakers of Soga and the group of dialects related to NKORE. RUNDI and Rwanda also belong to the same branch of Bantu languages. In northern Uganda, where unrelated languages are spoken, Ganda is less used as a lingua franca.

East Africa: Ganda, Soga and Gwere

'The Luganda word *èmmûndú* 'rifle' looks like a loan from Swahili *bunduki*, itself a loan from Turkish via Arabic. The ultimate origin of this word is the Greek *pontikòn* 'hazelnut', referring to the shape of a musket bullet.'

Jan Knappert in D. Dalby and others, *Language and history in Africa* (London: Cass, 1970) p. 82

Ganda, Soga and Gwere on the map

'If the political divisions could be disregarded, Luganda and Lusoga might just be considered to be very different dialects of the same language'
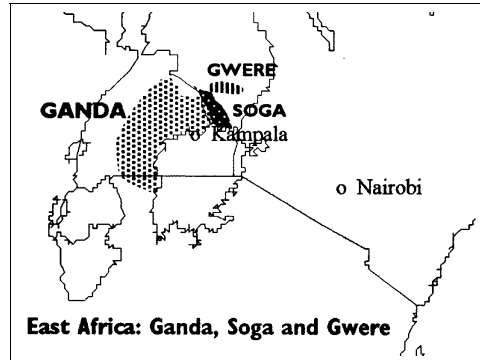
(P. Ladefoged (1972)). For all that, *Soga* speakers consider Ganda an alien language, one that they need to know for political and religious reasons. Their kingdoms were historically independent of Buganda. The two, together with Gwere, can be called the North Nyanza group of Bantu languages, and form a dialect continuum.

Ganda is best defined as the language of the Kingdom of Buganda, with its capital at Kampala. About 2,300,000 speak it as a mother tongue.

Soga is the language of the following former kingdoms: Bugabula, Luuka, Bulamogi, Bugweri, Bukoli, Busiki, Kigulu, Butembe (around the town of Jinja), Bunya and Bunyuli, with at least ten other very small states. There are perhaps 1,200,000 speakers. Two main dialects are recognised, *luPakoyo* in the north and *luTenga* in the south.

Gwere, with about 250,000 speakers, is the dialect of Bugwere and is the easternmost of the group.

Numerals in Ganda, Nkore and Rundi			
	Ganda	**Nkore**	**Rundi**
1	-mu	emwe	limwe
2	-biri	ibiri	kabiri
3	-satu	ishatu	gatatu
4	-nya	ina	kane
5	-taano	itaano	gatanu
6	mukaaga	mukaaga	gatandatu
7	musanvu	mushanju	karindwi
8	munaana	munaana	umunani
9	mwenda	mwenda	icyenda
10	ekkumi	ikumi	icumi

GARHWALI

1,200,000 SPEAKERS

India

One of the INDO-ARYAN LANGUAGES (see map there), Garhwali is the local language of the former state of Garhwal, now part of Uttar Pradesh. Garhwal has a place in three mythologies. To Hindus it is *Uttarakhāṇḍa*, 'the last stage' on the pilgrimage to the sources of the Ganges and the Jumna: both great rivers rise in this small Himalayan region. To the colonial British, Garhwal was the state in which the proud administrative centre of Dehra Dun and the fashionable hill station of Mussoorie lay close together, within easy reach of Delhi. And at Rishikesh, where the Ganges flows out of the mountains into the North Indian plain, the Beatles met their Indian teacher, the Maharishi Mahesh Yogi.

Garhwali, like its neighbour Kumauni, belongs to the Pahari subgroup of Indo-Aryan. Difficult communications in the Himalayan valleys have led to the crystallisation of several distinct languages with relatively small numbers of speakers. The Garhwali numerals, however, are identical with those of KUMAUNI.

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

GBAYA

PERHAPS 1,200,000 SPEAKERS

Central African Republic, Cameroun

Gbaya belongs to the Ubangi group of NIGER-CONGO LANGUAGES, and, like BANDA, has been spoken in its present location for about three thousand years.

Gbaya was apparently once the ruling language of a state in what is now the western part of the Central African Republic and south-eastern Cameroun; at any rate it became the second language for people of several different mother tongues in this region.

In the 19th century the Gbaya were subject to slave-raids. This was why Gbaya was also known in Ngaoundéré in northern Cameroun, the principal market for Gbaya slaves.

Gbaya speakers live on both sides of the River Kadei, which roughly corresponds to the frontier between Cameroun and the Central African Republic. Gbaya and Banda are the two major hinterland languages of the Central African Republic, now overshadowed linguistically by the national language, SANGO (see map there).

In Gbaya nouns and noun phrases a prefix marks the plural: *wi* 'man', *yowi* 'men'; *polombo wi* 'young man', *yopolombo wi* 'young men'. The first ten numerals are: *kpém, rííto, tar, nar, mɔɔrɔ́, 'dong kpém, 'dong rííto, nú-nãá, kusi, 'bú.*

GEORGIAN

3,500,000 SPEAKERS

Georgia

One of the Kartvelian or South CAUCASIAN LANGUAGES, Georgian is the national language of independent Georgia, in the Caucasus, formerly one of the constituent republics of the Soviet Union. Apparently unrelated to all other languages except its three close neighbours, Georgian is the vehicle of an ancient Christian culture and of a literature that goes back to the 5th century AD.

Modern Georgia (*Sak'art'velo*) includes both ancient *Iberia* (modern Georgian *K'art'li*) and ancient *Colchis* (later *Lazica* and *Abasgia*, modern Georgian *Egrisi*) from which, according to Greek legend, the Argonauts brought home the Golden Fleece. These territories were politically united in 1008 under King Bagrat III.

For some centuries before that unification, K'art'li had been under Muslim rule; it now became the heart of an independent kingdom, but one that was soon to fragment. Georgia allied itself with Russia in the late 18th century, for Russia was also a Christian state, a potential protector in view of the threat of Turkish conquest. In due course Russia became ruler rather than ally. Georgia asserted independence in 1918 after the Russian Revolution, but was reconquered by the Soviet Union in 1921. Georgians were prominent in Soviet politics: Yosif Vissarionovich Stalin (Joseph Djugashvili) was a Georgian; Aleksandr Shevardnadze, once foreign minister of the Soviet Union, was to become president of newly independent Georgia. Independence was achieved once more in 1991, since when Georgia has been beset by civil warfare.

Christianity in Georgia is traditionally dated to the conversion of King Mirian in 337 and Georgian literature began with religious texts, most of them translated from Greek: the 5th-century *Martyrdom of Shushanik* is identified as the earliest original work. Thus there is a very long tradition of native literature in Georgian, including royal chronicles and epic poetry. The 12th-century poem by Shota Rustaveli, *The Man in the Panther's Skin*, is the best known work of Georgian literature.

Georgian serves as the official and literary language for speakers of Svan and Mingrelian. Many Abkhaz speakers are bilingual in Georgian. The Jewish community in Georgia has had a distinct language, Judaeo-Georgian, which may still have as many as 10,000 speakers – many of them now in Israel.

There is also a population of Georgian-speaking Muslims: these are the *Ach'areli* (Russian *Adzhartsy*), about 150,000 of whom live in the Ach'ar autonomous republic on the Black Sea coast near the Turkish border. Their religious culture has been Turkish-speaking, and many Ach'ar are bilingual in Turkish. There are at least 50,000 Ach'ar in Turkey: their ancestors fled Christian Georgia in fear of religious persecution in 1877–8.

Georgian shows the influence of Greek, Turkish, Armenian, Persian and Russian. It remains a typically Caucasian language: although the Caucasian language families have not yet been shown to be related to one another, they share such features as a very large inventory of consonants and a complicated 'agglutinative' word structure.

For a table of numerals see MINGRELIAN.

The Kartvelian languages

Georgian is the national language of Georgia. The *K'art'li* dialect of Tbilisi and its region is the basis of standard Georgian. The western dialects, including *Ach'ar*, show the influence of MINGRE-LIAN and Laz, two related languages, which are spoken at the western extremities of the Georgian-speaking area.

Svan has about 40,000 speakers in Svanetia, the sequence of inaccessible high mountain valleys whose streams feed the Inguri and Kodori rivers, in north-western Georgia. Related to the other Kartvelian languages, Svan has probably been growing apart from them for over three thousand years.

Mxedruli: the Georgian alphabet

Mxedruli gradually came into use for Georgian between the 9th and 13th centuries. It replaced an older script, quite different in shape, now known as *asomtavruli*, which is first found in a

stone inscription of AD 430 and was clearly invented at about the time of the conversion of Georgia to Christianity in the 4th century.

The alphabetical order is that of Greek – but the additional consonant sounds of Georgian have been fitted in, with newly invented letters, towards the end of the sequence. The alphabet still begins with *a* (Greek *alpha*) and ends with o (Greek *omega*), though the last letter is now obsolete.

Mxedruli

ა ბ გ დ ე ვ ზ წ თ ი კ ლ მ ნ ჲ ო პ ჟ რ ს ტ ჳ უ ფ ქ ღ ყ შ ჩ ც ძ წ ჭ ხ ჴ ჯ ჰ ჵ

a b g d e v z ē t i ḳ l m n y o ṗ ž r s ṭ ü u p k ğ q̊ š č c j ç č̣ x q ǰ h ō

GERMAN

120,000,000 SPEAKERS

Germany, Austria, Switzerland, Russia, Kazakhstan, Romania, United States

One of the GERMANIC LANGUAGES, German is in origin the language of the tribes that stayed put – east of the Rhine, north of the upper Danube – while Germanic mercenaries and war-lords were parcelling out the Roman Empire among themselves between AD 300 and 500.

German has many different names. *Germania* was the Roman name for the country, and *Germani* for its people. They now call them-selves and their language *Deutsch* – a word first used in medieval Latin documents of the 8th century in the rather different form *theodisca lingua*, 'language of the tribes'. *Deutsch* is also the origin of the English word DUTCH.

In French, Germany is named *Allemagne* and its language *Allemand*, after the *Alemanni*, one of the prominent tribal groups of the 4th century AD. The usual eastern European term (Czech *Němec*, Romanian *Neamţ*) comes from a Slavonic word meaning 'mute', i.e. 'not speaking our language'.

The first records of a German language come from the 8th century. They amount to the so-called *Song of Hildebrand* (a few lines of an epic poem), some magical charms, and a collection of German glosses – translations between the lines – in Latin manuscripts. There is even a short Latin–German dictionary, now called *Abrogans* from its first Latin word, written in the 760s. But the usual written language in Germany, for many centuries yet, was Latin, and students of 'Old High German' (700–1100) have to study a fairly small number of texts. The emperor Charlemagne, around 800, ordered the ancient German epics to be written down in a book – but, if this was ever done, the book has long ago disappeared.

The Strasbourg Oaths

In AD 842 an alliance between Charles the Bald and Louis the German was sealed by the taking of mutual oaths by the French and German monarchs and by their armies. So that everyone could understand what was going on, the oaths were taken in the everyday languages of the two countries, and not in Latin. The official chronicler Nithard recorded the proceedings word for word. Charles took the oath in Ger-man:

In Godes minna ind in thes christianes folches ind unser bedhero gehaltnissi, for thesemo dage frammordes, so fram so mir Got gewizci indi mahd furgibit, so haldih thesan minan bruodher, soso man mit rehtu sinan bruher scal, in thiu thaz er mig so sama duo, indi mit Ludheren in nohheiniu thing ne gegango, the, minan willon, imo ce scadhen werdhen.

For God's love and for the salvation of the Christian people and ourselves, from this day onward, so far as God gives me knowledge and power, I will aid this my brother Charles, as a man in justice ought to do for his brother, so long as he does the same to me, and I will enter no talks with Lothar which might, by my will, damage him.

For the French text as sworn by Louis, see box at ROMANCE LANGUAGES.

Nithard, *Histoire des fils de Louis le Pieux* ed. Ph. Lauer (Paris: Les Belles Lettres, 1926) p. 106

High German was the language (already divided into regional dialects) of southern and central Germany. Northern Germany spoke Low German, which was already a distinct language.

High German first flourished as a language of written literature in the 12th and 13th centuries. Lyric poets – the *Minnesinger*, 'singers of love' – wrote songs of great beauty and freshness, inspired by OCCITAN poetry. Longer poems were also written, from the epic *Nibelungenlied* 'Song of the Nibelungs', which drew on German legendary history, to romances such as Gottfried von Strassburg's *Tristan und Iseult*, modelled on French originals. Although Germany was a far from united country, made up of large and small principalities, the poetry composed at numerous scattered courts was in a surprisingly uniform dialect, now known as the *mittelhochdeutsche Dichtersprache* 'Middle High German poetic language', which must have been maintained by travel and mutual influence among poets and audiences, since for official purposes the courts still used Latin.

There is no agreement about when the Middle High German period ends. The courtly poetry ceased to be written, but local governments gradually began to use German more frequently in official documents. This tended to be a regional language; but some courts began consciously to shape *das gemeine Deutsch* 'common German', and with the spread of printing, from 1445 onwards, there was every reason for a common language to emerge.

There is no doubt at all of the landmark that stands at the beginning of the Modern High German literary standard. This was Martin Luther's German translation of the Bible, completed in 1534. It had a huge sale in all parts of Germany, and its language – inspired by spoken German and partly shaped by the written language of the Royal Court of Saxony – was repeated in religious services, was avidly read and became the universally accepted model for German prose. It was perhaps Luther's Bible that sounded the death-knell of Low German as a literary language (see box).

Swiss German

Swiss German (*Schweizerdeutsch* or, locally, *Schwyzerdütsch*) has over 4,000,000 speakers. It is the everyday spoken language of the majority of Swiss – but it is not normally used for literature, though some well-known writers, such as Jeremias Gotthelf (1797–1854) have written their work in Swiss German.

Switzerland has four national languages, one of which is German – but it is standard German. The language of the German literature, newspapers and official documents of Switzerland is very little different from that of Germany.

The lands in which German was spoken remained politically disunited. They still are. The modern colloquial of Vienna, capital of Austria – and once the capital of an empire stretching across south-eastern Europe – is very different from the colloquial of Berlin. Munich, Hamburg and other German cities can claim a cultural life, and, naturally, a colloquial dialect, quite independent of Berlin's. German is also the majority language of Switzerland, and Swiss German is different again. Regional German has its distinctive accents and also has a surprising amount of variation in vocabulary. Some of this is visible in the written German of authors from the various regions; yet true dialect writing is almost as marginal a literary activity in German as it is in English.

Although English and German are quite closely related, at an English reader's quick glance a German text is not especially easy to understand. Nearly all the words look unfamiliar. There are two reasons for this. The High German Sound Shift, which (linguists have concluded) happened before the earliest German written texts, altered the consonant sounds in the German form of many common words that German and English share: *help*, *helfen*; *make*, *machen*; *apple*, *Apfel*. The other reason is that German does not borrow freely from Latin, whereas English has borrowed very heavily from both Old French and Latin. The typical way to make new words in Ger-

Europe and Central Asia: modern and [recent] German-speaking areas
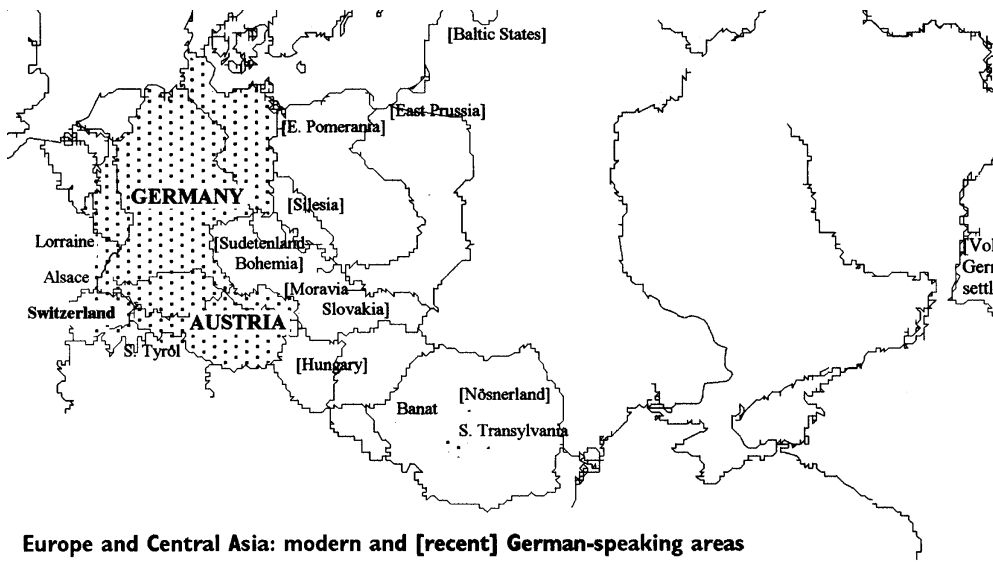
man has been not by borrowing but by compounding: English *television* (invented from Greek and Latin words meaning 'far' and 'vision') corresponds to German *Fernseher* (a purely German compound meaning 'far-seer'). However, this is now changing, and French and English words are multiplying in modern spoken German.

Written German looks unfamiliar for two other reasons. In German every noun is given a capital letter, while adjectives based on proper names have no capitals: *das deutsche Buch* 'the German book'. And while English makes noun phrases by chaining nouns together, or by using an adjective, German very often makes the chain into a single word. These very long words are a feature of technical writing and official documents: *Orangensaftkonzentrat* 'concentrated orange juice', and they can demand to be abbreviated, as *Nationalsozialistische* 'National Socialist' was abridged to *Nazi*.

The varieties of German

The far west

There are some extensive communities of German speakers in the rural United States, totalling at least 300,000 speakers. The majority are in south-eastern Pennsylvania.

Their language is locally *Pennsilfaanisch* or in full *Pennsilfaanisch-Deitsch*, hence the English name *Pennsylvania Dutch* – but it is a dialect of German, not of Dutch. Settlers came in 1683 onwards, mostly from the Rhenish Palatinate and Switzerland. Their first settlement was Germantown, Pennsylvania. Until about 1830 German speakers were in a majority hereabouts. They still form a close-knit community, held together by their religious beliefs: hence their dialect survives. It continues to show resemblances to that of the Palatinate, but also heavy American English influence: *büssig* 'busy' (standard German *geschäftig*); *Was zeit iss 's* 'What time is it?' (standard German *Wieviel Uhr ist es?*).

In the seven villages of Amana in Iowa, trilingual descendants of a Pietist community established in 1855 speak *Kolnie-Deitsch* ('colony German') as well as standard German and standard American English (which they typically speak with a slight distinct accent). In familiar conversation the local German is mixed with English words: *Ich pick dich uf! Hast du dein Test gepassed?* 'I'll pick you up!' 'Have you passed your test?'

mus' based on Afrikaans *seekoei*; *Armstuhl* 'arm-chair'.

Examples from Claus Jürgen Hutterer, *Die germanischen Sprachen* (Budapest, 1975) pp. 342–7; Philip E. Webber, *Kolonie-Deutsch: life and language in Amana* (Ames: Iowa State University Press, 1993); Elizabeth De Kadt in *Language and social history: studies in South African sociolinguistics* ed. Rajend Mesthrie (Cape Town: David Philip, 1995) pp. 107–15

The dialects of High and Low German

Low German is divided from High German by an 'isogloss', a line on a dialect map – or rather by a bundle of isoglosses, all following roughly the same path. The most important of them is the northern limit of the area in which the High German Sound Shift took effect.

KAZAKHSTAN

lga
man
ements]

The south

German also spread to other continents. Pidgin varieties of German existed briefly in short-lived German colonies such as North West New Guinea, Tanganyika, South West Africa, Togo and Cameroon, though they never displaced the already established lingua francas of these countries. German influence can be traced in TOK PISIN, the national language of Papua New Guinea: *mak* 'shilling' from German *Mark*; *gumi* 'rubber'.

Some Germans settled in these colonies: others emigrated to Australia, Chile, Argentina and Brazil, in all of which cohesive but small German-speaking communities still exist.

At the end of the 18th century more than half the European population of the Cape Town region was German in origin, and there are still about 40,000 German speakers in South Africa, with some German schools and churches, most of the latter originating in the 19th-century Lutheran missions in Natal. Most immigrants probably spoke varieties of Low German – but modern *Springbockdeutsch*, thanks to German schooling, has turned into a regional variant of standard High German. It has numerous loanwords and calques from Afrikaans and English: *Fenz* 'fence'; *Seekuh* 'hippopota-

Low German

Almost as old as the oldest literature in German are two poems in what is now called Old Saxon – the language of northern Germany in the 9th century AD. They are called *Heliand* (' The Healer') and *Genesis*, and were composed at the request of King Louis the Pious, son of Charlemagne. Even in those days Old Saxon was quite distinct from the 'High German' of the south.

In its later forms the speech of the lower Rhine and the north German plains is known as *Plattdeutsch, Low German* (because it is the language of the low-lying country). Until the 16th century it was used in official documents by northern courts, and in particular in the business of the Hansa cities, which dominated Baltic trade. Low German influence can be traced in Norwegian, Danish, Swedish, Latvian and other languages of northern Europe. It might easily have become the standard dialect, the German that everybody learnt, if German cultural history had taken a different course – if medieval German courts had not patronised the poetry of the south, and if Luther's Bible translation had not dominated writing and literature in the 16th century. But it was not to be: for centuries now the Low German dialects have been seen as homely, rustic

forms of speech, seldom written down except in folk tales and consciously 'folkloric' literature.

With one exception. DUTCH is in origin one of the Low German dialects: for centuries, Dutch has counted as a language in its own right.

Why did Louis the Pious commission *Heliand* and *Genesis*? 'He instructed a certain Saxon, considered by his own people no mean seer, to turn the Old and New Testaments into German poetry, so that the lessons of the Divine Teachings should be revealed not only to those of his subjects who could read but also to those who could not . . . The stories were expressed in German of such clarity and elegance that all who listened to them took no little pleasure in their style' (*Heliand*, Latin prose preface).

Low German consists of several quite distinct dialects, some of which were used as written languages in late medieval times. None ever became a standard language for the whole area. The main modern dialects are DUTCH (or Low Franconian), *Niedersächsisch* or Low Saxon and *Ostniederdeutsch* or East Low German. Where the isogloss bundle splays out – in real human terms this is along the Rhine valley, where north–south communication encouraged dialect mixture – some linguists distinguish two West Middle German dialects. These are *Mittelfränkisch* or Middle Franconian (which includes the Moselle Valley and LUXEMBURGISH) and *Rheinfränkisch* of the Palatinate. These are important dialects because they formed the basis for others: Pennsylvania Dutch; Transylvanian Saxon; Volga German; and also YIDDISH, the language of German-speaking Jews, a language that gradually spread from the medieval Rhine across large areas of eastern Europe.

High German dialects include Thuringian and Upper Saxon (but these are sometimes grouped as 'East Middle German'), Upper Franconian, Alemannic (including Swiss German), Bavarian and Austrian. Both Alemannic and Austrian dialects extend southwards across the Italian border.

The further east

The geographical range of Yiddish was merely one result of the vast, long-lasting migration that left central and eastern Europe, at the beginning of the 20th century, studded with large and small German-speaking communities of farmers, miners and others. There were heavy concentrations of German speakers in Silesia and Pomerania, now Polish-speaking; from Pomerania German speech extended by way of Danzig eastwards to East Prussia (now divided between Poland and Russia) and into the Baltic states.

Halbdeutsch 'Half-German' was the simplified German that was spoken by Estonians and Latvians who settled in the German-dominated towns of that region, until Russian influence increased in the 19th century.

The Sudetenland, the regions of Bohemia (now the Czech Republic) bordering on Germany, had a German-speaking majority. Unusually extensive also were the German settlements in Transylvania (*Siebenbürgen*: now in Romania) and on the lower Volga in southern Russia.

As the ruling language of the German and of the Austro-Hungarian Empire, the language of their armies, and the vehicle of higher education, science and scholarship, German was uniquely respected. Thus, however isolated, German settlements tended to retain their language for centuries.

The fate of German in the east

The Second World War was fomented by nationalist agitation concerning the German-speaking populations of Czechoslovakia and Poland. Ironically, in the course of the war eastern European speakers of the Yiddish form of German were almost completely wiped out as a matter of German policy. The outcome of the war was a catastrophe for most of the remaining German-speaking communities of eastern Europe. Many died; most others were uprooted. Many of these, as refugees in Germany, have learnt to lose their local dialects and culture.

Germans in the Baltic republics were moved en masse to Germany in 1939 after the German–Soviet Pact placed these republics in the Soviet sphere of influence.

German rural settlement in Russia itself had been concentrated in the lower Volga valley around Saratov. The Volga German Autonomous Republic, established soon after the Revolution,

was broken up on the German invasion of the Soviet Union in 1940 and its whole German-speaking population exiled to Kazakhstan and Siberia. There were still about a million German speakers in Kazakhstan, and over 800,000 in Russia (mostly in Siberia), in 1989; the numbers are steadily declining as German speakers are allowed to emigrate to Germany.

In 1945 the German population was expelled from East Prussia, which was divided between Poland and the Soviet Union. German speakers were also expelled en masse from western Poland up to the new Oder-Neisse border, and from Sudetenland.

Numerals in German and Luxemburgish

German		Luxemburgish
eins	1	eent
zwei	2	zwee, zwou
drei	3	dräi
vier	4	véier
fünf	5	fënnef
sechs	6	sechs
sieben	7	siwen
acht	8	aacht
neun	9	néng
zehn	10	zéng

German fonts

Printing in German

The two 'Gothic' font families shown in the box are known as *Fraktur* (lines 1–2) and *Schwabacher* (lines 3–4). Until the 1930s they were typical of German language printing. When quoting single words or longer texts in other languages, printers would normally switch to a Roman font.

Research showed that reading in Fraktur was significantly slower than in Roman, and by a decision of the National Socialist government the Gothic fonts were abandoned. Younger Germans now find pre-war printed books difficult to read.

GERMANIC LANGUAGES

Among the language groups that derive from proto-Indo-European, the Germanic languages have a very clear geographical base and several features that distinguish them from their relatives. English and German are among the major modern languages belonging to this group.

Germanic has long been the international term for this language family: it derives from the usual Latin name, *Germani*, for a group of peoples coinciding more or less with the speakers of early Germanic dialects. It is not a native but a Latin word: it meant 'relatives'. *Teutonic* has occasionally been used as an alternative: it is the name of one tribal group at the same period – but it is quite possible that this group spoke a Celtic language! Another alternative is *Gothic*, outdated in English (where it is the name of one particular, extinct, Germanic language) but, as *gottonsk*, favoured by some Scandinavian scholars.

Proto-Germanic, ancestor of all the modern Germanic languages, must have been spoken on the shores of the western Baltic in the last few centuries BC. There are no records of it, but Germanic scholars have been able to reconstruct its forms, probably with fair accuracy, through working backwards from the known Germanic languages of later periods. Additional evidence comes from loanwords in Finnish and the Baltic languages.

The early records of Germanic languages are inscriptions on stone and wood in Runic script. The very oldest Runic inscriptions, such as that on the horn discovered at Gallehus, go back to AD 200 or soon after, and they are very close indeed to the reconstructed proto-Germanic. A slightly later example appears in the box at OLD NORSE.

The Germanic languages in 1000 BC

In the first millennium BC the early Germanic dialects had been surrounded by other INDO-EUROPEAN LANGUAGES, Celtic, Baltic and perhaps Illyrian and Slavonic. To the north, the North Germanic dialect (prehistoric OLD NORSE) may already have been close to areas of SAMI and FINNISH speech.

Some important words were borrowed into Germanic from neighbouring languages: they include modern German *Reich*, Dutch *rijk* 'kingdom', a Celtic word in origin.

Germanic languages in Europe today

Icelandic

NORTH GERMANIC

Faroese

Norwegian

Swedish

Danish

WEST GERMANIC

Frisian

English   Low German

Dutch

German   SLAVONIC LANGUAGES

ROMANCE LANGUAGES

Germanic languages about AD 500

The Germanic languages in AD 500

By AD 500, East Germanic (GOTHIC) was being carried through the Roman Empire as a language of mercenary troops who were already beginning to conquer territory for themselves. Many West Germanic speakers, who included Angles, Saxons, Burgundians and Franks as well as the direct forerunners of modern Dutch and Germans, were also within the boundaries of the Roman Empire. Both these branches of Germanic show heavy influence from Latin. One early Latin loanword surviving in modern Germanic languages is Latin *vinum*, German *Wein*, English *wine* – but the name of a different fermented drink is inherited directly from Indo-European, German *Met*, English *mead*.

Speakers of North Germanic would soon begin their expansion across the north Atlantic to the Scottish and Irish coasts, the Isle of Man, the Faroes, Iceland and Greenland.

The Germanic languages today

The East Germanic branch has died out. To West Germanic belong GERMAN, YIDDISH, DUTCH, LUXEMBURGISH, FRISIAN, ENGLISH and the ENGLISH CREOLES AND PIDGINS. To North Germanic belong SWEDISH, DANISH, NORWEGIAN, FAROESE and ICELANDIC.

GILAKI

PERHAPS 4,250,000 SPEAKERS OF 'CASPIAN DIALECTS'

Iran, Azerbaijan

The 'Caspian dialects' are IRANIAN LANGUAGES quite distinct from Persian, spoken in northern Iran and Azerbaijan near the shores of the Caspian. None of them is recognised as an official language: many younger speakers are thus bilingual in Azeri or in Persian.

Some 16th-century poetry from Ardebil, in north-western Iran, is in a dialect very close to modern Talysh (see map). There is old poetry in Mazandarani too, by Amir Pazvari and others. In general, however, very little has been written in any of the Caspian dialects.

With the growth of education and press in Persian, these survivals of the very early spread and differentiation of Iranian languages are likely to disappear.

In Gilaki the first four numerals are *i, do, sə, čār*. In Mazandarani they are *atto, detto, se, čār*. The Caspian dialects retain more than Persian does of the noun declension system that was characteristic of older Iranian.

The Caspian and north-western dialects: the map

Talysh or *Taleshi* is the language of the Caspian coastal plain in southern Azerbaijan and north-west Iran. There may be 250,000 speakers.

Gilaki is spoken along the western half of the Caspian's southern coastline. Its natural centre is the city of Rasht. One estimate suggests 2,000,000 speakers.

Mazandarani borders the south-east Caspian, and includes *Gorgani*. There may be 1,500,000 speakers. The Elburz mountains mark the southern limit of Gilaki and Mazandarani.

Scattered among speakers of Turkic AZERI, in inland north-western Iran, are the Iranian dialects known confusingly as *Azari* or *Tati* (see box at PERSIAN for the other language known by this name). They include *Takestani*. There are perhaps 250,000 speakers.

Semnani is the language of the city of Semnan, east of Tehran.

There are further dialects of central Iran, insufficiently known, gradually retreating before the advance of Persian.

Iran: Caspian dialects

GONDI

3,000,000 SPEAKERS OF CENTRAL DRAVIDIAN LANGUAGES

India

The family of DRAVIDIAN LANGUAGES includes not only the state languages of southern India, but also a number of minority languages of hill and jungle peoples scattered over much of the subcontinent. *Gōṇḍī* is the most important of the Central Dravidian group, which must have separated from southern languages such as Kannada and Tamil well over two thousand years ago.

The hills and jungles where the Gōṇḍ tribes live were once called after them Gondwana (Sanskrit *Goṇḍavana* 'forest of the Gonds') – a name which, before it was forgotten, was borrowed for a geological stratum typical of the region, and then borrowed again by the proponents of the geological theory of continental drift. So Gondwanaland is now the agreed name of the supercontinent that existed hundreds of millions of years ago before South America split from Africa, Antarctica, Australia, Arabia and India. Most Gond tribes call themselves *Kōī*, *Kōītōr* and other similar names.

There is no written literature in Gōṇḍī. The language is spoken in several large enclaves of jungle country, separated from one another by settled lands where Marathi, Chhattisgarhi and Telugu are spoken. There is no doubt that some of these settlements are relatively recent: Gōṇḍī speakers, as they settle to agriculture, turn to speaking these languages of 'civilisation'. Gōṇḍī clearly forms the linguistic substrate underlying the distinctive features of the Bhatrī dialect of ORIYA and the Halbī dialect of MARATHI.

In these languages the three-way gender distinction of early Dravidian has given way to a two-way distinction between masculine (male humans) and non-masculine (all other nouns).

The first ten numerals in Gōṇḍī are: *undī*, *rend*, *mūnd*, *nālūng*, *siyāng*, *sārūng*, *ērūng*, *armur*, *anma*, *putth*. For the higher numerals – from '8' up – borrowed Marathi forms are usually used.

The Central Dravidian languages

The islands of *Gōṇḍī* speech, with about 1,900,000 speakers, are in the Indian states of Maharashtra, Madhya Pradesh and Andhra Pradesh. About 5,000,000 people in India class themselves as Koitor or Gonds, but many speak Hindi or Telugu and not Gōṇḍī. The Muria Gonds of the Bastar district are well known among anthropologists for the *ghotul*, the young peoples' dormitory, in which sexual promiscuity before marriage is encouraged.

Kui (500,000 speakers) and *Kuvi* (300,000 speakers), similar to each other and fairly close linguistically to Gōṇḍī, are spoken in a hill and jungle region of southern Orissa state.

Kōlāmī (90,000 speakers) and *Parjī* (or Duruwa, 90,000 speakers) form a separate subgroup, spoken in hill districts to the south of the Gōṇḍī-Kui area.

Central Dravidian languages
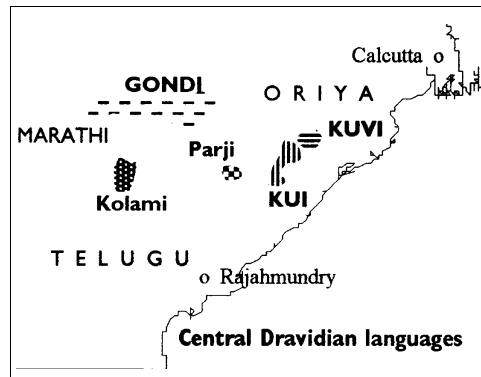
GONGA LANGUAGES

PERHAPS 200,000 SPEAKERS

Ethiopia

A group of now-scattered languages belonging to the OMOTIC group of Afroasiatic languages, Gonga is the speech of a series of old kingdoms of south-western Ethiopia.

> The best-known kingdom, the best-known language and a popular name for the whole group is *Kefa* or *Kaffa* (the double *f* shows this to be an Amharic form). However, *Gonga* is a name known ever since the oldest reports of travellers to the region.

Christianity was introduced to Kefa and other Gonga kingdoms from Ethiopia about 1590. In the early 17th century, when Europeans first travelled through the region, the Gonga kingdoms were already under attack from Oromo-speaking invaders. In the late 19th century they were conquered by the Ethiopian Empire of Menelek. Meanwhile, in 1855 a Roman Catholic mission entered the religious scene. An early wordlist of Gonga had been published by the explorer Charles Tilstone Beke in 1846.

> Kefa is remarkable among minor African languages in having had a detailed study made of its erotic vocabulary – could this have been a special enthusiasm of the Austrian explorer F. J. Bieber? He published his work in the pioneering journal of sexology, *Anthropophyteia*, in 1903.

Gonga languages have two or three tones, and also distinguish vowel length and syllable stress. Already in the 17th century Gonga languages included Amharic loanwords in fields such as politics and religion: the kingdoms were then self-governing but tributary to the King of Ethiopia.

The major Omotic languages

The Gonga languages include Kefa, Moca, Anfillo and Shinasha. *Kefa* is the language of the old kingdom of Kefa. *Moca* or Mocha is spoken to the west of Kefa. *Anfillo* or Southern Mao, a dying language, is spoken in Anfillo forest in an area now predominantly Oromo-speaking. *Shinasha* or North Gonga or simply Gonga is a group of languages spoken in the Blue Nile valley.

The Ometo languages include *Basketto*, of the east bank of the Omo; *Kullo*, the dialect of Jimma and its neighbourhood; and, WOLAYTTA or Welamo, the main dialect and now the basis of a literary language.

Ethiopia: Omotic languages

GOTHIC

EXTINCT LANGUAGE OF EUROPE

Gothic is the only member of the East Germanic branch of the Germanic languages in which any continuous texts are recorded.

The name of the Goths can still be traced in their earliest recorded territory, southern Sweden and the Baltic shores – it survives in the island of *Gotland* and the Swedish districts of *Götland*. The Goths of Roman history were divided into two 'tribes', *Ostrogothi* or *Greutungi* 'dune-dwellers' and *Visigothi* or *Tervingi* 'steppe-dwellers'.

Three peoples whose dialects were very close to one another – Goths, Burgundians and Vandals – were prominent in the invasions and mercenary warfare that marked the decline of the Western Roman Empire, between about AD 350 and 500 (see maps at GERMANIC LANGUAGES).

The major Gothic text is a translation of the Bible, the work of Bishop Ulfilas (*Wulfila*), the religious leader of the Visigoths when they first settled within Roman borders, in what is now Bulgaria, after 348. It is the earliest literary text in any Germanic language.

The Visigoths eventually established a kingdom in 5th-century Spain: their language had some influence on the late Latin of Spain, and thus eventually on Spanish. The Ostrogoths – who also used Ulfilas' Bible translation – ruled Italy about the same time, and again some Gothic vocabulary can be traced in modern Italian. The name of the French province of Burgundy, *Bourgogne*, is the last reminiscence of the Burgundians, who settled in that region. The Vandals briefly established an empire in North Africa; their earlier dominion in southern Spain was still remembered when the Arabs conquered the region and named it, after them, *al-Andalus* (modern Spanish *Andalucia*).

Some Goths, meanwhile, remained on the shores of the Black Sea, where their migrations had taken them before they entered the Roman orbit. Until the 15th century there was a Gothic principality in the Crimea. Even in the 16th and 17th centuries, European travellers were surprised to find *Crimean Gothic* still spoken by small communities. The language is now extinct.

The first ten numerals in Gothic are: ᚨᛁᚾᛋ ᛏᚹᚨᛁ ᚦᚱᛁᚷᚨ ᚠᛁᚨ�Yᚱᚱ ᚠᛁᛗᚠ ᛋᚨᛁᚺᛋ ᛋᛁᛒᚢᚾ ᚨᚺᛏᚨᚾ ᚾᛁᚢᚾ ᛏᚨᛁᚺᚢᚾ or, in transliteration, *ains*, *twai*, *thrija*, *fidwor*, *fimf*, *saihs*, *sibun*, *ahtau*, *niun*, *taihun*.

The Gothic alphabet

ᚨ ʙ ᚷ ᚷ ᛖ ᚢ z ʜ ᚦ ɪ ᚴ ᚦ ᛖ ᛖ ɴ ɢ ᚾ ᚾ ʏ ᚴ s ᛏ
ʏ ᚠ x ᚠ ᛖ ↑

a b g d e q z h th i k l m n j u p – r s t w f ch wh o –
1 2 3 4 5 6 7 8 9 10 20 30 40 50 60 70 80 90
100 200 300 400 500 600 700 800 900

Font: *gothicl. ttf* from
'Dr Berlin's Foreign Font Archive'

The Gothic script

'Gothic' is a word that has been used with many meanings, in history, in art, in architecture and in typography. Quite different from the black-letter scripts of medieval Europe, and the Fraktur typefaces of German printing, both of which have sometimes been called 'Gothic', what is shown in the box is the alphabet invented for the Gothic language by Ulfilas in the 4th century. It is based on Greek scripts of his day (with added letters from Latin and Runic alphabets). Just as in Greek, the 25 letters of the alphabet, with two special added characters ᚴ and ↑, served also as numerals from '1' to '900'.

GREEK

11,500,000 SPEAKERS

Greece and Cyprus

Greek is an INDO-EUROPEAN language, the sole descendant of one of the original dialects of proto-Indo-European. In its mountainous homeland in south-eastern Europe, from 700 to 300 BC and after, ancient Greek was the language of philosophers, historians, poets and playwrights whose works have been read ever since. Greek has had a continuous and eventful history from that time to the present day.

> Greek is known to its own speakers as *Elinikí* (the word is borrowed from ancient Greek *Helleniké*). This name is formed from the native name of the country, ancient *Héllas*, modern *Elládha*. The traditional term for Greek in medieval and early modern times was *Romaikí*, as the language of the 'Romans' of the later empire: fewer speakers use this term now. Most foreign names for the language and the country derive from the Latin word *Graecus*, 'Greek'.

3,300 Years of History

Greek was first recorded about 1300 BC on clay tablets in Linear B script (see box). Then there is a six-hundred-year gap in the record. Greek has an uninterrupted recorded history from about 700 BC – the date of the earliest surviving texts in the Greek alphabet – down to modern times.

By that time it was the majority language of the Greek peninsula and islands and a major language of Cyprus, and it has held these positions ever since. It was the language of scores of independent Greek cities founded along the Mediterranean and Black Sea coasts in the cen-

turies after 750 BC. It was the ruling language of the great 'Hellenistic' kingdoms of the Near East that lasted for several centuries after the death of Alexander the Great in 323 BC and stretched as far as the Indus valley. It was the lingua franca of the whole eastern half of the Roman Empire.

Periods of Greek

Mycenaean: about 1300 BC
Ancient: 700 BC to AD 500
(Archaic, 700 to 500 BC; Classical, 500 to 300;
Hellenistic, 300 to 1; Roman, AD 1 to 500)
Byzantine: 500 to 1450
(early Byzantine, 500 to 1100; late Byzantine:
1100 to 1450)
Turkish and Venetian: 1450 to 1800
Modern: from 1800

Even after the long domination of the Romans, Greek eventually supplanted Latin, and became once more the single ruling language of the 'Byzantine' Empire, which fell to the Ottoman Turks in AD 1453. Greek was at first the language of diplomacy for the Seljuk and Ottoman Turkish kingdoms. Greeks retained the status of an autonomous minority, *millet*, under the Ottoman Empire. Many ancient communities of Greek speakers, all over the empire and especially in the coastal regions of Anatolia, survived and prospered. Greek retained wide importance as the language of the Greek Orthodox Church; Greek literature flourished in Venetian-ruled Crete, and Greek culture and education prospered in Bucharest (Romania).

What was once a very large minority in Asia Minor, numbered in millions, declined rapidly with the 'population exchanges' that followed the disastrous Greek invasion of Turkey in 1923. Although the balance was redressed slightly by Greek-speaking Muslims who were expelled from Greece at the same time, the total Greek-speaking community in Turkey now is estimated at well under 50,000.

> The refugees from Greece to Turkey in 1923 included *Valakhádhes*, people of AROMUNIAN origin who had become town-dwellers, adopting Islam and the Greek language. Their dialect is unlikely to survive long. The same is true of several Greek communities of Asia Minor itself that adopted Islam in Ottoman times – they look on themselves as Turks and did not migrate in 1923.

The changing Greek language

There are great differences between Mycenaean and ancient Greek – differences that are only underlined by the utterly different writing systems. There is no surviving Mycenaean literature: the clay tablets are accounts and inventories. When ancient Greek emerges, around 700 BC, it is in the form of the great epics attributed to Homer, the *Iliad* and the *Odyssey*, which have influenced European literature ever afterwards.

Ancient Greek is recorded in several dialects (see map), but the Attic dialect of Athens of the 4th century BC gained cultural prestige and became a standard, adopted in the new Hellenistic monarchies of the Near East and, gradually, in Greece itself. Classical Attic continued to be the usual language of education and written culture until almost AD 1500. Its derivative, the koiné ('common') dialect, was the ordinary spoken language of Hellenistic and Roman times: the Hebrew Old Testament was translated into the developing koiné of 250 BC and the New Testament was written in the koiné of about AD 100.

Yet almost at the same date Lucian was writing satirical pieces and Plutarch was composing his biographies of Greeks and Romans, both in skilful pastiches of classical Attic. Meanwhile, narrative poetry – from the limpid, romantic *Hero and Leander* of Musaeus to the florid *Dionysiaca* by Nonnus, both written around AD 500 – would still be composed in a metre and a dialect close to those of the *Iliad* and *Odyssey*, which were by then twelve hundred years old.

'The Greeks had a word for it'

The proverb is a reminder of the richness of the Greek vocabulary: Liddell and Scott's *Greek–English Lexicon* has about 300,000 entries. Like modern German, ancient Greek was a language in which any speaker or writer might invent new compound words with the confidence that they would be understood.

In the Byzantine Empire, too, it came naturally to the educated to write in a kind of classical Attic, though writings for a less courtly milieu, such as saints' lives, were in a language closer to that of every day.

Through the Roman and Byzantine centuries the development of the spoken language can be traced in informal writings (for example, in papyri from Egypt). At last satire and fiction in a recognizably modern form of Greek emerge around 1150, throwing off the classical and ecclesiastical straitjackets. At the centre of the linguistic revolution, it has been argued, were French and German empresses and princesses, familiar with the new romantic literature of their home countries. Since that time, creative writers have increasingly used the language of every day.

Modern Greek looks very like ancient Greek, but it probably sounds very different. Greek as spoken in classical Athens in 400 BC had a front-rounded vowel like German *ü*, several diphthongs (*ai, au, eu, oi, ui*), aspirated stops (*ph, th, kh*, like those of Hindi), a distinction between long and short vowels, and, perhaps most important of all, a distinction between high and low pitch, which must have given a musical quality to the spoken language comparable to

that of Welsh. All these features disappeared in the centuries after Alexander. High pitch was replaced by syllable stress, and many changes occurred in vowel sounds. Modern spoken Greek has 5 vowels and, by one count, 22 consonants.

The basic vocabulary of Greek was inherited from Indo-European. The names of cultural innovations were borrowed, often from languages to the south and east: ancient *oînos* 'wine', *asáminthos* 'bathtub', *tyrannos* 'dictator'. Names of native plants and animals come from some other source, presumably from earlier languages of Greece or Turkey now lost: *amygdále* 'almond', *koríannon* 'coriander', *sepía* 'cuttlefish'. In later times Greek has borrowed words from Latin, French, Turkish and English: *tavérna* 'restaurant', *kafé* 'coffee', *yaúrti* 'yoghurt', *flért* 'flirtation'.

A form of ancient Greek was revived in the 19th century in the stiff and artificial katharé-vousa, the 'purified' language of modern officialdom, occasionally reimposed in schools by right-wing governments. It is used in some technical writing but not in literature, which is written in dhimotikí, the demotic or 'popular' tongue.

Greek in the wider world

Greek was one of the two great languages of the Roman Empire, very widely used among both the upper and the lower classes: it was the first or second language of a large proportion of slaves. Any educated Roman, between about 100 BC and AD 300, was expected to know Greek. Many Romans studied in Greek 'university' cities, especially Athens. Greek had a powerful influence on Latin and Arabic, the medieval languages of international learning and science. This influence continues today as ancient Greek words and affixes, often in Latinised forms, are borrowed and compounded to create new international technical terminology in pharmacology and medicine, in the sciences and social sciences (Greek *tékhne*, skill; *phármakon*, drug).
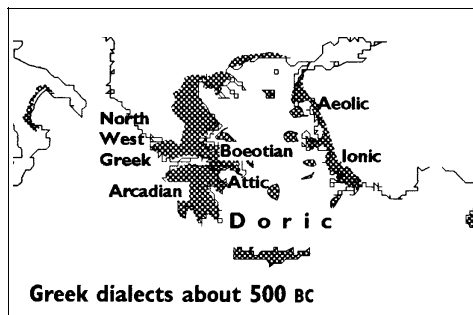
The majority language of Greece and Cyprus, Greek is spoken by minorities in Georgia, Egypt, Albania, southern Italy and many other countries. There were once flourishing Greek communities in Odessa (Ukraine), in Alexandria and Cairo (Egypt) and in the old Byzantine capital, Istanbul. These are now in decline: in recent years, particularly large Greek and Cypriot colonies have grown up in Chicago, London and elsewhere across the world. The Greek speakers of London, numbering well over 100,000, are mostly of Cypriot origin and came between 1955 and 1974 with the political upheavals that culminated in the Greek and Turkish interventions and the division of Cyprus.

The first ten numerals in modern Greek are: *ena, dhio, tria, tessera, pende, exe, efta, ohto, enya, dheka*.

Greek dialects: ancient and modern

During the early centuries of recorded Greek numerous dialects were spoken. Their mapping is a complex task as migrations and colonizations multiplied. At first no single dialect was standard. In official inscriptions recording laws and religious ceremonies, each city would use its own local dialect. So it was in literature, too, at first. The love poetry of Sappho was written in the Aeolic of Lesbos, her native island. For choral lyrics Doric seemed natural – the legacy of Alcman, who had written songs of matchless beauty for the girl choirs of Doric-speaking Sparta. Tragedy and comedy belonged to the Athenian stage, and were in Attic: so were the writings of the philosophers Plato and Aristotle, who taught in Athens. Medical treatises and notes, traditionally unsigned, were written in the Ionic dialect of Cos, even if Ionic did not

Greek dialects about 500 BC

Map: Greek in the 20th century. Labels include: Pontic, Pontic, [Pontic], BULGARIAN, ITALIAN, MACEDONIAN, ALBANIAN, Aromunian, Italiot, Istanbul, TURKISH, [Cappadocian], Tsakonian, GREEK, Cypriot.

come naturally to their authors: Cos was where Hippocrates, father of medicine, had lectured and practised around 500 BC. Thus the use of dialects easily slipped into a stylistic convention. Even the early epics, *Iliad* and *Odyssey*, are recorded in a mixed Ionic/Aeolic dialect which was no one's mother tongue.

The true ancient dialects disappeared as the koiné spread: they have only one modern survival, Tsakonian, descendant of the Laconian variety of Doric. But eventually dialect differences developed in the koiné itself, encouraged by political fragmentation as the Byzantine Empire declined and was extinguished. Modern standard Greek is based on the dialect of Athens. Cyprus has its own local standard. Modern dialects spoken outside Greece and Cyprus include Italiot and Pontic. Italiot is spoken in two small districts in southern Italy, south of Lecce in Apulia and near Bova in Calabria. Greeks had settled in southern Italy by the 7th century BC, but the Greek spoken there will have remained close to standard Greek until the Byzantine Empire lost its Italian foothold around AD 1100. Pontic and Cappadocian were spoken by Greeks who were expelled from northern and central Turkey in 1923, in the 'exchange of populations' that followed the disastrous Greek invasion of Asiatic Turkey. Pontic is still spoken by a few

communities in southern Russia, especially near Rostov. Again, the first settlements of Greeks in these regions took place in the 7th century BC and after: under the later Roman Empire almost all of Asia Minor (modern Turkey) was Greek-speaking. Cappadocian Greek communities were Turkish-ruled from 1071 onwards, and their dialect became strongly influenced by Turkish. Pontic grew apart from standard Greek with the decline of Byzantium and the establishment of the independent Empire of Trebizond after the Crusaders captured Constantinople in 1204.

Traditionally, students of Greek have been taught to speak and write classical Attic. An early textbook was composed around AD 170 by Julius Pollux for the future Roman emperor Commodus. Unfortunately, Commodus was more interested in ceramics, dancing and martial arts: 'subjects unsuitable for an emperor,' said his biographer.

In recent times, all over western Europe, Greek has been one of the two 'classical' languages of a full education, though Ben Jonson, a classical scholar, thought none the worse of his contemporary Shakespeare because the latter knew 'small Latin and less Greek'. The writing of classical Attic prose and verse is still a common exercise for students of classics at school and university.

Tsakonian: survival of an ancient dialect

Tsakonian is spoken in an inaccessible mountainous district of the Peloponnese (southern Greece), north of Leonidhi. It is the direct descendant of the Doric dialect of ancient Greek. The link can be seen in the preservation of the old long *a* sound so typical of Doric: Attic *mêlon*, modern standard Greek *mílo*, Doric *mâlon*, Tsakonian *máli* 'apple'. Development in isolation has made Tsakonian very different from modern Greek.

The Tsakonian-speaking villages were listed in a local song:

Alepú tta Síkina	Foxes at Sitena
ke liúko tta Kasténitsa	and wolves at Kastenitsa
ke kukuváya ttom Prasté,	and owls at Prastos –
ke més ta Meliyítika	and up at Meligou
pidhúnde fardakláne.	it's the frogs that dance.

Greek in Writing: 1. the Linear B script

Michael Ventris (1922–56) was an English architect with a flair for languages and cryptography. His discovery that Greek was the language of the Linear B tablets was the culmination of twelve years of interest in the mysterious script, beginning with a paper 'Introducing the Minoan language' published in the *American journal of archaeology* in 1940, when he was 18. Although it contradicted current opinions (classicists used to think that 'the Greeks' could not have invaded Crete until after the time of the tablets) Ventris's insight was soon taken up by others, notably John Chadwick of Cambridge (see his *The decipherment of Linear B*, 2nd edn, 1967) and Emmett L. Bennett of Yale. Linear B is a syllabary, combining consonant with following vowel in each character.

The Greek alphabet developed out of a Phoenician script that had symbols for consonants only. In the Greek alphabet, vowels were written as well – a spectacular advance. At first, local alphabets varied widely: almost every city seemed to have its own. But by about 400 BC most of them were already using something very like the present upper-case Greek alphabet of 24 characters.

The Greek alphabet

Α Β Γ Δ Ε Ζ Η Θ Ι Κ Λ Μ Ν Ξ Ο Π Ρ Σ Τ Υ Φ Χ Ψ Ω

α β γ δ ε ζ η θ ι κ λ μ ν ξ ο π ρ σ τ υ φ χ ψ ω

a b g d e z ē th i k l m n x o p r s t y ph kh ps ō

Greek in Writing: 2. From alpha to omega

Some argue that it was first used for writing down poetry (notably the epics, *Iliad* and *Odyssey*), while others think that laws and official inscriptions came first. But the earliest surviving examples of the Greek alphabet are actually 8th-century BC graffiti on a rock face on the island of Thera, mostly recording homosexual encounters.

GUARANÍ

3,000,000 SPEAKERS

Paraguay

Tupí and Guaraní form a closely related group, possibly belonging to the family of AMERIND LANGUAGES. When the Spanish and Portuguese began to explore the south-eastern parts of South America, they found that the languages of Para-

guay and coastal Brazil were no more different from one another than were Portuguese and Spanish. Guaraní, the language of Paraguay, was at the eastern end of a dialect chain that stretched unbroken from what is now southern

South America:
Tupí and Guaraní
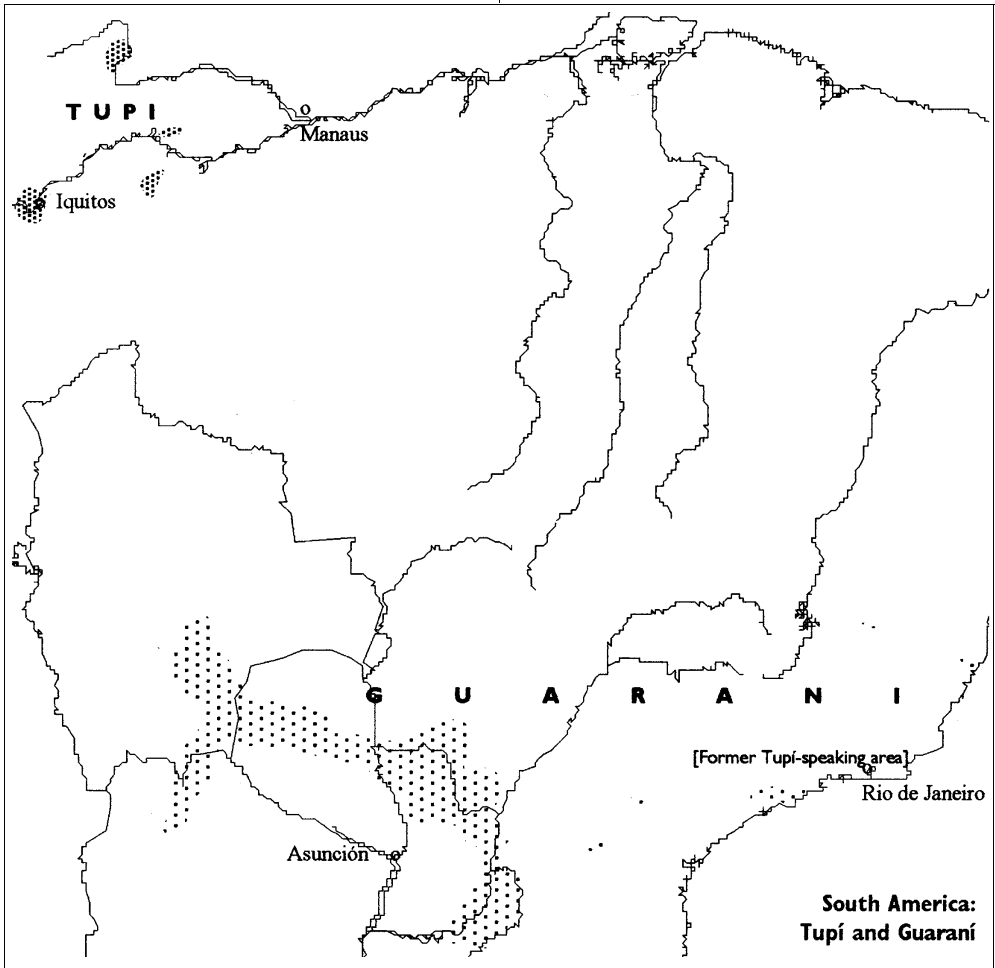
Bolivia to the southern Brazilian coast: the Tupí dialects extended from here north-eastwards.

The later history of these two important languages has been very different. As early as 1583 the Spanish-speaking Jesuits of Paraguay began a programme of settlement and education for the speakers of Guaraní, their converts and subjects. The long term result is that although Guaraní has practically disappeared from the rest of its old territory, Spanish and Guaraní are now the two national languages of Paraguay. Religious texts have been published in Guaraní since the 18th century.

Four centuries of coexistence, during all of which time Spanish has had a higher status, have led to considerable changes in what was at the outset a major language of inland South America, a language relatively uninfluenced by others. Guaraní is now stuffed with Spanish loanwords: in some subject areas the whole vocabulary is Spanish. Even the numeral system is largely Spanish. The common numerals in Guaraní are: *peteĩ, mokõi, mbohapy, irundy, sinko,* *sei, siete, ocho, nueve, diez*. '5' to '10', and all the higher numerals, are Spanish loanwords, though the old Guaraní word for '5', *pekua*, can still be heard.

As the major language of its region Guaraní has contributed loanwords to several European languages. Those to be found in English include *cougar, jaguar, toucan*. The Guaraní word *petun*, once a synonym for 'tobacco' in English, survives in the name of the botanical genus *Petunia*. *Para*, a common element in South American place names, is the Guaraní word for 'river'.

The Tupian languages on the map

Guaraní is one of the national languages of Paraguay. There are small communities of speakers in Argentina and Brazil.

Tupi, once a lingua franca of Brazil, now has only a few thousand speakers on the banks of the Vaupés, Içana and Rio Negro in Amazonas state.

Tupian languages once extended from the Amazon southwards to the border of Uruguay.

GUJARATI

45,000,000 SPEAKERS

India, Great Britain

Gujarati is the Indo-Aryan language of the Indian state of Gujarat – and of numerous migrant communities in many parts of the world.

The name is undoubtedly that of the *Gujjars*, a widespread caste, traditionally cattle-farmers. Nomadic members of the community are nowadays most commonly found far to the north, in Kashmir and Himachal Pradesh, where they speak *Gujuri* (see INDO-ARYAN LANGUAGES).

Old Gujarati, ancestor of both Gujarati and Rajasthani, is recorded in texts of the 12th to 14th centuries. Gujarati literature really begins in the 15th century, however, with mystical poets such as Mirabai. Both Persian-Urdu and Sanskrit traditions of poetry continue to influence Gujarati authors. The statesman and writer Mohandas Karamchand Gandhi (1869–1948), born in the Gujarat town of Porbandar, helped to inspire a 20th-century renewal in the literature of his mother tongue.

Although the major regional dialects differ little, there are also social dialects of several kinds. There are strong differences in intonation between men's Gujarati and women's Gujarati. Among tribal dialects are the mountain languages Bhili and Khandesi (see map). An important caste dialect is that of the Parsis, clustering in the southern part of the state. They have a press and literature of their own, quite distinct from standard Gujarati in style and idiom.

Muslim speakers of Gujarati – now mostly settled in Pakistan, Bangladesh and Britain – tend to be at least trilingual, being familiar also with Urdu and having studied classical Arabic. Their Gujarati, like that of the Parsis, contains many Persian loanwords. Another dialect of historical importance is that of Kathiawari sailors, who travelled all over the world as steamship crewmen. Their speech included loanwords from Urdu and several European languages.

In Kutch, in north-western Gujarat state, the 500,000 speakers of the Sindhi dialect *Kacchī* use Gujarati as their literary language.

Outside the borders of Gujarat state, Gujarati speakers are widely scattered: in Maharashtra state; in Pakistan and Bangladesh; in South Africa; in Singapore. One of the oldest communities of the diaspora is in East Africa, where Indian traders (Gujarati, Kacchī and Konkani) were already established in the 1490s when Portuguese explorers reached the Indian Ocean. This linguistic enclave was reinforced when Gujaratis were recruited in large numbers by both British and German governments to work in East Africa around 1900. East African Gujarati became a distinct dialect. Many East African Gujaratis, faced with an increasingly uncertain future and mounting discrimination in newly independent East African countries – especially Uganda – migrated to Britain, where there are now a third of a million speakers of the language, many of them in the London areas of Wembley, Harrow and Newham and in Leicester, Coventry and Bradford.

The Gujarati alphabet

અ આ ઇ ઈ ઉ ઊ ઋ એ ઐ ઓ ઔ કખગઘઙ ચછજઝઞ ટઠડઢણ તથદધન પફબભમ યરલવ શષસહ

a ā i ī u ū ṛ e ai o au k kh g gh ṅ c ch j jh ñ ṭ ṭh ḍ ḍh ṇ t th d dh n p ph b bh m y r l v ś ṣ s h

Numerals in Gujarati			
એક	૧	1	ek
બે	૨	2	be
ત્રણ	૩	3	traṇ
ચાર	૪	4	cār
પાંચ	૫	5	pāñc
છ	૬	6	cha
સાત	૭	7	sāt
આઠ	૮	8	āṭh
નવ	૯	9	nav
દસ	૧૦	10	das

The Gujarati Script

The Gujarati script, which is similar to Devanagari but without the headstrokes or 'washing line', was standardised in its present form in the 19th century. The usual transliteration is given in the box. Both script and transliteration are poor at representing the vowel sounds of Gujarati, which actually has 8 normal vowels əaɛeiɔou, no distinction of vowel length, but (especially in the standard and central dialects) a range of breathy vowels.

Gujarati and Rajasthani on the map

Standard literary Gujarati is based on the central dialect of Baroda and Ahmadabad. Other dialects of Gujarati are Pattani (north), Surati or Surti (south) and Kāṭhiyāwāḍī of Saurashtra. They differ relatively little from the standard.

Bhīlī is spoken by the mountain tribes of Bhil, to the east and north-east of Gujarat state. It has numerous dialects: Wāgḍī alone claims 750,000 speakers. The whole dialect group may have as many as 2,500,000. Bhīlī is often counted a separate language.

To the south of Bhili, *Khandeśī* is another group of hill dialects, with up to a million speakers. Khandeśī is heavily influenced by Marathi.

RAJASTHANI, by contrast with Gujarati, has no strong modern literary tradition. Its main dialects are Marwari, Jaipuri, Mewati and Malvi. Nimadi, an isolated southern dialect in a mainly 'tribal' area, has developed special peculiarities.

Gujuri is the language of nearly half a million semi-nomadic herdsmen in Indian and Pakistani Kashmir and in Himachal Pradesh. Gujjars keep buffaloes and cows, while Bakarwals, who speak the same language, keep goats and sheep. Most speakers own their winter pastures, moving to high meadows, *tok*, during the summer season, April to August. The language is close to Rajasthani.

Parya is a recently discovered Indo-Aryan language spoken by about 1,000 people in the Hissar valley, Tajikistan. Its history is unknown – but linguistically it seems to belong with Rajasthani and Panjabi.

Rule of thumb

Aro gharo ne, nhālli va'w par āvine paro.
Whatever happens, the youngest
daughter-in-law gets the blame.

Shrikrishna N. Gajendragadkar,
Parsi-Gujarati (Bombay:
University of Bombay, 1974) p. 153
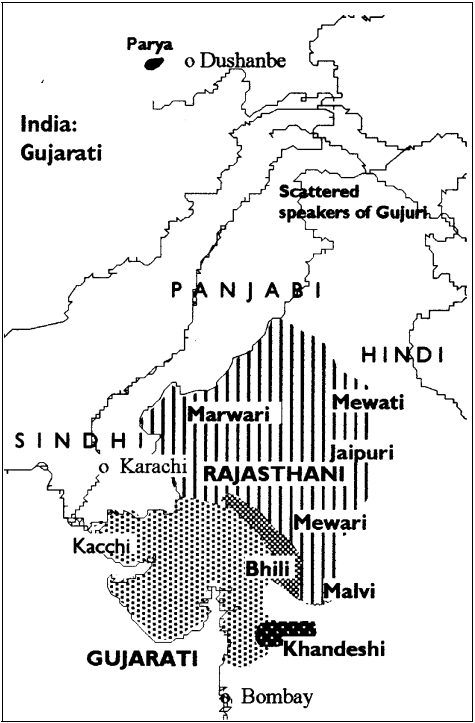
GUSII

2,000,000 SPEAKERS

Kenya

Gusii or Kisii is one of the BANTU LANGUAGES, spoken in a heavily populated region of Kenya, the Kisii highlands – two thousand metres high – of Nyanza province. Its speakers trace their history to a migration five hundred years ago. They are predominantly arable farmers and cattle-breeders.

Gusii is remarkable for the low number of words that it shares with other members of the Bantu family, even with neighbours to which it is considered closely related, such as LUYIA (see map there). In fact Gusii has been described as having 'one of the most interfered-with vocabularies in East Africa' (Derek Nurse). It seems to have borrowed from Cushitic languages at some earlier period, and more recently from Nilo-Saharan languages – indeed its territory is entirely surrounded by those of LUO, Kipsigis (see KALENJIN) and MASAI.

HAITIEN

6,000,000 SPEAKERS

Haiti

One of the FRENCH CREOLES, and very close historically and linguistically to the forms of Creole spoken in Guadeloupe, Martinique and other Caribbean islands, Haitien has developed a separate identity thanks to the long independent history of Haiti. It is the one essential everyday language, used throughout the island, but until very recently it had no official recognition, French being regarded as the sole language of government and culture in Haiti. The Haitien spoken in the capital, Port-au-Prince, still shows relatively greater influence from French because of the concentration there of educated speakers and of government offices.

In Haiti, the educated language of the urban elite is French. To reach a wide audience, you write Creole. But a writer's Creole will be influenced, more or less intrusively, by French. For example, in US Embassy publicity, the phrase 'to the United States' may appear in a form closer to standard French, *Ozetazini* (French *aux Etats-Unis*), rather than in the everyday Haitien *peyi Zetazini* (from French *pays* 'country' + *les Etats-Unis* 'the United States').

In the same way, spoken Creole may include more or fewer French features – such as the front-rounded vowel in elevated Creole *zeu* (ordinary Creole *ze*), from French *les oeufs*, 'eggs'.

The Haitian audience

Literacy remains low. Yet Haitian Creole authors still reach a mass audience, as Franketienne, author of *Dezafi* (Port-au-Prince, 1975), explained in an interview:

'Bon, dapre ou menm, konben moun kapab li sa – bagay tankou *Dezafi*, sa ki fèt an kreyòl?'
'Sa k'fèt an kreyòl? M fè eksperyans deja. E gen anpil zanmi ki fè eksperyans nan sèten pwovens. Pèp-la pa konn li. Yo li l pou yo. Yo li *Dezafi* nan sètenn rejyon pou de san, sen senkant moun ki reyuni: yo konprann tou sa k'ladan yo.'

'Well, according to yourself, how many people are capable of reading it – something like *Dezafi*, that's written in Creole?'
'Written in Creole? I have already proved that. I have plenty of friends who have proved that, in various parts of the country. People there can't read. They read it for them. They read *Dezafi* in some places to two hundred, a hundred and fifty, people at once. They understand everything in it.'

From *Les créoles français entre l'oral et l'écrit* ed. Ralph Ludwig (Tübingen: Narr, 1989) pp. 53–4

If you know French, you will find that this text begins to seem familiar if you read it aloud, making sure that the stress falls on the last syllable of each word as it does in French. A vowel followed by *-n* is nasalised. Notice the following words: *zanmi* (from French *les amis*) 'friends'; *moun* (from French *du monde*) 'people'; *tankou* (from French *tant que*) 'like'; *de san* (from French *deux cents*) 'two hundred'.

HANI

PERHAPS 1,300,000 SPEAKERS

China and south-east Asia

Hani is spoken by mountain-dwellers who live in a compact region of western Yunnan, China, and in scattered villages in four other countries – Vietnam, Laos, Burma and Thailand. It is the main language in the 'southern Loloish' subgroup of SINO-TIBETAN LANGUAGES (see map at BURMESE). Hani speakers are an official nationality of China, with a population there of over 1,000,000. It is only recently that the size of this language community has been realised: earlier estimates were closer to 200,000.

> *Hani* is the official Chinese name for the language and people best known internationally as *Akha*. In Burmese and Shan (and in older English writings) Hani speakers are called *Kaw*, in Lao (and French) *Kha Kaw* or *Kha Ko*, in Thai *Ikaw*, in Vietnamese *Hanhi*.

Hani speakers in southern Yunnan have traditionally been rice farmers, building narrow, irrigated terraces on steep mountainsides; they are also among the growers of Pu'er tea, so named because it is traditionally marketed through the city of Pu'er.

Their migration southwards from Yunnan appears to be quite a recent phenomenon. In the south-east Asian countries Hani speakers are often bilingual in the local majority language such as Shan or Thai. Hani men are often familiar with LAHU, using Lahu loanwords freely in their own language. Roman Catholic missionaries have had some success among Hani speakers and have published religious texts in the language.

For a table of numerals see LAHU.

Hani genealogies and personal names

'Akha culture differs in many ways from that of other Loloish groups such as Lahu. The Akha are preoccupied with religious ritual. The religious leader or *dzŏema* is usually a person of considerable importance, as his title, which includes the verb ''rule'', implies.

'The village is sited above, not near or below, its water source. At the main entry there is a gate, renewed each year with considerable ritual. Near the centre of the village there is a ''swing'', which is used only at the New Year, when the gate is renewed. [Villagers take turns swinging, in a ceremony lasting three days, while sacrifices are performed to expel evil spirits.]

'The Akha clans are named according to their common ancestor, and members of a clan are thought to constitute a more or less homogeneous group, speaking a distinct dialect. Every individual has his own genealogy, which includes a series of mythical ancestors preceding the clan founder, then the ''common ancestor'', and then a series of increasingly more recent ancestors, who are more and more likely to be real, named individuals, until the father is reached. As in the case of genealogies among other Loloish groups, the final syllable of one name is the initial syllable of the following name – unless an individual has been renamed. Thus the son's name normally begins with the last syllable of the father's name.'

David Bradley, *Proto-Loloish* (London: Curzon Press, 1979) pp. 33–4, abridged

Hausa

25,000,000 SPEAKERS OF HAUSA AS A MOTHER TONGUE

Nigeria, Niger, Cameroun

Hausa is the most widely known of the CHADIC LANGUAGES, which belong to the Afroasiatic family. Once the language of the Empire of Sokoto, it is now spoken in the Northern Region of Nigeria and across most of neighbouring Niger.

In Kanuri the Hausa language is *Afūno* – hence its name in some medieval Arabic and 18th-century European sources, *Afnu*. The term *Hausa* is first recorded, as a tribal name, in a 13th-century Arabic text. Speakers call themselves *Háusáawáa* (singular *Bàháushè*).

In medieval times Hausa was the language of seven states of the 'central Sudan' – in terms of modern political geography, northern Nigeria. They and their neighbour Bornu, where KANURI was spoken, were scarcely known to the outside world except for the description by the young Arabic-speaking traveller 'Leo Africanus', written in the early 16th century when he was a captive in Italy. Some Europeans probably lived as slaves in the Hausa states in recent centuries, but none returned.

However, there have always been trade routes across the Sahara, and in North Africa the devoutly Muslim Hausa are still renowned as traders. Their home territory has been a crossroads. Trade goes from the Niger valley eastwards and (even more important) from the West and Central African forests northwards, along the Sahara caravan routes to the Mediterranean. Hausa themselves, in considerable numbers, have settled in West and North African cities at the nodes of these trade routes.

Thus the extent and diversity of the Hausa country is as striking as the diaspora of Hausa communities elsewhere. The language is an indispensable medium of communication for all inhabitants of northern and central Nigeria and Niger and for many in Benin, Togo, Ghana, Chad and Cameroun.

Groups of the essentially nomadic FULANI pastoralists had long settled in the Hausa country. In 1804 a Fulani priest, Osman dan Fodio, founded the Empire of Sokoto. Through the 19th century this ruled what is now northern Nigeria. Although Fulani became the dominant language of the state of Adamawa, Hausa, language of the great majority elsewhere, maintained its position as the ruling language of the new empire – and flourished as a language of written poetry, in the Arabic script known as *ajami*.

Under British rule Hausa retained official status – jointly with English – in northern Nigeria, and with their departure it became one of the official languages of Nigeria. In the early years of independence there was even a movement to make Hausa the single national language, though against the opposition of most Ibo and Yoruba speakers (not to mention the many smaller linguistic groups) this was doomed to fail. It is, however, used as a second or third language by a large proportion of the population of northern Nigeria. As many as 40,000,000 people probably use Hausa on a daily basis.

Hausa has a long tradition of literacy and of written (mostly religious) literature in ajami script. Under British rule a Roman alphabet for Hausa was developed and assiduously encouraged: this *boko* script (from English *book*) was made official in 1930. The Hausa Language Board, established in 1955 on the model of other British colonial committees, worked towards the standardisation of
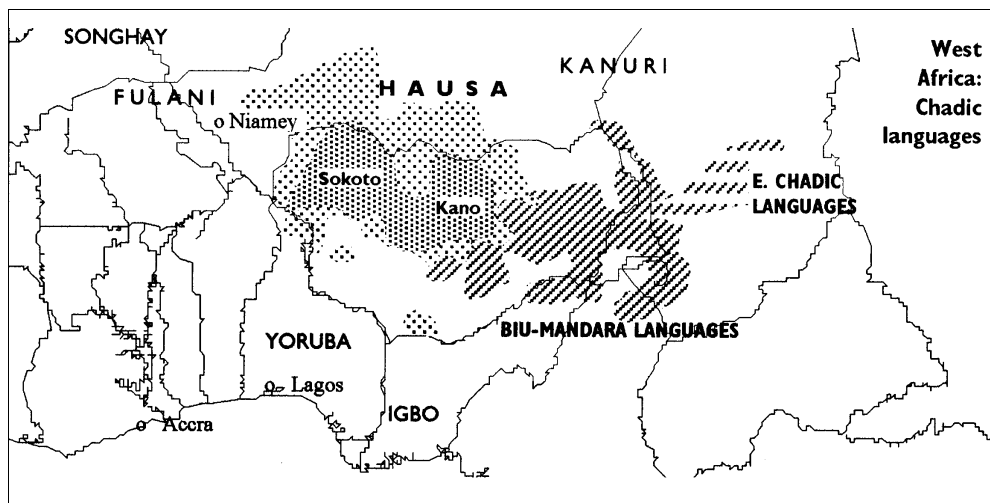
the written language on the basis of the Kano dialect. Literary Hausa is enriched with idioms and phrases drawn from other dialects. There is now a strong press in Hausa and the language is well represented in the media.

Barikanci, a pidgin form of Hausa, grew up around army barracks in northern Nigeria in the early twentieth century and came to be used as a lingua franca by the armed forces. Simplified Hausa is also used on the desert route to Mecca.

Hausa has three tones, high, low and falling. Like other Chadic languages it has a characteristic set of verbal extensions, the effect of which has been compared to the phrasal verbs of English such as *give away, finish up*. There are six verbal extensions in Hausa, including *-a* 'down, properly', *-e, -nye* 'completely, all over', *-o, -wo* 'towards', *-n* 'to, for'. The first ten numerals are: *ɓaya, biyu, uku, huɓu, bìyar, shidà, bakwày, takwàs, tarà, gomà*. Some higher numerals are Arabic loanwords.

Pure Hausa – impure Hausa

As far back as tradition serves, the language was spoken in the 'Hausa seven', *Hausa bakwai*: these seven states are Biram, Daura, Rano, Kano, Zazzau (Zaria), Gobir and Katsena. Here dialect differences are few: the main dialects are those of *Sokoto* and *Kano*.

Nowadays Hausa is spoken, or at least serves as a lingua franca, far beyond this region. Tra-

ditionally Hausa speakers give a list of the *banza bakwai* or 'false seven' states that have adopted something of Hausa language and culture: Zamfara, NUPE, Kebbi, Gwari, Yauri, YORUBA and Kororofa (the *Jukun*-speaking region).

There are long-established Hausa communities in many Ghanaian towns, speaking their own dialect. The Hausa used as a trade language in Ghana is a pidginised form, largely lacking the masculine/feminine distinction (as does the usual spoken English of the area). It is one of the three most important lingua francas of Ghana, along with AKAN and English.

The subgroups of Chadic

The Biu-Mandara languages include *Bura* (250,000 speakers in Bornu State, Nigeria), *Kilba* (100,000 in Gongola State), *Mafa* or Matakam (125,000 in Far North State, Cameroun), *Margi* (200,000 in Bornu and Gongola States, Nigeria), and many smaller languages.

The numerous East Chadic languages are spoken in southern Chad, Cameroun and the Central African Republic. None has more than 50,000 speakers.

West Chadic languages, all spoken in Nigeria, include *Angas* (100,000 speakers), *Bade* (100,000), *Tangale* (100,000) and *Hausa*. Hausa has 20,000,000 speakers in Nigeria, 2,000,000 in Niger and probably 3,000,000 in other countries.

HAWAIIAN

2,000 SPEAKERS

Hawaii

One of the AUSTRONESIAN LANGUAGES, the most northerly by far of the Polynesian group, Hawaiian was still the language of an important independent kingdom in the early 19th century. Newspapers and a range of publications appeared in it. A Bible translation was published in 1839. It was a language of oral literature, too, notably the creation myth *Kumulipo*.

Now it has been swamped by English, and is spoken only by the two hundred people on Ni'ihau island and by not more than two thousand people in total. Even these are bilingual in English, or in the Hawaiian Creole variety of English.

Hawaiian is studied – along with other languages of Oceania – at the University of Hawai'i. The language is still used in some state ceremonies and in church services, and Hawaiian phrases still impress tourists.

Hawaiian, like other languages of the Polynesian group, has a simple sound pattern, with five vowels and only eight consonants, *' h k l m n p w*. The first ten numerals in Hawaiian are: *kahi, lua, kolu, haa, lima, ono, hiku, walu, iwa, 'umi*.

HEBREW

3,000,000 SPEAKERS

Israel

One of the SEMITIC LANGUAGES, Hebrew has an astonishing history. It was the language of the early Jews, the language in which the books of the Old Testament were written. Two thousand five hundred years ago, ultimately as a result of the Jewish 'captivity in Babylon', Hebrew ceased to be an everyday spoken language – replaced, naturally enough among a displaced and scattered people, by Aramaic, the lingua franca of the Babylonian and Persian empires.

But Hebrew remained the sacred language of the Jewish religion: and in the early 20th century, as Jews from all over the world began to settle once more in Palestine, Hebrew was revived, from the old texts, as a spoken language. It became an official language in 1922 and is now the everyday speech of Israel and the mother tongue of a growing number of Israelis.

The English name *Hebrew* comes, by way of Latin and Greek, from Aramaic *'Ebrāyā*. The original is ancient Hebrew *'ibrī* 'from the other side of the river'. Some linguists prefer to distinguish modern Hebrew under the name *Ivrit* (which is the modern pronunciation of its name in Hebrew).

Modern Hebrew necessarily differs from the ancient language. It differs in pronunciation, since it had to compromise with the speech habits of its first new users, mainly Yiddish-speaking, and of more recent immigrants to Israel; in any case the details of ancient pronunciation are not fully known, and it was necessary to build on traditional Hebrew pronunciations used in the religious schools. It differs from ancient Hebrew in vocabulary, too, because

modern life demanded numerous new words including many English loans.

The writings included in the Jewish Bible (the Christian 'Old Testament') were first composed between 1200 and 200 BC, and there are also a few Hebrew inscriptions from this period.

The Hebrew literature of the next period includes the Mishna, Tosephta, Midrash, and various texts called 'apocryphal' (which is essentially a negative term meaning that, though in some ways comparable to Biblical texts, they were not included in the Bible). Although Hebrew had ceased to be used as the mother tongue, it was still the language of scripture and law, and for many centuries also of poetry and philosophy. There was a notable flowering of Hebrew poetry and scholarship in Muslim-dominated Spain, in the 10th and 11th centuries.

By now Jews were writing in other languages too: not only Aramaic and Greek but also JUDEZMO (the Romance language of the *Sephardim* exiled in 1492 from Spain to Turkey) and YIDDISH (the Germanic language of the *Ashkenazim* who spread eastwards from Germany into central and eastern Europe). There are several more Jewish languages, including varieties of ARABIC, PERSIAN and OCCITAN. As migration to Israel has increased, so modern Hebrew is supplanting these other languages, none of which has official status in Israel. There is a growing modern literature in Hebrew.

Like Arabic, Hebrew is a language whose vocabulary is largely built on three-consonant roots. Because of its status as the language of the Old Testament, details of Hebrew and its structure permeate into modern Western lan-

guages. Many are aware of the typical Hebrew noun plural suffix -im (cherub, cherubim).

The first ten numerals in Hebrew are: *akhat, shtayim, shalosh, arba, khamesh, shesh, sheva, shmone, tesha, esser.*

The script

The earliest Hebrew alphabet was derived directly from a Phoenician script. The familiar square Hebrew alphabet of classical and modern times developed out of an Aramaic script, itself originally based on Phoenician.

Vowels can be fully marked in Hebrew writing by the use of additional signs, but this is normally done only in texts of the Bible, where correct pronunciation is essential for religious reasons. Even when using these vowel signs, no distinction is made between the central vowel ə and the absence of a vowel.

The Hebrew alphabet

ת ש ר ק צ פ ע ס נ מ ל כ י ט ח ז ו ה ד ג ב א

ʼ b g d h w z ḥ ṭ y k l m n s ġ p ṣ q r š t

HILIGAYNON

4,600,000 SPEAKERS

Philippines

One of the Bisayan languages, like CEBUANO (see map there), and thus a member of the larger family of AUSTRONESIAN LANGUAGES, Hiligaynon is spoken on the islands of Panay and Siparay (Negros) in the southern Philippines.

There are about 4,600,000 speakers of Hiligaynon. The language is close to *Capiznon* (450,000) of the north-eastern shore of Panay, and to *Masbateño* (350,000) of Masbate, which faces Panay and Siparay across the Visayan Sea. Masbateño in turn is linked with *Northern* or *Masbate Sorsogon* (100,000) of the city of Sorsogon in south-eastern Luzon.

Hiligaynon has become the dominant language in the whole of Panay Island. Speakers of Kinaray-a, to the south-west, and Capiznon, to the north-east, tend to be competent also in Hiligaynon and the national language, Tagalog.

Hiligaynon soldiers were recruited by the Spanish in their conquest of the southern Philippines in the 17th century. This explains the Hiligaynon influence that is evident in the SPANISH creole now spoken in Zamboanga at the western end of Mindanao.

The first ten numerals in Hiligaynon are: *'isah, duhah, tatloh, 'apat, limah, 'anom, pitoh, waloh, siyam, napulo'*.

HINDI

180,000,000 SPEAKERS

India

One of the INDO-ARYAN LANGUAGES, Hindi is cultivated in two forms which may be called two languages, though there are no significant differences of grammar or pronunciation between them. The twin language of Hindi is URDU (see table of numerals there). Hindi contests with English the position of India's principal language.

> *Hindi* is so called as the language 'of India', from the Mughal point of view (Persian *Hind* 'India') – because it was the vernacular of the Delhi region, from which the Mughal Sultans ruled.

The history of Hindi, with the closely related dialects Braj and Awadhi, can be traced to the 12th century: earlier writings from the Hindi-speaking region are in middle Indo-Aryan dialects (Prakrits and Apabhramsas). Until the 19th century Hindi was important only as a spoken language and in literature – nearly all of it in verse, for Hindi has 'a system of versification which for inexhaustible variety, as well as for its intrinsic beauty, is probably unsurpassed' (S.H. Kellogg, *A Grammar of the Hindi Language* (London: Routledge, 1938). In the government of northern India Persian ruled. Under the British Raj Persian eventually declined, but, the administration remaining largely Muslim, the role of Persian was taken not by Hindi but by Urdu, known to the British as Hindustani. It was only as the Hindu majority in India began to assert itself that Hindi came into its own.

As spoken languages, Hindi and its close relatives meanwhile continued to grow as the north Indian population grew. Under British rule millions of speakers found work abroad, and many of them established permanent colonies which still prosper. However, linguists regard the Indo-Aryan languages of these overseas settlements as closer to Bhojpuri than to any other Hindi-like language (see BHOJPURI).

Standard Hindi and Urdu are structurally alike. Both are literary languages based on the spoken dialect of Delhi, known as *Dihlavi* or *Khari Boli* 'standard speech'. But in terms of vocabulary there are strong differences. Urdu draws official, scholarly and religious vocabulary from Persian and, through Persian, from Arabic. Hindi draws on the resources of Sanskrit for learned and technical vocabulary of all kinds. Thus the difference between them is essentially a religious and cultural one, underlined by the fact that they use different scripts.

While the independence movement attempted to unite Hindus and Muslims politically, Mahatma Gandhi urged speakers of Hindi and Urdu to merge their languages in a shared 'Hindustani' – a name which continued to be emphasised in linguistic writings of the period. However, when Pakistan and India were separated on religious lines, Urdu was left as a relatively minor language within India, and Hindi became paramount.

Hindi (*Modern Standard Hindi*, to be precise – *High Hindi*, as it used to be called) is the standard and official language of a vast inland region of India: Delhi and the states of Uttar Pradesh, Bihar, Madhya Pradesh, Rajasthan, Haryana and Himachal Pradesh. Throughout these states it is the language of administration, the press, schools and literature. More than that, Hindi has official status, alongside English, throughout India – unlike all the other official languages of

India, which are recognised only in their own states or districts. This special position of Hindi is a compromise. Proponents of Hindi wanted it to take the place of English as India's lingua franca. It may still do so: but as yet speakers of the other state languages of India generally prefer English, not Hindi, to be their second language.

Hindi literature is taken, these days, to include the classics written in the older standard languages Braj and Awadhi. Old Hindi literature, antedating these, is relatively small in bulk (it includes two short pieces in the *Adi Granth* of the Sikhs, attributed to the Sufi saint Baba Farid, who died in 1266). Braj was the language for praises of Krishna; traditionally associated with the cities of Mathura and Vrindaban, it was used for poetry from the 16th to the early 20th century. Awadhi belongs to literature in praise of Rama, legendary king of Ayodhya, and was used for poetry between the 16th and 18th centuries. The great *Ramcaritmanas* of Tulsidas is in Old Baishwari or Old Awadhi of the late 16th century. The oral epic *Alhakhand* is traditionally sung in Banaphari, a mixed Bundeli-Bagheli dialect of Hamirpur where the action of the epic takes place.

The *Sadhu Bhasha*, the fluid mixed literary language of the Nirguna tradition of mystical poets, beginning with Kabir in the late 15th century, incorporated some elements of *Khari Boli*, the local language on which Hindi is directly based. But the creation of modern 'High Hindi' as a literary language in the early 1800s is credited to Lallu Lal, author of *Prem Sagar* and *Rajniti*, retellings of Sanskrit classics which immediately achieved classic status themselves. Hindi is now the language of a prolific literature in prose as well as verse.

Hindi -*vālā*, English *wallah*

Hindi has the unusual distinction of having contributed a (now rather old-fashioned) loan suffix to English: this is the form -*vālā* added to a noun, English *wallah*. Hindi examples: *dudhvālā* 'milkman', *Dihlīvālā* 'man from Delhi', *Kābulvālā ghoṛā* 'horse from Kabul'. One of the oldest Anglo-Indian examples is *competition-wallah* 'one who got his job by competitive examination'.

Hindi is an easy language for the average West European learner. It is clearly Indo-European in much of its structure, but, as with English, un-

India: Hindi area languages
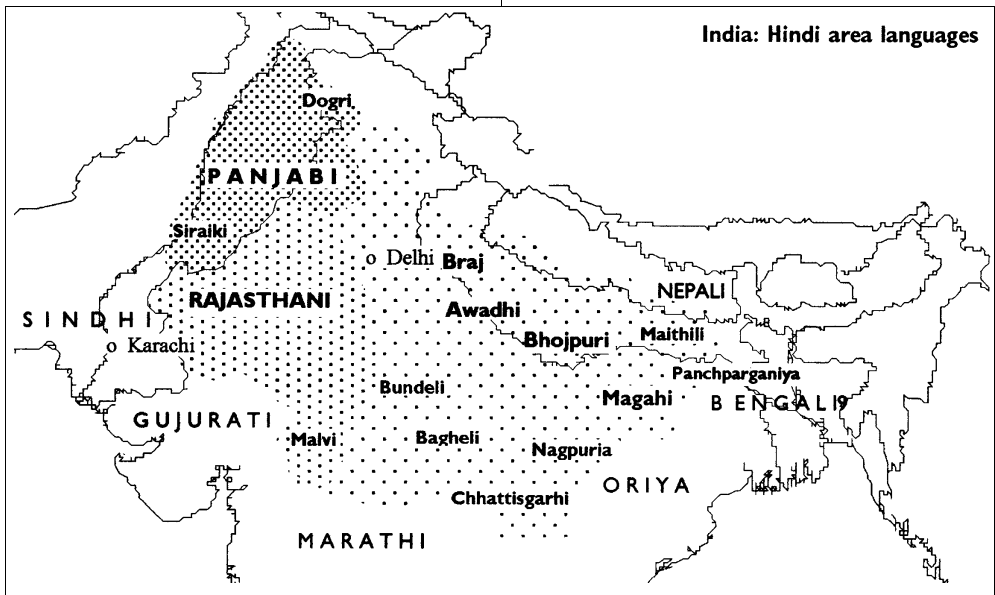
varying prepositions and particles tend to replace the complex noun and verb inflections of older forms of the language. It has two genders (Sanskrit and Pali still retained three from Indo-European).

The Hindi language area

The linguistic map of the north Indian plains cannot be drawn in sharp lines. In the neighbourhood of Delhi, the local speech varieties (*Khari Boli*) would be regarded by everybody as dialects of Hindi. Almost as close to Standard Hindi is Eastern PANJABI, yet that is treated – here and elsewhere – as a separate language. As one moves outwards from Delhi, one comes to other dialects which have had claims to the status of 'language': *Braj* (centred on Agra, capital city from 1566 to 1658) and *Awadhi* (of the old kingdom of Oudh). Then come some language varieties that are rather more distinct from Hindi: BHOJPURI, MAGAHI and MAITHILI, the three once grouped together as *Bihari*, but that was a name dreamed up by a linguist which has no cultural validity; CHHATTISGARHI; RAJASTHANI.

Bhojpuri is spoken in the middle Ganges valley, both north and south of the historic city of Benares. Sadani or *Nagpuria* may be regarded as a dialect of Bhojpuri: it is the vernacular and lingua franca of the eastern Chota Nagpur plateau.

Magahi is spoken in the southern half of Bihar, notably the cities of Patna and Gaya. *Panchparganiya* is the eastern dialect of Magahi.

Maithili is the language of northern Bihar and southern Nepal.

Chhattisgarhi is spoken in south-eastern Madhya Pradesh in the districts of Raipur, Bilaspur, Raigarh, eastern Balaghat and northern Bastar.

Rajasthani is itself a grouping of quite varied dialects, a reflection of the long political fragmentation of what is now the state of Rajasthan.

The script

The most striking difference between Hindi and Urdu is in their scripts. Urdu uses a form of Arabic script. Hindi is written in the Devanagari alphabet which is also used for Nepali and Marathi and is the usual choice for printing Sanskrit.

The Devanagari alphabet for Hindi

अ आ इ ई उ ऊ ऋ ऌ ए ऐ ओ औ कखगघङ चछजझञ टठडढण तथदधन पफबभम यरलव शषसह

a ā i ī u ū ṛ ḷ e ai o au k kh g gh ṅ c ch j jh ñ ṭ ṭh ḍ ḍh ṇ t th d dh n p ph b bh m y r l v ś ṣ s h

---

Like all the major languages of south Asia, Hindi can be written in any south Asian script. The Indian prime minister H. D. Deve Gowda, a native Kannada speaker, gave the traditional Independence Day speech in 1996 in Hindi – but he is said to have read from a text converted for him into Kannada script.

Hiri Motu

PERHAPS 250,000 SPEAKERS OF HIRI MOTU AS A LINGUA FRANCA

Papua New Guinea

Motu, which belongs to the Oceanic group of AUSTRONESIAN LANGUAGES, is one of the two local languages of the Port Moresby region in Papua New Guinea. In this linguistically fragmented region the speakers of Motu appear to have developed in contact with their immediate neighbours a pidginised or simplified form of their language which was used in local trade.

'This Foreigner Talk or Simplified Motu is presumed to have been taken up, used and spread by relatively large numbers of foreigners who settled or traded in the Port Moresby area in the late 1870s and throughout the 1880s, some . . . later employed by the Government as interpreters, guides, boatmen and unofficial policemen' (Dutton).

Police Motu was traditionally looked down on by external observers, who viewed it as a 'corrupt' Motu. But Seventh Day Adventist missionaries found it indispensable in missionary work. The first study of Police Motu, published as late as 1962, found that the so-called Central dialect, in general used by speakers of Austronesian languages related to Motu, differed from the non-Central dialect used by speakers of Papuan languages both in phonetics and grammar – for example, in having a range of possessive suffixes, *-gu* 'my',*-na* 'your', *-na* 'his, her, its'.

The first three numerals are: *ta, rua, toi*. Numerals above '3' are English loanwords.

When Papua became a British colony in 1888, this language soon became indispensable to the local administration and particularly to the police force that was established in 1890. *Police Motu* – as it was still called until recently – gradually spread all over Papua with government control and peacekeeping.

Until recently, Police Motu had no official recognition or encouragement. It is now accepted alongside English and TOK PISIN, the local English-based creole, as a national language of Papua New Guinea and an important medium of communication across language boundaries. However, it is in practice less used officially than the other two, and is seldom heard in Parliament. The new official name, *Hiri Motu*, comes from the probably mistaken belief that Motu speakers once used this language on their long distance trading voyages, *hiri*.

Based on Thomas E. Dutton, 'Police Motu of the Second World War' in *Pacific linguistics*, A 76 (1988) pp. 133–79 and other sources

HITTITE

Extinct language of Turkey

'Hittites' were known only from references in the Bible and other ancient Near Eastern sources until the beginning of the 20th century. Then clay tablets written in an unknown non-Egyptian language were found at Amarna in Egypt. Soon afterwards, great numbers of tablets in the same language were found in the excavation of what was evidently a royal capital, Boğazköy in north central Turkey. The history of this kingdom can now be traced in documentary and literary texts in its local language from about 1780 to 1200 BC.

An unrelated people of central Turkey are called *Hatti* in the texts: the capital city now known as Boğazköy had the related name *Hattušas*. But the name that speakers of what we call 'Hittite' gave to their language is not known.

The tablets could not at first be read. The Czech scholar Bedřich Hrozný showed in 1915 that they were a very ancient record of an Indo-European language, from a branch hitherto completely unknown.

Hittite texts include laws, myths (some of them Hattic and Hurrian in origin), historical texts and poetry. The Hittite law texts are organised in two series, the first known as *If a man . . .*, the second *If a vineyard . . .*, from the first words of each. Hittite religious poetry includes passages in the older Hattic language.

The Hittite kingdom in its heyday was in close touch with other Near Eastern peoples. The Assyrian trading colony of Kanesh was not far to the south. The Hurrians to the east (see CAUCASIAN LANGUAGES) were conquered around 1380 BC, and many Hurrian loanwords are found in Hittite texts. By about 1300 the Hittite kingdom dominated nearly all of Anatolia, and contested Syria with the Egyptians.

After the destruction of Hattušas by invaders, around 1200, Hittites gradually disappear from the historical and linguistic record.

The language of the tablets is usually divided into three periods, Old Hittite (to about 1550 BC, coinciding with what historians call the Old Kingdom), Middle and Late Hittite. Hittite was in many ways very close to reconstructed proto-Indo-European. Famously, it alone retained the postulated laryngeal consonants of the ancestral language (see box at INDO-EUROPEAN LANGUAGES). However, Hittite does not have the dual number or the masculine/feminine gender distinction which are otherwise typical of Indo-European languages, especially at earlier periods.

The numerals in Hittite are not fully known, because they were usually written in 'figures' rather than words in Hittite texts.

The Anatolian languages

There is evidence of two other languages in the Hittite texts. *Palaic* was spoken in northern Anatolia in Hittite times. *Luwian*, to the south, is known from cuneiform texts and also from a few hieroglyphic inscriptions (sometimes called 'Hieroglyphic Hittite'). Palaic and Luwian were both related to Hittite, and are grouped with it as 'Anatolian languages' within the INDO-EUROPEAN family (see map there). At least two languages contemporary with classical Greek – *Lydian* and *Lycian* – also belonged to the Anatolian branch; Lycian was in essence a later form of Luwian.

Hittite cuneiform

Hittite cuneiform script originated as an adaptation to a completely different language of the script that was first devised for Sumerian and

later used for Akkadian. The script carries its history with it: partly phonetic, it also includes word forms taken directly from Akkadian cuneiform script, and more than a thousand word forms ('Sumerograms') inherited from Sumerian by way of Akkadian. The way that Hittite scribes and readers would have pronounced these words is often unknown.

The cuneiform script was never well suited to Hittite, whose sound pattern was evidently very different from that of Akkadian. The native 'hieroglyphic' script actually survived longer, continuing to be used by speakers of Hittite in northern Syria long after the Hittite kingdom and Hittite cuneiform had passed into history.

HUNGARIAN

15,000,000 SPEAKERS

Hungary, Romania, Slovakia

One of the URALIC LANGUAGES, Hungarian is separated by thousands of miles from its close linguistic relatives, Khanty and Mansi, the 'Ob-Ugric' languages of the Ob valley in western Siberia.

Hungarians may be identifiable in history in the 5th century AD, as *Onogouroi*, a tribe driven from their Siberian home by a series of migrations of other peoples. If the identification is correct, then for some centuries thereafter Hungarian horsemen roamed the Russian steppe and the plains of central Europe, fighting variously for Romans, Khazars, Franks and others. At any rate, Hungarian speakers conquered the Pannonian plains, the country now called Hungary, in 895–6.

Here the Hungarian language developed under the influence of Old Slavonic, of medieval Latin and of German. In the 16th and 17th centuries Hungary was divided between the Holy Roman (German) and the Ottoman Empires. As Ottoman power waned, an Austrian Empire grew, an em-

pire which eventually had two governing languages, German and Hungarian, and many less privileged minorities. A policy of 'Magyarisation' imposed Hungarian as the first language of the south-eastern half of the empire: but this policy remained unpopular with most of the minorities, and in practice German remained the first lingua franca of Austria-Hungary. Many Hungarians have themselves been bilingual in German.

Hungarians call themselves *Magyar* (and their country *Magyarország*). This may be in origin a tribal name – identical with *Mansi* and with *Måsy*, which is the name of one of the two formerly exogamous clans of the Khanty.

All three peoples, Mansi, Khanty and Hungarians, were called *Ugri, Yugra* in early medieval times. From this useful word come the linguists' terms for several language groupings – Finno-Ugric, Ugric, Ob-Ugric – and also the usual foreign name for *Hungarian*.

Hungarian and the Ob-Ugric languages

After 1918 Austria-Hungary was broken up. The boundaries then drawn, which left many Germans and Hungarians in the novel position of minorities in new nation-states, have been a source of dissatisfaction for Hungarians ever since. Several major Hungarian settlements exist well beyond the borders of modern Hungary. The largest such group lives in northern and eastern Transylvania (*Erdély*, 'the forest country' in Hungarian). Hungarian speakers there number well over 1,500,000, the majority in a compact enclave in the south-east: these are known as *Székely* in Hungarian, *Secui* in Romanian. There are others in north-eastern Serbia, Ukraine and southern Slovakia. Those who lived in the *Csángó* enclave in Moldavia migrated en masse to western Hungary early in the 20th century, but perhaps 50,000 remain.

There are now many Hungarian émigrés in Britain, other European countries, Australia and the Americas. Mass emigrations took place after 1918, during the 1930s (particularly of Jews), after 1945, and after the 1956 uprising.

The early literature of Hungary was in Latin, from the Anonymous Chronicle of the 12th century to the classicising poetry of Czesmicze János (1434–1472; his Latin name is Janus Pannonius, 'John the Hungarian') under King Matthew Corvinus, a patron of scholarship. But some texts in Hungarian survive from as far back as the 13th century. The first Hungarian printed book appeared – at Krakow in modern Poland – in 1527.

In Hungarian, surnames precede Christian names; *-né* means 'Mrs'. *Plaget Jánosné*, 'Mrs John Plaget', was how one 19th-century authoress, who happened to be married to an Englishman, signed her novels.

Traditionally a Hungarian's full name included the village of origin. *Misztótfalusi Kis Miklós* is the full native name of the 17th-century typographer known abroad as 'Nicholas Kis', who was born in the small mining town of Misztótfalu (*Tăuţii*) in present-day Romania.

Literature in Hungarian flourished in the late 18th and 19th centuries. Hungary's 'national poet', Petöfi Sándor (1823–49), a fanatical demo-

crat, rose to the rank of major in the revolutionary army of the 1848 uprising and was killed at the battle of Segesvár (Sighişoara) in Romania.

Hungarian has 7 short vowels *a e i o ö u ü* and 7 long vowels *á é í ó ő ú ű*. Nouns and verbs are highly inflected, nouns having at least 238 possible forms.

Loanwords in Hungarian

Older Turkic loanwords in Hungarian are evidence that Hungarian nomadic horsemen worked with and learnt from the Volga Bulgars, who spoke a Turkic language (see CHUVASH). They include *alma* 'apple', *árpa* 'barley', *ökör* 'ox', *bor* 'wine', *tyúk* 'hen'.

Slavonic loanwords are a reminder that Pannonia had been Slavonic-speaking when the Hungarians occupied it, and that Slavonic-speaking missionaries converted them to Christianity: *káposzta* 'cabbage', *pohar* 'glass', *szent* 'holy', *templom* 'church'.

German loanwords (for example, German *Graf*, Hungarian *gróf* 'Count') demonstrate Hungary's long political coexistence with Germany and Austria. *Kávé* 'coffee' is a later loanword from Turkish: it arrived under Ottoman rule.

Hungarian and the Ob-Ugric languages

Speakers of Mansi, only about 4,000 in total, live among the eastern foothills of the Urals in the valleys of the Sos'va, the Konda and other tributaries of the Ob. They were traditionally called *Vogul* by outsiders after another river, the Vogulka, flowing through their territory.

Khanty or Ostyak speakers occupy a much larger but even more thinly populated region to their east, in the middle and lower Ob basin. They number over 15,000. *Khanty* means 'people': it is probably related to Hungarian *had* 'fighting men'. *Ostyak*, from Tatar *istäk*, was an outsiders' name for all the indigenous peoples of Siberia and their languages, sometimes applied also to Selkup (see SAMOYEDIC LAN-

GUAGES) and to one of the PALAEOSIBERIAN LANGUAGES.

Linguistic renewal – 18th-century style

There are fewer German words in Hungarian than there used to be, thanks to a movement begun in the late 18th century to reduce the German element in the language. 'Old words were revived, and full use was made of the resources of derivation. New compounds were fabricated, sometimes by mutilation, e.g. *csőr* ''bill, beak, nib'' from *cső* ''tube'' and *orr* ''nose''. Folk etymologies were invented, e.g. *szivar* ''cigar'' from the verb *szív-* ''suck''. The most remarkable example of successful rashness in these endeavours is *minta* ''sample, model, pattern'', which is a most current and indispensable word in modern Hungarian. It was taken from a Lapp [SAMI] dictionary – it is, as a matter of fact, a Norwegian loanword in Lapp, now obsolete in both languages – under the false pretence that it has a Hungarian etymon: *mint* ''as, like'' + *a* ''that one''.'

Bjorn Collinder, *An introduction to the Uralic languages* (Stockholm: Almqvist & Wiksell, 1960) pp. 32–3

Two thousand years ago, early Hungarian speakers, too, must have lived in the neighbourhood of the Urals. They occupied modern Hungary after a series of epic wanderings.

Numerals in Hungarian, Khanty and Mansi

	Hungarian	Khanty	Mansi
1	egy	ĭt	akwa
2	kettő	katən	kitïgh
3	három	khutəm	khūrŭm
4	négy	nyătə	nyila
5	öt	wet	at
6	hat	khut	khōt
7	hét	tapət	sāt
8	nyolc	nyŭwtə	nyollow
9	kilenc	yĭryan	ōntəllow
10	tíz	yang	low

IBAN

Perhaps 1,200,000 speakers

Indonesia, Malaysia

Iban, formerly known as *Sea Dayak*, is the mother tongue of about a third of the population of Sarawak, and of a large, scattered population in western and southern Kalimantan. It is one of the AUSTRONESIAN LANGUAGES, close to Malay and the Malayic languages of Sumatra. For a map and table of numerals see MALAY.

> *Dayak* was once used to denote several peoples of the rivers of Borneo, whose languages are not closely related. *Land Dayak*, as the name suggests, is an inland language of the valleys where Sea Dayak or Iban is spoken: it has about 50,000 speakers. *Ngaju Dayak*, a Barito language related to MALAGASY, is a language of 250,000 speakers in the valleys of south-western Kalimantan.

Speakers of Iban have traditionally settled in river valleys and are traders along river and coastal routes. Their language is a lingua franca throughout these valleys. They have a rich oral literature of ritual and epic, not all of which has been recorded. Their religion is strongly influenced by the Hinduism that came to the Malay Archipelago two thousand years ago, and the language shows Sanskrit influence.

Dialects of Iban

'It is homogeneous: dialect words are few and differences between areas of settlement are differences of accent, and of detail in custom and ritual. The extremes are, on the one hand, the so-called ''Saribas'' Iban of the northern Second Division, who have combined with the previous inhabitants, have long been closely associated with Malays, and were the first to profit from extensive rubber plantations and the opportunities of literacy afforded by mission schools; and, on the other, the pioneer communities of the Ulu Ai and Bajleh on the eastern flank of migration whose manner is bluff and whose language is short and rapid. The latter tend to regard the former as precious and a little comic; the former see the latter as uncouth and rustic. In between are the great majority known by the names of the other rivers of the Second Division.'

Anthony Richards, *An Iban–English dictionary* (Oxford: Clarendon Press, 1981) p. ix

---

Ngaju Dayak and the language of praise

Ngaju communities hold ceremonies at which oral poetry is performed in a traditional language 'for discussing the gods', *Sangiang*. Verses are spoken by an expert poet, and repeated in turn by four or six apprentices, who also accompany the words with a rhythmical drum beat. This poetry, usually celebrating or mourning a contemporary, is marked by two-line parallelism and by pervasive internal rhyme:

lii . . .

Nyaho hai mamparuguh tungkupah,	The great thunder exerts its power,
Kilat panjang mamparinjet ruang!	The long lightning ignites the heaven!

lii . . .

Tesek bewey kalingun Sambang hariak nanjulu, kilaw riak kalawaw kabantukan danaw, Pandang tege karendem garu haringki, tingkah pahi laut Bukit Liti!	Memory of Sambang begins to ripple like the ripples of a fish on the surface of a lake, Knowledge of him starts to well up like the waves of a river fish off Bukit Liti!

From Kma M. Usop, 'Karunya: the Ngaju Dayak songs of praise' in *Papers from the Third International Conference on Austronesian Linguistics* ed. Amran Halim and others (Canberra: Australian National University, Research School of Pacific Studies, 1982) vol. 3 pp. 319–24

ICELANDIC

250,000 SPEAKERS

Iceland

A descendant of OLD NORSE (see map at NORWE-GIAN), Icelandic developed from the speech of the Vikings who settled Iceland from 870 onwards and officially adopted Christianity in the year 1000.

The island was self-governing until 1262. Thereafter it was ruled by Norway, and along with Norway was transferred to Danish rule in the 14th century. Iceland became fully independent only in 1944. The revival of Icelandic as a literary and official language is thus a very recent event. The language is still in general closer than any of its relatives to Old Norse.

Icelandic names

Most Icelanders do not have surnames but patronymics. Forenames are followed by 'son of . . .' or 'daughter of . . .' and the father's name. The system is no different from that to be found in the sagas, and it was once used throughout Scandinavia.

In the absence of surnames, the Icelandic telephone book lists subscribers under their forenames.

The great literary works of Iceland – its prose sagas – date from the 12th and 13th centuries. They were written for recitation, in a simple and vivid language that had as yet hardly differentiated from Old Norse. Many, including *Brennu-Njals saga* 'The story of the burning of Njal' and *Laxdaela saga*, are anonymous. Ari the Learned (1068–1148) and Snorri Sturluson (1179–1241) were the greatest named authors of medieval Iceland. Snorri, with his *Heimskringla*, was responsible for the gathering of the histories of the Norwegian kings; in the so-called *Prose Edda* he set out a schema of Norse mythology, with many illustrative quotations from Old Norse poetry.

The study and printing of medieval Icelandic literature began in the 18th century. The priceless manuscripts collected by Arne Magnusson (the Arnamagnaean collection) were for a long time held at the Royal Library in Copenhagen, but have now returned to Iceland.

For a table of numerals see SWEDISH.

Entertainment with a purpose

Ketill Flatnefr hét einn ágætr hersir i Noregi. Hann var sonr Bjarnar Bunu, Gríms sonar hersis ór Sogni. Ketill var kvángaðr; hann átti Ingvildi dóttur Ketills veðrs, hersis af Raumaríki . . .

Ketil 'Flatnose' was a famous chieftain in Norway. He was the son of Bjorn 'Roughfoot', the son of Grim, a chieftain of Sogn. Ketil was married; he had as his wife Ingvild, daughter of Ketil 'Wether', chieftain of Raumarik . . .

Eyrbyggja saga, 1

Icelandic sagas kept alive the knowledge of how, and by whom, the country had been settled, and how each farming family was related to its neighbours. Typically a saga covers several generations of the local history of a district of Iceland – and deals with disputes over land ownership and status that result in litigation, feud and murder.

Igbo

12,000,000 SPEAKERS

Nigeria

Igbo is one of the four national languages of Nigeria, alongside Hausa, Yoruba and English. It belongs to the family of NIGER-CONGO LANGUAGES, and it is likely to have been spoken in its present location for many hundreds of years.

The meaning of *Igbo* or *Ibo* is uncertain: perhaps 'people', or possibly 'slaves', a name that might have been applied to Igbo speakers in the neighbouring Igala language. Certainly many of them suffered enslavement; Igbo speakers were already, perforce, crossing the Atlantic in the 17th century.

The absence of a centralised state among earlier Igbo speakers has contributed to dialect differentiation (see map). Neighbouring languages, including Isoko, Edo, Igala, Idoma, Ibibio and Ijo, have influenced the dialects to which they lie closest. English, more recently, has been a powerful influence on all Igbo dialects and on the standard language.

Retaining power

ọkwụrụ akaa onyẹ kụrụ ya, 'Okra never grows taller than the one who plants it'. A useful proverb, but is it literally true? Yes – because the farmer can always bend the plant to a convenient height.

The dialect used for the earliest religious publishing was not the mother tongue of any Nigerian. Originating in Freetown, Sierra Leone, it was a mixed dialect or lingua franca spoken among ex-slaves of Igbo origin who worked with European missionaries to bring Christianity to the Igbo speakers of Nigeria, beginning in 1841. The first printed book in Igbo, a primer and prayer book compiled by the YORUBA linguist Samuel Crowther, appeared in 1857 in this *Isuama* dialect, which was difficult for native Igbo speakers to understand.

Igbo in writing

For a long time two different Igbo spellings were in competition. The Old orthography, developed by Samuel Crowther and J. F. Schön, represented only six vowels, *a e i o ọ u*. The New orthography of Ida Ward and R. F. G. Adams, developed in the 1930s, gave all eight, but used three characters that were difficult to type and print, *a e i o u ɛ ɔ ɸ*. The Catholics accepted it but the Protestants did not. Finally, in 1961, the Official orthography devised by S. E. Onwu and others satisfied all difficulties. It represents all eight vowels, and its dotted vowels mark Igbo vowel harmony effectively: the 'odd' vowels *i e o u* have no dots, while three of the four 'even' vowels *ị a ọ ụ* are dotted.

Southern dialects have a separate set of nasal consonants, but these are usually not written. Nor are the high and low tones that are an important feature of the language.

'Union' Igbo was devised soon after 1900, as a mixture of five dialects, and was used for a Bible translation. 'Central' Igbo was a conscious development of the 1930s: it was another artificial standard, but a more successful one, based essentially on the Umuahia and Owere dialects.

This has been used even by some speakers of Onitsha Igbo, though Onitsha has a rather different consonant pattern: *l* for Central *r, n* for *l, f* for *h*. Standard Igbo, encouraged by the Society for Promoting Igbo Language and Culture, has been the most successful written standard of all. It shows relatively greater influence of the Onitsha dialect.

Since the introduction of the Official Orthography in 1961 (see box) the Igbo press and publishing have grown rapidly. Onitsha, centre of the earliest missionary activities, is a major market city and still a focus of Igbo written culture.

Igbo has three verbs 'to be': *-bu* to identify, *-nọ* to locate animate subjects, *-dị* to locate inanimates. The first ten numerals are: *otu, abụa, atọ, anọ, isẹ, isi, asa, asatọ, itẹghẹtẹ, iri.*

Igbo and its dialects

The major dialect division is between *Onitsha* to the north and *Owere* to the south. The dialect of Onitsha city is distinct, and has strong links with those of the west bank of the Niger. The old city of Bonny, on the coast, has its own Igbo dialect: it was once a major centre for the export of slaves.
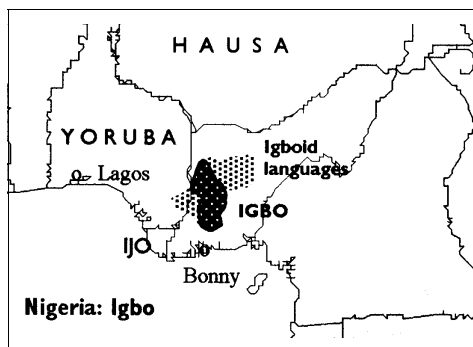
Nigeria: Igbo

IJO AND KALABARI

600,000 SPEAKERS

Nigeria

Ijo, one of the most important regional languages of southern Nigeria, belongs to the NIGER-CONGO LANGUAGES.

The name has been spelt in many ways: *Ijo, Ijaw, Izon, Udso, Djo* will all be found. Some members of the Ijo dialect continuum have been regarded as separate languages, including *Brass* or *Nembe* (see map) and the *Kalabari* of the old trading ports of Calabar and its neighbourhood.

Ijo speakers live in the Niger Delta, where travel is often easiest by water. Inland there is rich farmland, replenished by the Niger's regular floods; in the tidal regions near the coast the local economy depends on fishing and on trade.

The language is usually regarded as a separate branch of Niger-Congo, not closely linked to any of its neighbours. Its early history is unknown. The first recorded Ijo word occurs in Duarte Pacheco Pereira's geographical survey, *Esmeraldo de situ orbis*, in 1500: by this time, at any rate, Ijo speakers already lived where they do now. From the 16th to the early 19th centuries their trade flourished, as they held the middle ground between European mariners and the inland kingdoms. Calabar is only the most famous of the Ijo city states. It was also a slave port, from which many, Ijo and others, were sold across the Atlantic, where they were called *Caravali, Kalbary* or *Calabars*. With the British annexation of Nigeria, Ijo prosperity declined.

Writing and publishing in Ijo was begun by missionaries, who have used many dialects. Ibani or 'Bonny' was the first dialect to be used for Bible translations, beginning in 1870. In 1956 the complete Bible appeared in Nembe, which

was then expected to become the universal written language for all Ijo. This did not happen, but it remains a major standard alongside Kolokuma and Kalabari.

Kalabari itself has nearly a million speakers and is one of the languages used in broadcasting in Nigeria. There is some publishing in Kalabari, which is a lingua franca for speakers of smaller neighbouring languages such as Ido.

Ijo speakers are traditionally traders, and frequent travel has helped to keep dialect divergences within bounds. There is also intermarriage, and thus bilingualism, between Ijo and IGBO. English, and Nigerian Pidgin English, are widely known.

The first ten numerals in Ijo are: *kẹnẹ, mamọ, taro, nẹyi, shọnrọ, shendiye, shonoma, ningina, essiẹ, oyi.*

Dialects of Ijo

Ijo is spoken in many dialects, most often known by the name of their *Ibe* 'clan'. In general the dialects are mutually intelligible, but, for example, while Nembe speakers understand Kalabari

Nigeria: Ijo
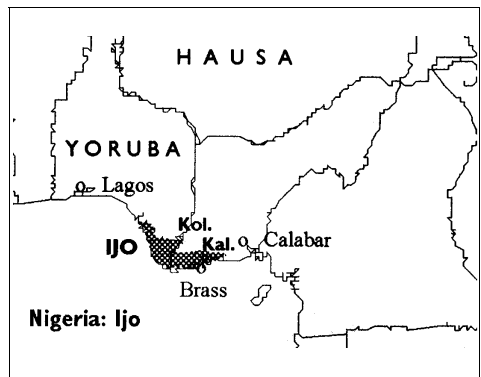
without difficulty, Kalabari speakers have more difficulty with Nembe.

Nembe or Brass-Nembe is the Southern Ijo standard language. *Akassa* is close to Nembe.

Kolokuma or Patani is the Northern Ijo standard. Two related dialects are *Gbanran* and *Ekpetiama*.

Kalabari forms a single dialect group, called 'Eastern Ijo', with *Okrika*, *Ibani* and *Nkoro*. Ibani, important in the 19th century, is now spoken only in part of Bonny town and a few villages; it has been supplanted by Igbo. In Calabar town EFIK is also spoken.

South Eastern, South Central, North Eastern and North Western dialect groups are also distinguished.

ILOCANO

5,000,000 SPEAKERS

Philippines

One of the AUSTRONESIAN LANGUAGES, Ilocano is the major language of the northern provinces of the Philippine island of Luzon (see map at TAGALOG). Its closest linguistic relative is Pangasinan.

> *Ilocano* – a Spanish form of the word – is the language of *Ilocos*. There are several variant spellings of the language name, including *Ilokano, Iloko*. An alternative name in local literature is *Samtoy*.

Ilocano is the everyday language in the provinces of La Union, Ilocos Sur, Ilocos Norte and Abra, but not in Batanes, where Ivatan is the local language. It is also widely spoken in Pangasinan, except the central region, where PANGASINAN retains its dominance. Ilocano had been studied by Spanish missionaries and was naturally one of the first languages in which the United States Army, newly charged with the rule of the Philippines, took an interest. Henry Swift, in his *Study of the Iloko language* (Washington, 1909), wrote of the people of the Philippines, 'he is best equipped and prepared to do his work, especially a government official, who can meet them on their own ground', and the army offered incentives to its officers to study Ilocano and other Philippine languages. Later, the syntax by Leonard Bloomfield, published in the journal *Language* in 1942, was a landmark in the study of Ilocano.

Ilocano is unique among the major Philippine languages – those of large, partly urbanised and mainly Christian peoples – as being the language of a traditional oral epic which survived long enough to be recorded by modern folklorists and anthropologists. Five versions of the story of *Lam-ang* were collected between 1889 and 1947.

Why did *Lam-ang* survive? Its setting is the frontier between Ilocano speakers and the 'blackest mountain', home of the Igorot tribe, who were never subdued during the four centuries of Spanish rule – and who had killed Don Juan, Lam-ang's father, before the hero was born. The threat of the Igorots was a real one until quite modern times, and the epic was as relevant as ever.

The first ten numerals in Ilocano are: *maysa, dua, tallo, uppat, lima, innem, pito, walo, siam, sangapulo.*

Ilocanos and Igorots

Ta idintocan a magtengnan	And when he reached
ti bantay a cangisitan	The blackest mountain
idiai Mamdili ken Dagman	In the towns of Mamdili and Dagman
dimmagus iti tallaongdan.	He went straight to the gathering.
Ket inna met nakitan	There he saw
daydi bangabanga ni amanan	His father's skull
ta napanda met sinarucang	Which had been placed on a stick
nga impasango iti dayaan.	And made to face the feast.

Text edited by Leopoldo Y. Yabes, 1935. From *Epics of the Philippines* ed. Jovita Ventura Castro and others (Quezon City: ASEAN Committee on Culture and Information, 1983) p. 72

Indo-aryan languages

The history of the Indo-Aryan group of languages goes back four thousand years. About that time, somewhere in central Asia, the speakers of 'Indo-Iranian' – one of the early INDO-EUROPEAN LANGUAGES or dialects – divided. The divergent group was to make its way towards northern India. Its dialect was an ancestral form of the modern Indo-Aryan languages, which are now spoken by over 600,000,000 people, in the northern part of the Indian subcontinent and in many other parts of the world.

The dialects of those left behind, who continued to occupy the central Asian steppe, and eventually spread into Iran, were the earliest distinct form of the IRANIAN LANGUAGES. The name *Ārya* was used by both groups. *Indo-Aryan* was invented by modern scholars to mean 'the Aryan languages of India' as opposed to those of Iran.

> *India* itself is a Greek and Latin term for 'the country of the River *Indus*'. The official name of India is *Bhārata*, the land of the Bharata tribe whose legendary story is told in the Sanskrit epic *Mahābhārata*.

The Indo-Aryan languages occupy a region where DRAVIDIAN LANGUAGES and AUSTROASIATIC LANGUAGES (and perhaps others now quite unknown) were spoken before them. They show influences from both these families. Development of written culture influenced other south and south-east Asian cultures as far away as Cham and Javanese.

Old, Middle and New Indo-Aryan

SANSKRIT is the earliest Indo-Aryan language of which texts survive. It may possibly date from 1000 BC or even earlier (though the written records are not so old). As a language of culture it has continued to influence the modern languages, its descendants, to this day.

Sanskrit was standardised by about 500 BC, and was no longer current in everyday speech. By that time spoken dialects, 'Prakrits', were developing very distinct forms. Prakrit dialects were written in the inscriptions of the Emperor Asoka (see below). Some of them also became literary languages in their own right, notably PALI and ARDHAMAGADHI. There are also texts in Niya Prakrit, the language of a community in Chinese Turkestan in the 3rd century AD, and in Maharashtri. This was the language of the Andhra or Satavahana Empire, centred on Paithan, in the 1st century BC to 2nd century AD.

Two of the Prakrits, called *Sauraseni* and *Magadhi*, are known for a special reason. In Sanskrit plays these dialects were put into the mouths of low class characters, who were not supposed to have learnt the 'Perfect' language of the well educated. They may be as close to the real speech of their day as are the country dialects of modern radio drama.

The root of all languages

Magadhi, one of the Prakrits, was the mother tongue of the Buddha according to a Pali commentary. It must therefore logically be the *mulabhasa*, the language which a child would naturally speak if it heard no other language spoken (*Vibhaṅgatthakathā* 387–8).

Some features of Maharashtri reappear in modern Marathi. Maharashtri was used in lyric poetry and in Jain literature. Sauraseni Apabhramsa, based on a dialect of the Hindi area of about 600 AD, was later used for literature all the way from Bengal to Gujarat, with some local variations. The period of the Prakrits and their

successors, from 500 BC to AD 1000, is known to linguists as Middle Indo-Aryan.

New Indo-Aryan is the period of development of the modern languages of the group. They begin to occur in literary texts from AD 1000 or soon after.

Modern Indo-Aryan languages cover a vast region of inland northern India across which dialects shade gradually into one another so that language boundaries are difficult to draw. For more information on the languages in this central zone see MAITHILI and MAGAHI (together called *Bihari*), BHOJPURI and CHHATTISGARHI (together called *Kosali*), HINDI, URDU, PANJABI, DOGRI. On the edge of this central zone to east, south and south-west are ASSAMESE, BENGALI, ORIYA, MARATHI, KON-KANI, GUJARATI, RAJASTHANI and SINDHI. Geographically separated, SINHALA and DIVEHI also belong to the group. So does *Saurashtri*, spoken by a colony of silk-weavers at Madurai.

Several Indo-Aryan languages are difficult to map as they are spoken by nomadic peoples. ROMANI, language of the Gypsies of Europe and America, is the most important. For *Gujuri*, spoken by 500,000 seasonal nomads of Indian and Pakistani Kashmir, see map at GUJARATI.

Lamani, also called Banjari, is spoken by well over a million nomadic people who call themselves Gormati – now mostly construction workers, noted for the colourful dress of their women – in Andhra Pradesh, Karnataka and Maharashtra, and elsewhere in India.

Linguistic influences are so complex in India that Indian linguists long ago set up a classification of word origins in the spoken languages of India. *Tadbhava* words are inherited from Sanskrit and have gone through normal processes of historical change; *deśya* 'country' words are at home in later Indo-Aryan yet they do not derive from Sanskrit; *tatsama* words are medieval or modern scholarly borrowings direct from Sanskrit. These distinctions are basic to modern research into linguistic history in many parts of the world – wherever a widely known language of culture, such as Old Slavonic, Latin or Sanskrit, has continued to influence the speakers of modern languages related to it.

In Himalayan valleys

In the wide north Indian plains easy communications have helped to reduce language differentiation. But in deep Himalayan valleys in Nepal, India, Pakistan and Afghanistan there are many little-known Indo-Aryan languages, spoken individually by small numbers.

Possible local origins of early Indo-Aryan languages
These languages were not all contemporary: Vedic Sanskrit may be as old as 1000 BC, while Apabhramsa Sauraseni dates from not long before AD 1000.

[Niya Prakrit]
[Vedic Sanskrit]
[Sanskrit]
[Ardhamagadhi]
[Sauraseni]
[Pali]
[Magadhi]
[Maharastri]

Indo-Aryan languages of the Himalaya
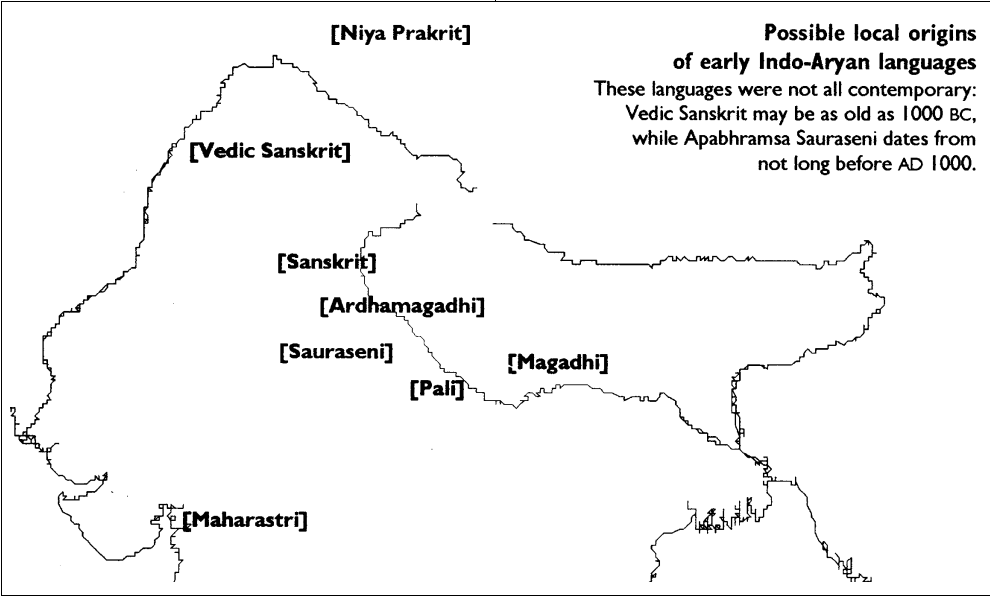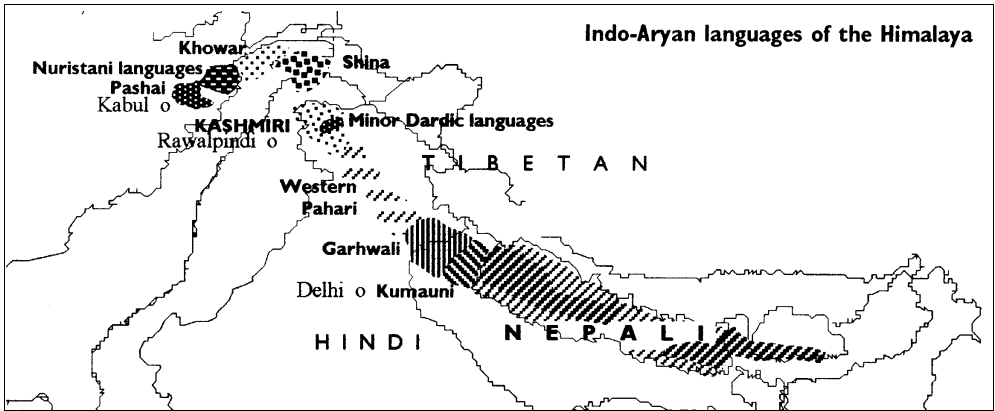
Pahari means 'of the hills', and is the convenient name for a group of Indo-Aryan languages from the Himalayan foothills north of the Ganges plain.

Languages of the Himalaya

The easternmost is NEPALI, and next to it lie the 'Central Pahari' languages, KUMAUNI and GARHWALI. The 'Western Pahari' languages have fewer speakers. Several were once the languages of small independent principalities. They include Jaunsari, Sirmauri, Baghati, Kiunthali, Sodochi, Mandeali (the largest, with over 200,000 speakers), Kului, Chameali and Bhadrawahi.

The Dardic languages, of which KASHMIRI has a separate entry, are spoken by small communities in Kashmir and north-eastern Afghanistan. The dialect of Srinagar is the main ingredient of Standard Kashmiri; to north and south, *Maraz* and *Kamaraz* dialects are distinguished. Other Dardic languages are Shina, with about 300,000 speakers, Kalami, KHOWAR and Torwali. Khetrani of Baluchistan may also be a Dardic language. For a table of numerals in Dardic languages see KASHMIRI.

Pashai, a group of mutually unintelligible dialects spoken in the smaller valleys north of Kabul, had a total of about 100,000 speakers before the current civil war.

The *Nuristani* languages, so different that they are usually considered a separate branch parallel with Indo-Aryan and Iranian, are spoken in remote valleys of north-east Afghanistan by a total of about 60,000 speakers. They include Kati or Bashgali, Waigali, Prasun and Ashkun.

Greeks in the Hindu Kush?

Nuristani (or *Kafir*, 'pagan') languages may belong to Indo-Aryan – or they may have separated even earlier from the mainstream, forming a separate small branch midway between Indo-Aryan and Iranian languages. Their discovery in the 19th century caused excitement.

'The more or less exact rumours about their fair complexion and hair and ''Nordic'' type, their wine-drinking and other European customs, their age-long resistance to the advance of Islam, have stirred the imagination of others beside the specialists; missionaries have cherished the hope that they were Christians of sorts; others have seen in them the descendants of Alexander's Greeks – an imaginative German traveller and scientist has [spoken of] ''this Homeric island in the interior of Asia'', and Kipling has based a romantic story, ''The Man Who Would Be King'', on the rumours abroad in India about these little known, but fascinating tribes . . . The geographical seclusion in which the Kafirs live in their inaccessible alpine valleys, and the religious and racial opposition between them and their Islamic neighbours, have deeply affected [their] social structure.'

Georg Morgenstierne, *Irano-Dardica* (Wiesbaden: Reichert, 1973) p. 298

Indian onomatopoeia

Among the words that are most clearly shared between Dravidian and Indo-Aryan languages are onomatopoeic terms. These are clearly not freshly invented in each language: their sound changes are regular, and in general they appear to be borrowed by early Indo-Aryan from an early Dravidian language. Some examples:

'Chew, grind the teeth'

Dravidian	Indo-Aryan
Kannada *kaṭakaṭa kadi* 'grind the teeth', Tulu *kaṭukaṭu* 'crunching noise'	Pali *kaṭakaṭāyati* 'snaps, crushes', Hindi *kaṛkaṛānā* 'chatter [of teeth]', *kaṭkaṭī* 'grinding of teeth'

'Jingle, tinkle'

Tamil *kaṇakaṇa-* 'to sound, rattle, jingle', Kannada *gaṇagaṇa* 'ringing of bells'	Sanskrit *kaṇakaṇāyita-* 'tinkling', Nepali *khankhanāunu* 'jingle', Marathi *ghanghan* 'jingle'

'Rustle, murmur'

Tamil *kalakalam* 'chirping of birds, murmur of a crowd', Kannada *kalakala* 'confused noise of a crowd'	Sanskrit *kalakala-* 'confused noise', Gujarati *kaḷkaḷvuṃ* 'to murmur', *khaḷkhaḷvuṃ* 'to rumble'

Examples from Murray B. Emeneau, *Language and linguistic area* (Stanford: Stanford University Press, 1980) pp. 269–71

Numerals in Nuristani languages and Pashai

	Bashgali	Waigali	Prasun	Pashai
1	ē, ev	ī, ek	ipin	ī
2	diu, du	du	lūe	dō
3	treh	trē	chī	hlə, hlē
4	što	štā	cipū	cār
5	puc	pūc	uc	panj
6	šo	šū	ušū	sə, khē
7	sut	sōt	sete	sat
8	ošṭ	ošṭ	aste	ašt, akht
9	noh	nū	nūh	nō
10	dič	dōš	leze	dē

From G. A. Grierson, *The Pisaca languages of north-western India* (London: Royal Asiatic Society, 1906) p. 37

Indo-Aryan in writing:
1. The Edicts of the Emperor Asoka

Asoka (c. *274–32* BC), grandson of Chandragupta Maurya, was one of the greatest of all rulers of India, and became a fervent believer in Buddhism. In many parts of India he erected stone inscriptions in the colloquial dialects of his time, the Prakrits, with instructions on *dharma*, or, to use a roughly equivalent Christian term, 'righteousness'. The Edicts of Asoka are the first surviving contemporary records of any of the Indo-Aryan languages of India.

In north-west India Asoka's stonecutters used an alphabet called Kharoṣṭhi, derived from an Aramaic script. Soon afterwards this is found on coins issued by Greek and Iranian monarchs in Central Asia. Kharoṣṭhi fell out of use long ago.

Elsewhere Asoka's inscriptions are in the Brahmi script. This survived. Gradually modified and adapted to different languages and writing materials, Brahmi is the ancestor of many alphabets used in south and south-east Asia.

Indo-Aryan in writing:
2. The children of Brahmi

The scripts that descend from Brahmi are more logical than any other. Devised over two thousand years ago on the basis of an accurate analysis of the sounds of Sanskrit and Prakrit, these alphabets have been adapted to very different languages, Indo-Aryan, Tai, Sino-Tibetan, Austroasiatic and Austronesian. All follow a similar alphabetical order, which is itself systematic and thus easy to remember. The box shows the basic

Devanāgarī alphabet of Hindi (almost identical with those of Marathi and Nepali) alongside corresponding characters in Bengali, Gujarati and *Gurmukhi* (for Panjabi). Ten vowels come first (there are additional vowel signs for *r* and *l* in most of the languages, but these are not shown here); then 25 stop and nasal consonants, in groups of five according to point of articulation, from the velars at the back of the mouth to the labials formed by the lips; then four fricatives; then four sibilants.

अ आ इ ई उ ऊ ए ऐ ओ औ कखगघङ चछजझञ टठडढण तथदधन पफबभम यरलव शषसह

অ আ ই ঈ উ ঊ এ ঐ ও ঔ কখগঘঙ চছজঝঞ টঠডঢণ তথদধন পফবভম যরলব শষসহ

અ આ ઇ ઈ ઉ ઊ એ ઐ ઓ ઔ કખગઘઙ ચછજઝઞ ટઠડઢણ તથદધન પફબભમ યરલવ શષસહ

ਅ ਆ ਇ ਈ ਉ ਊ ਏ ਐ ਓ ਔ ਕਖਗਘਙ ਚਛਜਝਞ ਟਠਡਢਣ ਤਥਦਧਨ ਪਫਬਭਮ ਯਰਲਵ ਸ਼ਸਹ

All scripts of the Brahmi family are based on the concept of one complete character per syllable: usually a central consonant shape to which vowel and second-consonant signs may be added, above, below or beside. The reader soon learns to dismantle these compound characters.

In Sanskrit, in each syllable in which no other vowel sign was added, the so-called 'inherent vowel', short *a* or *ə*, was to be read. This is one way in which the modern scripts are not so easy as they might be: nowadays one needs to know some fairly complicated rules, differing for each language, before one can tell whether an inherent vowel is to be read, or no vowel at all.

The Sanskrit word कात्स्न्र्य, *kārtsnya* 'fully', written in Devanagari script, is made up of two compound characters: the first is क *k* + आ *ā*, the second is र *r* + त *t* + स *s* + न *n* + य *y*. To the second character no vowel sign is added, so the vowel *a* is to be read. The *r* is written as a semicircle above the line.

The standard romanisation for Indic scripts, agreed by a congress of Orientalists in the late 19th century, can be applied to Sanskrit, to Pali and to many of the modern languages written in these scripts.

The standard romanisation

अ आ इ ई उ ऊ ए ऐ ओ औ कखगघङ चछजझञ टठडढण तथदधन पफबभम यरलव शषसह

a ā i ī u ū ṛ ḷ e ai o au k kh g gh ṅ c ch j jh ñ ṭ ṭh ḍ ḍh ṇ t th d dh n p ph b bh m y r l v ś ṣ s h

Indo-european languages

Indo-European languages are spoken on every continent and by members of every racial group. English, Spanish, French and Russian have some official status in dozens of countries worldwide. Hindi, Bengali, Portuguese and German are just as essential in many multinational contexts. Latin, Greek, Sanskrit, Pali and Avestan are classical languages of religion, philosophy and culture. Greek, Armenian, Yiddish and Romani are the languages of worldwide diasporas. There are many other Indo-European languages, from Icelandic to Italian and Persian, whose literatures enrich humanity.

How is it known that all these languages are related to one another?

In the 17th and 18th centuries a European scholar, having begun to master Latin and Greek and one or two modern European languages, might well go on to learn Hebrew – the language of the Old Testament – and perhaps Arabic. These were readily seen to be wholly different in their structure from the ancient and modern languages of Europe. Then European interest in India began to grow, and there was the opportunity to study Sanskrit, the classical language of India, under the guidance of Indian teachers. It was realised that Sanskrit – quite unlike Hebrew and Arabic – showed pervasive similarities with Latin and Greek and other languages of Europe. How did this come about? The speakers of these languages lived thousands of miles apart, and history told of no early contact between them. It was a puzzle that Indo-European scholars, ever since, have continued to explore.

Though not the first to notice the link between European and Indian languages, Sir William Jones put it most clearly into words, in his presidential address to the Asiatic Society of Bengal in 1786:

'The *Sanscrit* language, whatever be its antiquity, is of a wonderful structure; more perfect than the *Greek*, more copious than the *Latin*, and more exquisitely refined than either, yet bearing to both of them a stronger affinity, both in the roots of verbs and in the forms of grammar, than could possibly have been produced by accident; so strong indeed, that no philologer could examine them all three, without believing them to have sprung from some common source, which, perhaps, no longer exists: there is a similar reason, though not quite so forcible, for supposing that both the *Gothick* and the *Celtick*, though blended with a very different idiom, had the same origin with the *Sanscrit*; and the old *Persian* might be added to the same family'. (Sir William Jones, 'Third anniversary discourse, 1786' in *Asiatick researches* vol. 1 (1788) pp. 415–31).

The 'common source' that Jones deduced is now called proto-Indo-European. This may be the first time in English that the word *family* was applied to a group of languages.

Persian and Gothic had already been compared, as long ago as 1723, by the Swedish linguist Olaus Odhelius Andersson.

To the branches of the Indo-European family that Jones listed, later research has added Baltic, Armenian and Albanian; also some extinct languages rediscovered in the 19th and 20th centuries, such as Hittite and Tocharian.

In the 18th century the languages of Europe had sometimes been grouped together as *Japhetic*, named after one of Noah's three sons in the Biblical story of the Flood and its aftermath. To replace this name and to make clear the newly recognised link between European and Indian languages, the term *Indo-European* was coined by Thomas Young in 1818 in a review of Adelung and Vater's multilingual compilation *Mithridates*. German scholars have traditionally preferred *Indogermanische*. In the excitement engendered by the decipherment of Hittite, the term *Indo-Hittite* was briefly used as a statement that Hittite had separated from the Indo-European dialects slightly earlier than the remaining branches of the family. *Aryan* ('noble', the term that Vedic Sanskrit speakers used for themselves) was once popular as an alternative to Indo-European – but no longer, because of the racial connotations it picked up in European extremist politics.

It was clear that the known Indo-European languages must have gradually grown apart through changes in sounds: hence the difference in the initial consonants of English *brother*, Latin *frater*, Greek *phrater*, Sanskrit *bhrātr*, words that were clearly related both in sound and meaning. The essential breakthrough in Indo-European research was the gradual realisation that these sound changes were *regular*. In the development of each particular language or dialect they took place invariably, whenever a certain sound occurred in defined surroundings; if there appeared to be exceptions, the exceptions should in principle have explanations. In this way Indo-European studies helped to galvanise historical work on other languages too, for the same principles can be applied in all language families.

Once the regularity of sound change was accepted, it was possible to set up hypotheses about the sounds and the words of the parent language, 'proto-Indo-European': to make a formula (for example, **bhrāter* for 'brother') to which word forms in the known languages could be traced back.

The rebuilding of proto-Indo-European

One line of this research led to the gradual reconstruction of proto-Indo-European as an apparently natural language, with a complete sound system and a vocabulary to match. This is not a simple task. Some of the pitfalls become obvious as one studies the result.

The formulae or reconstructed forms are, in general, the most economical possible: such that the simplest possible series of changes will link them with known language forms. This makes them efficient as formulae, but it does not prove that they really existed in a parent language. The sound system of proto-Indo-European, variously reconstructed by generations of scholars, and sometimes given as many as four 'laryngeal consonants' (see below), differs noticeably from the sound systems of natural languages. The real proto-Indo-European can never have been a unitary form of speech but always a bundle of varying speech forms, 'sociolects' and dialects, as all real languages are.

The reconstructed vocabulary of proto-Indo-European is a collection of words that continued to be used, for some thousands of years, in *at least two* of the later branches of Indo-European. In the case of words that continued to be used in only one branch, there can usually be no proof of their ultimate origin, so they have to be discounted. Vocabulary is never static, especially as speakers migrate and as their economy and their society changes. Words are forgotten, new words are invented, and words are given new meanings. It is not surprising, after all this, that there are far fewer words in the proto-Indo-European dictionary than there are in any natural language.

Finally, proto-Indo-European itself was the product of a history, probably involving earlier migrations, certainly involving changes in society and economy. In all natural languages, history is reflected in inconsistencies, words whose meanings are undergoing change, words that some speakers misunderstand.

Still, linguists and archaeologists do try to pin down, from the vocabulary of proto-Indo-European, where its speakers lived, what kind of society they lived in, how they farmed, how

[Germanic]

[Baltic]

[Celtic]

[Slavonic]

Names of languages not

[Italic]

[Illyrian]

[Thracian]

Palaic

Hittite

Lydian

I r a

[Armenian]

Greek

Luwian/Lycian

they lived and even (Emile Benveniste, *Indo-European language and society*, 1973) how they thought. They try to identify a 'homeland' or (to use the German term) *Heimat* for the Indo-Europeans.

Heimat: the 'homeland'

Very different results have been reached. In the early 19th century some thought that Sanskrit itself, first spoken in north-west India, was the parent tongue. Sir Monier Monier-Williams, author of the great Sanskrit dictionary, saw the vast ruins of Balkh in Afghanistan as the city from which all Indo-Europeans traced their origin. Others looked to the north German plain.

Opinions shifted considerably in the next hundred years. For several decades now the majority of scholars has placed the 'homeland' on the south Russian steppes. The theory is particularly identified with the work of Marija Gimbutas, who presses the identification of the proto-Indo-Europeans with the builders of *kurgans* – burial mounds – in the eastern Ukraine.

A persistent minority looks south of the Black Sea. But these do not agree among themselves. The archaeologist Colin Renfrew (*Archaeology and language: the puzzle of Indo-European origins*, 1987) identifies the proto-Indo-Europeans with the earliest neolithic farmers of central Asia Minor, between 7000 and 6000 BC. Tomas Gamkrelidze and Vyacheslav Ivanov argue for an origin in what is now Armenia, perhaps equally early, followed by migration to the Ukraine, which then became a secondary 'homeland' and point of dispersal for the speakers of early Indo-European dialects. Robert Drews (*The coming of the Greeks*, 1988) also looks to Armenia – but as late as 1700 BC, when, he suggests, a chariot-riding elite spread their Indo-European speech across southern Europe and southern Asia.

The dialects of proto-Indo-European

Meanwhile, there has been work on the lines of descent from proto-Indo-European: how early, unrecorded dialects gradually differentiated into the widely different languages later known, and

Indo-European languages about 1250 / 750 BC
yet recorded at this period are in square brackets

[Tocharian]

m   i   a   n

Avestan

Sanskrit

Old Persian
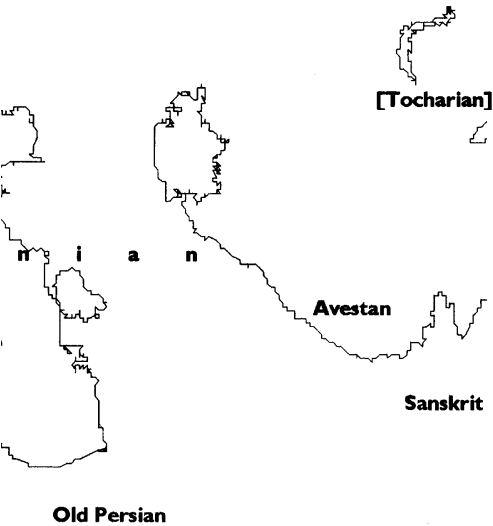
how these dialects were interlinked. They did not suddenly split. Just as with modern languages, there must for a long time have been a dialect continuum within Indo-European, marked by a succession of changes that affected different groups of dialects. On a modern dialect map the divisions are called *isoglosses*.

One, immediately noticeable, is called in linguists' shorthand the *centum/satem* split (see below). The innovation of the *satem* dialects – a very common one in language history – was the regular change from velar stops to dental fricatives when followed by a front vowel: thus, from proto-Indo-European **kṃtom*, Avestan *satəm* contrasts with Latin *centum* (pronounced *kentum*).

Keywords of proto-Indo-European research

Laryngeals: The Swiss linguist Ferdinand de Saussure (1857–1913) developed a theory that proto-Indo-European had a series of throaty consonants, 'laryngeals', which had disappeared in all the descendant languages but had left indirect evidence in the length and

---

type of vowels that had preceded them. Saussure could not specify how his laryngeals would have been pronounced, and the theory remained controversial – until the decipherment of Hittite, which turned out to have voiced and voiceless consonants (transliterated *ḫ*, *ḫḫ*) in exactly the position where Saussure had predicted his laryngeals.

*Loḱsos, salmon or trout? This word is found in many Indo-European languages, including German *Lachs*, Icelandic *lax*, Polish *łosoś*, Russian *losos'*. In these it means 'salmon'. If proto-Indo-European had a word for 'salmon', surely the language must have been spoken somewhere where salmon were to be found, that is, in northern Europe or Siberia. The argument has sometimes been thought powerful. But in other modern languages the same word means 'a kind of trout', Latvian *lasis*, Ossete *læsæg*, or simply 'fish', Tocharian A *laks*. This example shows how difficult it can be to place the proto-Indo-Europeans on the map: too often it depends on arbitrary decisions as to whether the meaning of words has shifted or remained fixed.

*Eḱwos, horse: The proponents of 'linguistic palaeontology' (research on proto-Indo-European society through the rebuilt vocabulary of the ancestral language) are roused to enthusiasm by the evident existence of a word for 'horse' in pIE, reconstructed from Sanskrit *aśvah*, Greek *hippos*, Latin *equus*, Old Irish *ech*, Anglo-Saxon *eoh*, Tocharian A *yuk*. They visualise the horse as a major element of pIE culture – a culture perhaps spread by warriors who conquered from horse-drawn chariots (like the Greek heroes imagined in Homer's *Iliad*) and who sacrificed horses (like those for whom a Vedic Sanskrit poet wrote the 'Hymn to the Horse', *Rgveda* 1.163). Some more cautious scholars consider that the horse was known, but not especially important, to the speakers of proto-Indo-European – and may have been known at first as a wild rather than a domesticated animal.

*kṃtom, the centum/satem split: The word for 'hundred' is an example of a sound change shared by almost half the descendant

languages of proto-Indo-European (see above). *K* followed by a front vowel became *s* or *sh* in Indo-Aryan (Sanskrit *śata*), Iranian (Persian *sad*), Slavonic (Russian *sto*), Baltic (Lithuanian *simtas*), Albanian (*qind*, pronounced *chind*) and Armenian. It remained *k* in Celtic (Welsh *cant*), Italic, Tocharian (*känt*), Greek (*hekaton*) and Germanic (with a subsequent move to *h*, English *hundred*).

This is a very common sound change in linguistic history: in fact a similar change has actually taken place, later, in many languages of the *centum* group. For example, Latin *centum* has become French *cent* (pronounced *sã*) and Spanish *cento* (pronounced *thento*).

Germanic, Baltic and Slavonic share some interesting features, such as the way in which the numerals '20' to '90' are built up. In the other Indo-European languages old compound words exist, such as Latin *triginta*, Greek *triákonta* for '30'. The corresponding terms in these north European languages are newly coined, and they have the obvious meaning 'three tens': Gothic *threis tigjus*, Lithuanian *trýs dẽšimtys*, Old Slavonic *triye desęte*, *tri desęti*.

What was proto-Indo-European like?

Like Sanskrit, Old Slavonic, Latin and Greek, proto-Indo-European had numerous forms for nouns (incorporating the notions of number and 'case') and for verbs (marking 'person', number, time and 'mood'). At some time in distant prehistory, an agglutinating language – in which these notions were separately identifiable as affixes – had become a synthetic language, in which they were fused with one another and with the noun or verb root. What is more, the fusion took different forms depending on the shape of the root. Students of Latin and Greek are all too familiar with the resulting noun declensions and verb conjugations whose forms cannot be wholly predicted and must be separately learnt. Proto-Indo-European, it seems, was already like this.

The later history of all the languages of the family has been of a gradual reduction in this complexity, and thus of a long term change of character. While the early Indo-European languages were largely *synthetic*, English and Hindi (to take two examples) are largely *isolating* languages.

The first ten numerals in proto-Indo-European may be reconstructed as *oinas, *dwō, *treyes, *qwetwōr, *penkwe, *sweḱs, *septm, *oḱtō, *enewen, *deḱm.

How did Indo-European languages spread so widely?

Colin Renfrew has argued that the spread of neolithic farming from central Anatolia to northern Greece and the Balkans is the archaeologically visible sign of the earliest Indo-European expansion. Most Indo-European specialists have not yet accepted Renfrew's argument, but it cannot be said to have been disproved. Renfrew's more tentative hypothesis that Indo-European languages spread also to Iran and northern India at this early period, along with neolithic farming, goes against the grain of the evidence.

The theory accepted by many linguists, though not by all, is that around 4000 BC an early group of Indo-European dialects was spoken across a wide swathe of central and eastern Europe, perhaps extending into southern Siberia and central Asia. The theory is often linked to the domestication of the horse – and certainly horsemanship has helped the spread of languages and empires, in this same steppe region, ever since.

Not long after 2000 BC the first surviving written records of Indo-European languages document HITTITE – the language of a powerful kingdom in Asia Minor – and the influence of an early form of Indo-Aryan on the Hurrian-speaking Mitanni in the Middle East. It is likely, meanwhile, that other Indo-European dialects were gradually spreading further into western Europe. Outlying and less accessible regions – Spain, Italy, southern Greece, northern Scandinavia – retained their non-Indo-European speech for rather longer. The speakers of early

Indo-Iranian dialects must by now have been approaching Iran and north-western India from the north. The hypothesis of their long-distance migration seems necessary to explain the links between Indo-Iranian languages on the one hand and Slavonic and URALIC LANGUAGES on the other.

At the time when the Romans were establishing their empire, early texts from Europe and Asia show with fair certainty the location of most of the then surviving branches of Indo-European: CELTIC LANGUAGES, GERMANIC LANGUAGES, BALTIC LANGUAGES, SLAVONIC LANGUAGES, LATIN and the other Italic languages, Illyrian (possibly the ancestor of ALBANIAN), Thracian, GREEK, ARMENIAN, the IRANIAN LANGUAGES, the INDO-ARYAN LANGUAGES and – far away to the north-east – the TOCHARIAN languages. The latter are not especially similar to Iranian and Indo-Aryan: their geographical location must result from a quite different migration.

Latin, spread widely by trade and empire, gradually evolved into the ROMANCE LANGUAGES. More recently English, French, Spanish and Russian have spread across the world: will they, in turn, split into daughter languages?

INUIT

ABOUT 65,000 SPEAKERS

Greenland, Canada and Alaska

Inuit is the only language of the Eskimo-Aleut family that is in widespread everyday use. It is often called Eskimo or Greenlandic, and is the language of one of the few peoples who have developed a way of life admirably suited to Arctic conditions. Their habitat, untrammelled by national frontiers, stretched from Greenland (where they are the majority population) across the northern edge of Canada and Alaska to the Diomede Islands (for map see ESKIMO-ALEUT LANGUAGES). The Eskimos of Big Diomede were moved to mainland Siberia and dispersed in 1948.

The East Greenland dialect, spoken around Angmagssalik and Scoresbysund, far across the icecap from the settled districts on the west coast, differs quite strongly from West Greenland – because of a persistent custom of word avoidance. Taboo has meant that new expressions are invented, old words are forgotten, and the whole vocabulary is renewed at a rapid rate. Knut Bergsland has used East Greenland to show how risky is the statistical approach to language history called glottochronology, a method that relies on counting similar words to estimate how many hundred years ago a pair of languages began to grow apart.

The West Greenland dialect is one official language of Greenland, the other being Danish. It is a language of education, publishing, broadcasting and administration, though it does not threaten the East Greenland (3,000 speakers) and Polar (700 speakers) dialects, which remain in daily use in their regions. In Greenland, Inuit is officially *Kalaallit Oqasii*, 'Language of the Greenlanders'.

Publishing in Canadian Inuit began in Labrador in the 18th century. At first it was entirely religious. There are now newspapers, government publications, schoolbooks and some creative literature.

A few Inuit words have been borrowed into English, including *kayak* and *anorak*. On maps of the Arctic, Inuit place names mingle with names imposed by explorers from the south. The northern Greenland settlement Thule (so named by classically minded Europeans after the legendary 'farthest north' of ancient Greek geography) is called *Qâmâq* in Inuit.

Numerals in Alaska and Greenland Inuit

Alaska Inuit		Greenland Inuit
atautseq	1	ataasiq
matleruk	2	marluk
pingayun	3	pingasut
stauman	4	sisamat
tatlîman	5	tallimat
aravinligin	6	arvinillit
matlerunligin	7	arvini-marluk
pingayunligin	8	arvini-pingasut
qulingúneritâran	9	qulingiluat
qulin	10	qulit

In Greenland, numerals above '12' are now usually borrowed from Danish.

The Inuktitut syllabary

In Greenland, Inuit is written in the Roman alphabet: so it is in Alaska and by some Canadians. In these countries there are several dif-

ferent spelling systems depending on the nationality of the missionaries who first began to write the language and on later changes of practice – linguists in Alaska and Canada have treated Inuit as a laboratory for spelling reforms. In Siberia, meanwhile, Inuit is (or was) written in Cyrillic. In Canada, since the mid 19th century, a syllabary (based on the one used for CREE) has competed with the Roman alphabet. The syllabary is most used by writers of the Inuktitut dialect as spoken in Quebec. The current version, called ᑎᑎᕋᐅᓯᖅ ᓄᑖᖅ *Titirausiq Nutaaq* or 'New Writing System', was adopted by the Inuit Cultural Institute in 1976.

ᑎᑎᕋᐅᓯᖅ ᓄᑖᖅ **The Inuktitut syllabary**

ᐃ ᐱ ᑎ ᑭ ᒋ ᒥ ᓂ ᓯ ᓕ ᔨ ᕕ ᕆ �qᖀ ᖏ ᖠ
i pi ti ki gi mi ni si li ji vi ri qi ngi łi

ᐅ ᐳ ᑐ ᑯ ᒍ ᒧ ᓄ ᓱ ᓗ ᔪ ᕗ ᕉ qu ᖑ ᖢ
u pu tu ku gu mu nu su lu ju vu ru qu ngu łu

ᐊ ᐸ ᑕ ᑲ ᒐ ᒪ ᓇ ᓴ ᓚ ᔭ ᕙ ᕋ qa ᖓ ᖤ
a pa ta ka ga ma na sa la ja va ra qa nga ła

H ᑉ ᑦ ᒃ ᒡ ᒻ ᓐ ᔅ ᓪ ᔾ ᕝ ᕐ q ᖕ ᖦ
h p t k g m n s l j v r q ng ł

A dot above a symbol stands for a long vowel: ᐲ *pii*, ᖖ *ngaa*. The last row, in the table above, gives the small symbols that are used for consonants not followed by a vowel.

Iranian languages

As the INDO-EUROPEAN LANGUAGES gradually differentiated, several thousand years ago, the earliest precursors of Balto-Slavonic and Indo-Iranian languages must have been spoken side by side. Already distinct from one another, the two groups show evidence of long term mutual influence.

If Indo-European speakers in general were an agricultural people, the ancestors of Indo-Iranian language speakers, occupying the plains north of the Black Sea, may have adopted the nomadism of the steppes.

All this is hypothetical. What is known is that when they first emerge into recorded history, over 2,500 years ago, *Old Iranian* languages were spoken and written in Persia, thousands of miles to the south-east of the Baltic and Slavonic peoples, their former neighbours. Some of them, not long afterwards, were certainly spoken by nomadic peoples. At the same date the INDO-ARYAN LANGUAGES come to light even further off, in northern India. And the agricultural terminology that is shared by most of the other Indo-European languages seems to be forgotten, or to have changed meaning, in Iranian and Indo-Aryan.

> *Īrān* was so named, long ago, as the country 'of the Arya', *erānām*. This name *Ārya* was used of themselves by speakers of several Iranian and Indo-Aryan languages (and was more recently taken up in Europe by racial theorists). The use of 'Iran' in place of 'Persia' as the name of the modern state is slightly misleading: Iranian languages have been, and still are, spoken over a much wider area than this.

The first Iranian people recorded in history are the Medes, who must have been in north-western Iran by the beginning of the first millennium BC. Almost all that remains as evidence of their language is a corpus of place names and personal names found in Akkadian and Greek texts. Around the same time the *Parsa*, Persians, are heard of, first in Kurdistan, then in south-western Iran: there are important inscriptions in Old PERSIAN, dating from the 5th and 4th centuries BC, when the Persian Empire, efficient and decentralised, extended its rule from Greece to the borders of India. Some AVESTAN texts – the sacred books of Zoroastrianism – are at least as old as this, though the manuscripts in which they now survive were copied many centuries later. Meanwhile the Scythians and Saka – still nomadic steppe-dwellers – ranged an even wider swathe of territory from the Black Sea coast to the borders of China.

The fall of the Persian Empire to Alexander the Great (died 323 BC) is a convenient beginning for the *Middle Iranian* period, which is considered to last until the Islamic conquest in AD 640. Texts are known in several Middle Iranian languages, including Middle PERSIAN, Parthian, Chorasmian, Bactrian and SOGDIAN. Further east, KHOTANESE is considered a late variety of Saka, the only one in which literary texts are known.

The modern Iranian languages

New Iranian languages are still spoken in Iran and central Asia. Nearly all of them have absorbed Arabic influences through the medium of Islamic culture. Those of Iran are now heavily influenced by Persian. The basic dialect division is between East and West Iranian.

Old Persian is already identifiable as a Western language. Its best known modern relative is PERSIAN, an official language in Iran, Afghanistan and Tajikistan. The provincial languages of Iran

are numerous – it is wrong to class them simply as dialects of Persian. This book deals with them under the headings LURI and GILAKI. Important beyond the borders of Iran are KURDISH and BA-LUCHI.
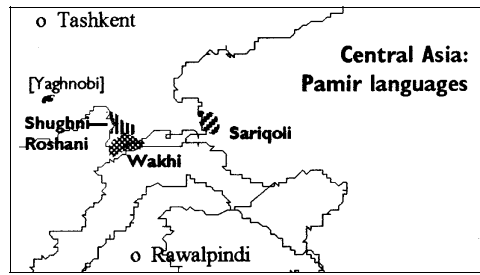
The Eastern group includes Avestan, Sogdian and Khotanese. Its major modern representatives are PASHTO, official language of Afghanistan, and the long-isolated OSSETE, which descends from the language of the Alans, a Scythian people. Sogdian's modern descendant is Yagh-nobi, now found in Tajikistan. There are many other East Iranian languages spoken by small communities in the valleys of the Pamir mountains.

The Pamir languages and Yaghnobi

Spoken in the mountains of north-east Afghanistan, Tajikistan and a tiny corner of China, the Pamir languages have a total of fewer than 100,000 speakers.

Shughni has at least 40,000 speakers in the valley of the Panj, which divides the Mountain Badakhshan district of Tajikistan from Badakshan province of Afghanistan. It is a second language in many of the Pamir valleys. *Roshani*, mutually comprehensible with Shughni, has about 12,000 speakers downstream.

Sariqoli is the Pamir language of China, with 10,000 speakers in Tashkurghan and the villages around. Their ancestors migrated from Shugh-

nan in Tajikistan. The language is sometimes misleadingly called *Tajiki*. It is historically close to Shughni, but is influenced by Uighur, the administrative language of Xinjiang.

Wakhi is spoken by about 25,000 people settled in the upper valley of the Panj, in south-east Tajikistan and the 'Wakhan corridor' of Afghanistan. The language is also spoken in high valleys of Hunza district in Kashmir.

The Pamir languages belong to East Iranian, but they have archaic features indicating that they split from the main body of Iranian dialects in very early times.

Yaghnobi, the modern descendant of SOGDIAN, was until 1971 spoken by about 2,500 people in the upper Yaghnob valley. The speakers then moved en masse to Zafarobod, still in Tajikistan, where they retain their language.

There are several other Pamir and East Iranian languages with no more than a few hundred speakers each.

Numerals in the Pamir languages				
	Shughni	**Roshani**	**Sariqoli**	**Wakhi**
1	yīw	yīw	iw	yiw
2	dhu	dhaw	dhɛw	buy
3	aray	aray	aroy	truy
4	cavōr	cavūr	cavur	cəbır
5	pīnzh	pīnzh	pinzh	panzh
6	khōgh	khūw	khel	shad
7	wūvd	wūvd	ıvd	ıb
8	wakht	wakht	wokht	at
9	nōw	nāw	new	naw
10	dhīs	dhos	dhes	dhas

From John Payne, 'Pamir languages' in *Compendium linguarum iranicarum*
ed. Rüdiger Schmitt (Wiesbaden: Reichert, 1989) pp. 417–44

IRISH

PERHAPS 80,000 SPEAKERS

Ireland

Irish is the only survivor among the CELTIC LANGUAGES (see map there) whose speakers remained outside the Roman Empire. It has the oldest literature in Europe after Greek and Latin.

> The language is often known as *Irish Gaelic* or simply *Gaelic*: the Irish spelling is *Gaeilge*. In this book GAELIC is used as the name of its Scottish offshoot.

Christianity was introduced to Ireland from Roman Britain: St Patrick, who lived in the early 5th century, was himself a Briton, and wrote in Latin. Irish authors were prominent in the Latin literature of Europe in the next three centuries. Meanwhile a literature in Irish was developing. Early hymns, including one attributed to Patrick, seem to go back to the 5th century. Heroic tales in prose, quite non-Christian in inspiration, were being written down by the 7th century, no doubt on the basis of earlier oral traditions. But most surviving manuscripts were written after the destructive Viking invasions of the 9th and 10th centuries.

Early Irish literature is prolific and varied. The best known of the prose tales (often called sagas) is *Táin Bó Cuailnge*, 'the *Táin*' for short, the 'Cattle Raid of Cooley'. The oldest form of this survives in the Book of the Dun Cow or *Lebor na hUidre*, compiled at the Monastery of Clonmacnoise in the 12th century. Alongside many other tales and rich lyric poetry there are chronicles, historical narratives, law codes and texts on grammar and poetics. *Dinnseanchus* is a book of place names – their etymologies and their legends – and there are other traditional dictionaries.

English colonisation began in the 16th century. Many chiefs, at whose courts the language and its literature had been cultivated, emigrated after the defeat at Kinsale in 1601. In the 17th to early 20th centuries Irish gradually gave way to English, which was now the only language of government. In the 19th century, with very few remaining monolingual speakers, schools worked to eradicate Irish. Famine and labour migration contributed to its decline, for English was needed by migrant workers.

At the division of Ireland in 1922 Irish gained the position of official language. As a symbol of independence it has had strong official encouragement. Government and civil service are, at least theoretically, bilingual. The language and its distinctive uncial script are widely seen in public. The Irish form of public titles and names has been borrowed into Irish English (and into the international English of television reporting): *Garda* 'police', *Taoiseach* 'Prime Minister', *Dail* 'Parliament'. Around 1,000,000 people claim to speak Irish.

But the officially defined *Gaeltacht*, the area where Irish is the majority everyday speech, is small and fragmented, with about 80,000 inhabitants. Even here modern population mobility threatens the survival of the language. The only Irish weekly (Sunday) newspaper, *Anois*, has a circulation of 5,000.

In 1922 Irish was still spoken by 20,000 people in remote areas in the six counties of Northern Ireland that remained British. Very few speakers remain there in the 1990s. There are communities of Irish speakers in London, Liverpool and Birmingham.

Sheldru or *Shelta* (also known as *Cant, The Old Thing* and *Gammon*) is a secret language that has long been used by Irish travellers and their descendants in England (where there has certainly been an Irish 'tinker' community since the late 16th century) and also in the United States. Its structure is English: its vocabulary is largely Irish, and many words are altered to ensure that outsiders cannot understand it.

Irish and English have coexisted for centuries and have influenced each other. Irish loanwords in English include *slogan* from *sluagh ghairm*

'army shout, war cry'; *whiskey* from *uisge bheatha* 'water of life, aqua vitae'; *bard*, a traditional oral poet.

According to Richard Stanyhurst in the *Description of Ireland* (1577), 'As the whole realme of Ireland is sundred into foure principal parts, so eche parcell differeth very much in the Irish tongue, euery country hauing his dialect or peculiar maner, in speaking the language.' There was a common saying, he added, that Ulster has only the right grammar, Munster has only the true pronunciation, Leinster has neither, and Connacht has both.

Sweeney mad

Ó'dchuala trá Suibhne sésdan na sochaidhe & muirn an mór/lúaigh nostógbaidh uime asin mbile re fraisnéllaibh na firmaiminti ós mullaighibh gacha maighni & ós fhégi gacha ferainn.

   Baoi fri ré chéiniarsin seachnóin Érenn ag tadhall & ag turrag a sgalpaibh cruadhcharrag & a ndosaibh crann urard eidhneach & i ccuasaibh caolchumhguibh cloch ó inber do inber & ó binn do binnd & ó glinn do glionn go ráinic Glenn mbitháluinn mBolcáin.

When Sweeney heard the shouts of the soldiers and the big noise of the army, he rose out of the tree towards the dark clouds and ranged far over mountains and territories.

   A long time he went faring all through Ireland, poking his way into hard rocky clefts, shouldering through ivy bushes, unsettling falls of pebbles in narrow defiles, wading estuaries, breasting summits, trekking through glens, until he found the pleasures of Glen Bolcain.

The legend of Sweeney, who was defeated at the battle of Moira in AD 637 and was driven mad, is told in mixed prose and verse in a 17th-century manuscript – but it is of much older origin. The prose passage quoted here is marked by rhythm and parallelism. Translation after Seamus Heaney, *Sweeney Astray* (Derry: Field Day Theatre Co., 1983).

Ogam script

b l v s n h d t c q m g ng z r a o u e i ch

Font: *Beth Luis Fearn, blf.ttf*
by Curtis Clark (jcclark@csupomona.edu)

Ogam script

Apparently based on late Latin handwriting, Ogam script may have been developed for writing on wooden message-sticks: straight lines are easier to carve than curves. About

a hundred stone inscriptions from Ireland are known in Ogam script, probably dating from the 5th and 6th centuries. In Britain some bilingual Ogam/Latin inscriptions have been found. The script may be called *beth luis fearn* and *beth luis nion* after the names of the first few letters.

   In the version shown in the box the letters ⊥, ⧣ and ⧤ are given their values in later manuscript texts: their values as originally used on stone inscriptions are not known. The script was usually written vertically, bottom to top, in inscriptions, horizontally in manuscripts. Ogam was still studied in scribal schools in Ireland down to the 17th century.

The Irish alphabet

Modern Irish script is a local variant of the Latin alphabet, originating in medieval manuscripts. Irish can also be printed in Latin typefaces, and often is, but the script is seen on road signs and public notices throughout Ireland.

Upper and lower case scarcely differ except for ꞃ and ꞅ. An acute accent on a vowel marks length. Consonants may be written with a dot above to mark aspiration. When Irish is written in Latin script the dot is replaced by *h*, and there are other differences between the usual spellings in Irish and Latin scripts.

Modern Irish script

ᴀ ƀ ᴄ ᴅ ᴇ ꜰ ᵹ ʜ ı ʟ ᴍ ɴ ᴏ ᴘ ꞃ ꞅ ᴛ ᴜ

ᴀ ƀ ᴄ ᴅ ᴇ ꜰ ᵹ ʜ ı ʟ ᴍ ɴ ᴏ ᴘ ꞃ ꞅ ᴛ ᴜ

ITALIAN

60,000,000 SPEAKERS

Italy, Switzerland, Eritrea

One of the ROMANCE LANGUAGES, Italian is a direct descendant of Latin – and is spoken in the same region where Latin first spread, in the peninsula of Italy, the Po river valley to the north, and the island of Sicily to the south.

> The peninsula was known as *Italy* even before Rome's power had spread across it, over two thousand years ago. *Italian* has become the usual term for the dialects spoken here, and for the standard language of united Italy – though this standard language is still sometimes called *Tuscan* and was once better known as *lingua Fiorentina*, the language of Florence, the cultural and political centre from which the standard variety of Italian has gradually spread.
>
> In Polish the term *Wloch* is used for 'Italian' – a word of the same origin as *Welsh* and *Vlach*, historically meaning 'speakers of a strange language'.

In the first few centuries after the collapse of the Roman Empire, those who wrote anything in Italy continued to write Latin. In fact there are few signs that a new language was developing: yet it was, for in the tenth century a few short texts, inserted in Latin documents (see box), show how far the language of everyday speech had changed from that of the Roman Empire.

Thus Italian emerges as fragments of real speech in records of lawsuits – and then as poetry. Led by King Frederick II, himself a poet, the 'Sicilian school' of the 11th century composed lyric verse that was influenced by the emerging poetry in OCCITAN. In the following centuries there was at first no standard Italian language. Instead, a succession of writers in different regions formed for themselves written dialects that were more or less strongly influenced by one another, by Latin (which continued to be the language of the great majority of texts) and by other languages.

Although it appears to outsiders to be a united country, Italy is historically a group of small and large states which were not politically united and some of which were under foreign rule or influence for long periods. There was every reason for local dialects to continue to differentiate. Besides, in northern Italy, often French-ruled, both French and Occitan were cultivated as literary languages.

The impulse towards an Italian linguistic standard was cultural rather than political. Crucial was the work of Dante Alighieri (1265–1321), who not only wrote the *Divine Comedy* – still the classic of Italian literature – but also, in his Latin study *De Vulgari Eloquentia* 'On Vernacular Literature', explored, in a wholly original way, the spoken languages of medieval Europe, especially the dialects of Italy, and their suitability for literature. Perhaps predictably, his choice fell on his own – the speech of Florence, capital of a small state and cultural focus of Tuscany in central Italy. Dante, Petrarch and Boccaccio, author of the *Decameron*, all wrote in the Tuscan of Florence. It was no coincidence that Tuscan gained respect in the 14th century, and that political and cultural circles elsewhere in Italy began to speak and write in varieties of Tuscan.

Latin was still pre-eminent (all three of these authors also wrote in Latin). It was only in 1495 that Leon Battista Alberti compiled the first grammar of Italian, a manuscript to which he

gave the Latin title *Regule lingue florentine* 'Rules of the Florentine language'.

In the 15th and 16th centuries Italian and Latin were both used for technical and scientific writing. Scholars spoke both languages: macaronic poetry made fun of the unskilled ones whose Latin was full of Italian words, while the Fidenzian style (named after the 16th-century pedant Pietro Giunteo Fidenzio of Montagnana) was an Italian full of Latin words and Latin forms. But the spread of printing slowly gave the advantage to Italian, which many more people could read and understand.

At the same time, Italian itself was becoming more standardised. The bitter argument of the 16th century – whether to write in Tuscan or in the 'language of the courts' – was a non-argument, for the courts were using a form of Tuscan. It was no longer the language of Florence, or even *lingua fiorentina in bocca sanese* 'Florentine words in a Siena mouth' (one 16th-century definition of the best Italian): Tuscan had become a lingua franca for cultured Italy, as it remains today. Out of many experiments with writing and printing Italian came the idea of using the Greek epsilon, ε, for the open *e* sound, a usage introduced by Giangiorgio Trissino in his play *Sophonisba* in 1524 and nowadays widespread among linguists.

Music, painting and architecture, fields in which Italy was at the forefront of European culture, had emerged as subjects for which Italian was the best language to choose. Italian terms are still used internationally in music – *cantata, allegro, andante* – and have to be learnt by students as they learn musical notation.

With the political upheavals of the Napoleonic period (when northern Italy once more came under French rule) and the unification of Italy around 1860 language questions were renewed. Once more, no alternative existed to Tuscan as a national language. Tuscan is in general the language of modern broadcasting and the press: of films, too, though the film industry, centred on Rome, has raised the profile of the city dialect of Rome. Local and regional forms of speech, within the spectrum of Italian, remain more distinct and more

Europe: Italian

important in Italy than in some other European countries.

Italian is now the official and everyday language of Italy itself and of the two small enclaves of the Vatican City and San Marino. It is one of the official languages of Switzerland, where there are half a million speakers. Italian is still spoken as a second language by many people in Malta (where it was once official).

Italian is still an important language of Eritrea, an Italian colony from 1890 to 1941. The press and education used standard Italian until, as an Islamic country, independent Eritrea recently began to favour Arabic. The Italian of everyday conversation in Eritrea is rather different from the Italian standard. Phonetically it resembles TIGRINYA: *borta* for *porta* 'door'; *tərobbo* for *troppo* 'very'.

Italian has spread widely across the world as the language of Italian émigré communities. The largest and longest-established are those of Chile, Argentina, Brazil (where *Fazendeiro* used to be the creole of a mixed Italian-Negro community in Sao Paulo), Australia, Canada and the USA. The Italian population of Britain numbers 200,000, and has been building up for several centuries, though larger numbers arrived from the mid 19th century onwards. Clerkenwell is the centre of the bilingual Italian community of London.

A distinct Jewish variety of Italian once had

many speakers in the Jewish quarters of north Italian cities such as Modena. There are now no more than a few hundred speakers at the most. A Jewish variety of the Venetian dialect used to be spoken by the Jews of Corfu: this community was wiped out during the Second World War.

Italian is the source of many loanwords in English and other European languages. They include military terms such as *colonello* 'colonel', *sentinella* 'sentinel'; words linked to the arts like *sonetto* 'sonnet', *pantalone* 'pantaloon' (originally a character in comedy); and many others, such as *cortigiana* 'courtesan' and *influenza*, literally 'influence'.

For a table of numerals see SARDINIAN.

Italian, Sardinian and Corsican on the map

Italian is the majority language of Italy, also spoken in part of Switzerland. Standard Italian is based on the Tuscan dialect of Florence. Regional varieties of Italian shade into the city dialects (including that of Rome) and into strongly marked local dialects. The northern dialects share some important features with French and Occitan. Sicilian, sometimes regarded as a distinct language, has a literary tradition older than that of Italian itself.

Sardinian is the regional language of Sardinia, though it has little official recognition. The major dialects are Campidanese and Logudorese. Gallurese has many links with Corsican, while Sassarese is influenced by the mainland dialects of Pisa and Genoa. In Alghero the local language is Catalan; in Calasetta and Carloforte the Ligurian dialect of Italian is spoken.

Corsican is the regional speech of Corsica, where the official language is French. In Bonifacio in Corsica a Ligurian dialect of Italian is spoken, implanted long ago by colonisation from Genoa.

The Voice of the Lombard People

The 'Northern League', which presses for autonomy or independence for the Italian North, has an ambiguous attitude to the northern dialects. These are satisfyingly distinct from standard Italian – but lack the cultural cachet of the language of Dante. Some separatist writing does appear in dialect form, such as the small-circulation paper *Lombardia Autonomista: la vus del popul lumbard*. One 1986 issue opened with the following clarion call:

Lumbard, muvemas tücc, e drelamen, perchè Roma l'autonomia ghe la regala minga de sücür. O semm bun de cunquistala nun, o'l noster popul el scumpariss de la storia.

Lombards, let us move forward, and quickly, because Rome will not serve us autonomy on a plate. Either we are up to conquering it for ourselves, or our people will be erased from history.

Indovinello Veronese: the 'Veronese riddle'

The oldest Italian text is generally said to be the riddle that a scribe of the 9th century inserted in the margin of a religious manuscript:

Se pareba boves,	The oxen were driven out;
alba pratalia araba	they ploughed a white field
e albo versorio teneba	and held to a white furrow
e negro semen seminaba.	and sowed a black seed.

The answer, appropriately, is the scribe's own hand as he writes his text.

JAMAICAN CREOLE

2,200,000 SPEAKERS

Jamaica

One of the ENGLISH CREOLES AND PIDGINS of the Caribbean, Jamaican Creole originated in the heavy immigration of African slaves to Jamaica after the English took the island in 1655. The slaves came from many different African cultures: at certain times, to judge by word origins in the modern Creole, AKAN and TEMNE speakers must have been strongly represented. But the English pidgin of the Atlantic trade was their only means of communication, and this soon developed into a new language that mixed English elements with Portuguese, African and others – a language as necessary to the white inhabitants of Jamaica as to the slaves.

Through all the centuries that followed, English has been the language of officialdom and educated culture on the island. The creole has historically been despised by teachers and by educated Jamaicans, who, in speaking to outsiders, often found it best to deny any knowledge of it. In fact there are many language varieties (see box) between 'pure creole' and standard English, and skill in switching between them is learned early in life: children speaking to a teacher will use neither creole nor standard, but something between.

A sociolinguistic continuum

'The varieties of Jamaican English differ to the point of unintelligibility. There are many middle-class St Andrew housewives who claim that they can speak the broad creole because they can converse with their maids, yet they can understand very little of the conversation if they hear the maid talking with the gardener.

'In Jamaica there is no sharp cleavage between creole and standard. Rather there is a continuous spectrum of speech varieties ranging from the 'bush talk' or 'broken language' of Quashie to the educated standard of Philip Sherlock and Norman Manley. Many Jamaicans persist in the myth that there are only two varieties: the patois [Creole] and the standard. But one speaker's attempt at the broad patois may be closer to the standard end of the spectrum than is another speaker's attempt at the standard. Each Jamaican speaker commands a span of this continuum, the breadth of the span depending on the breadth of his social contacts; a labor leader, for example, commands a greater span of varieties than does a suburban middle-class housewife.'

David DeCamp in *Pidginization and creolization of languages* ed. Dell Hymes (Cambridge: Cambridge University Press, 1971) p. 350, abridged

JAPANESE

120,000,000 SPEAKERS

Japan

Japanese is something unique: one of the major languages of the world, spoken by well over a hundred million people, yet with no known linguistic relatives.

Naturally, attempts have been made to find links between Japanese and other languages. Research has concentrated on an apparent, very distant, relationship that might exist with the ALTAIC language family. But the results so far are anything but certain, as can be seen from a recent survey article which speculates whether Japanese may be 'a mixed language, deriving its lexical and grammatical properties from both Austronesian and Altaic, possibly with additions from Austro-Asiatic and Dravidian' (M. Shibatani). To have undergone such varied influences a language would have to be sited at a crossroads of Asia, rather than on an offshore archipelago. In fact, in historic times, the Japanese islands have been involved rather little in migration and ethnic mixture – a fact that certainly helps to explain why Japanese has no obvious links to other known languages.

Japanese is known to its own speakers as *Nihongo*.

Although the two languages are utterly different from each other, Chinese has exerted enormous cultural influence on Japanese throughout its known history – from the very beginning, the 8th century AD, when the first two Japanese texts were written down in Chinese script. These were *Kojiki*, 'Records of Antiquity', a miscellany of myths and legendary history; and *Manyōshū*, 'Collection of Ten Thousand Leaves', an anthology of poetry. Some of the texts in these collections are clearly rather older than the 8th century. Japanese has been a written and literary language for about as long as English.

In later centuries Chinese loanwords multiplied, yet the styles and genres of Japanese literature tended to remain distinct from those of Chinese – indeed, Japanese intellectuals also wrote in Chinese. Medieval Japanese poetry and prose have one striking feature which makes them a landmark in world culture: many of the greatest masterpieces were written by women. This includes the single greatest classic of Japanese literature, the 11th-century *Tale of Genji* by Lady Murasaki.

Japanese and Chinese continued their symbiosis until the mid 19th century: Chinese loanwords are said to amount to more than half of the Japanese vocabulary. Then Japan opened itself to Western influence, almost for the first time. It now has numerous loanwords from European languages, notably English: tēburu 'table', wāpuro 'word processor', *rampu* 'lamp'. Kōhī 'coffee' is an older loanword, from Dutch.

With political and commercial expansion, Japanese became a major language in the western Pacific and eastern Asia at this period. It was an official language in Korea, Malaysia, Indonesia, Singapore and other countries while they were under Japanese rule.

In spite of the collapse of Japanese imperialism at the end of the Second World War, Japanese remains an important language of worldwide trade: two Japanese daily newspapers, *Asahi shinbun* and *Nihon keizai shinbu*, are now published simultaneously in Europe.

Japanese is a language in which politeness,

Japanese
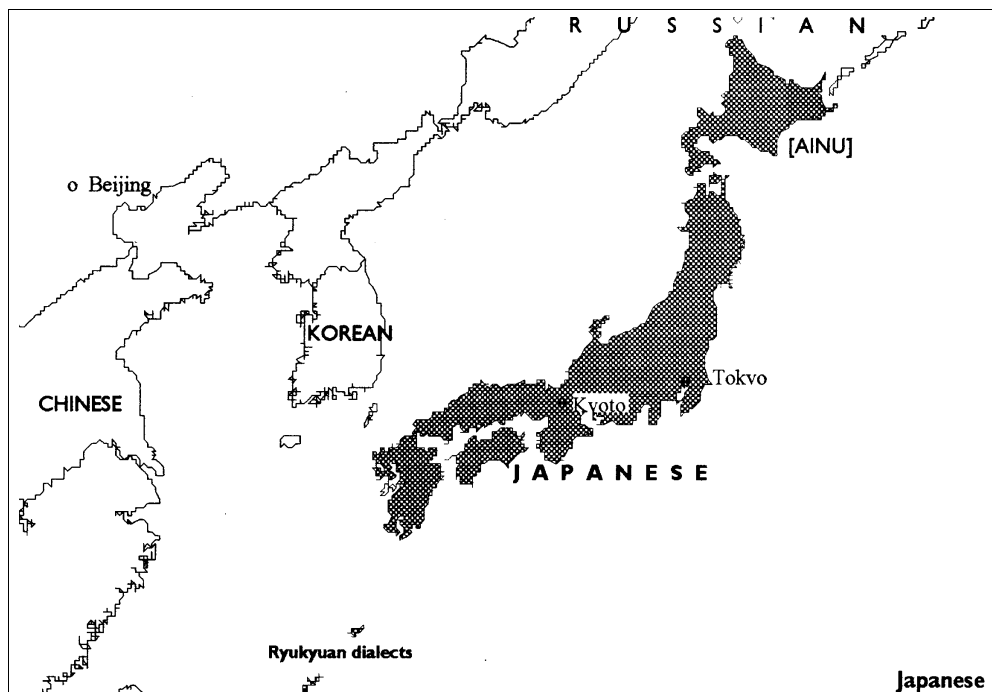
deference and euphemism are built in very firmly. So are distinctions between men's and women's speech; there are words categorised as 'rough' which women do not use, and women's speech tends to have fewer Chinese loanwords than men's.

One feature of the Japanese sound pattern is well known: it has only one phoneme corresponding to the *l* and *r* of many other languages. Japanese speakers of foreign languages find the distinction between these two sounds difficult to make.

Japanese dialects

Over most of Japan dialect differences, though significant, do not impede understanding because a standard form of Japanese (*kyōtūgo* 'common language'; *hyōzyungo* 'standard language') based on the Tōkyō dialect is used in broadcasting and education and is generally understood. The major dialect division is between Western dialects, including that of the old capital Kyōto, and Eastern, including Tōkyō.

Ainu

A second language native to Japan, once spoken in northern Honshū and the islands of Hokkaido and Sakhalin but now practically extinct, Ainu is quite unrelated to Japanese. Like Japanese it has no proven links with any other language of the world. The Ainu were distinctive in their language, their culture – and in being hairier than any other human population. Once an independent fishing and hunting people, they were conquered by Japan in 1669–72 and have very slowly assimilated to Japanese culture.

There was no written literature in Ainu but rich collections of oral epic poetry have been made. Ainu epics were usually in the first person, the imagined speaker being sometimes a god or goddess, sometimes a human being or an animal. They were most often performed by women: one, Imekanu (1875–1961), after her retirement from work as a Christian missionary, wrote down more than 20,000 pages of epics from her own repertoire.

Ryukyuan dialects are quite different from those of the main islands, though still recognisably related to Japanese. They are on the way to extinction, as young people no longer learn them.

Japanese numerals

Chinese		Japanese
ichi	1	hitotsu
ni	2	futatsu
san	3	mitsu
shi, yo, yon	4	yotsu
go	5	itsutsu
roku	6	mutsu
shichi, nana	7	nanatsu
hachi	8	yatsu
ku, kyū	9	kokonotsu
jū	10	tō

The native Japanese numerals for '1' to '10' survive in some traditional uses. Numerals borrowed from Chinese are much commoner: they are always used for counting people, and always for counting above ten. But *shi* for '4' and *shichi* for '7' are often avoided because they sound like *shi* 'death': short forms of the Japanese numerals are substituted.

Japanese in writing

Between the 9th and 12th centuries a syllabary of 46 characters was devised for Japanese, on the basis of Chinese characters. From this point on, syllabic script was intermixed with Chinese ideograms in every Japanese text. Although all Japanese words could be written in syllabics, this does not happen in practice. The syllabics were used to write prefixes and suffixes (the most obvious feature in which Japanese differs from the essentially monosyllabic Chinese), and nowadays they are also used for loanwords from English and other Western languages.

There are in fact two syllabaries, *hiragana* and *katakana*, related in their origin and identical in their structure but kept apart in their use. Japanese words are usually written in a combination of ideograms and *hiragana*, so these are found together in every text. *Katakana* is used in modern loanwords and in spelling out unfamiliar sounds – and is the script used in telegrams.

The characters that are used in the Chinese way as ideograms are known as *kanji*. About 2,000 of these form part of the official modern script: they are real ideograms, typically read with several different sounds depending on the context. Children learning to write begin with the syllabaries, but they are expected to learn nearly 900 *kanji* characters during the six years of primary school.

For a hundred years the Hepburn system of Romanisation (*rōmaji*) has been the standard way of converting Japanese to the Latin alphabet.

JAVANESE

75,000,000 SPEAKERS

Indonesia

Three AUSTRONESIAN LANGUAGES are commonly spoken on the heavily populated island of Java. Malay is the lingua franca of Indonesia's national capital, Jakarta: its official form, Indonesian, is the language of government and the press. Outside the capital, Sundanese is the language of the west of the island, Javanese of the centre and the east.

Records of Java's history begin with an exciting series of 8th-century stone inscriptions in Sanskrit – for Java, whether seven kingdoms or one, was already a Hindu realm, Indian in religion, Indian in culture, Indian in the language of high culture and government. Buddhism, too, found its way to Java. The Indian influence is as clear as the astonishing original inspiration in the ruins of Borobudur, which date from this period.

The oldest inscription in Javanese is dated to 806. Scholars call the script of these early inscriptions *Kawi*. This is properly the name of something quite different – the poetic language of the *wayang* puppet plays and other traditional literature of modern Java – but scholars often get things wrong. With the arrival of Islam, Arabic script was introduced to Java by about 1400, and for a long time this script was used regularly for Javanese.

Javanese written literature, beginning with *kakawin* and *prawa*, poetry and prose of Sanskrit inspiration, has flourished for a thousand years, but the oldest texts are known only from later copies. The earliest surviving palm leaf manuscripts date from the late 14th century. Several layers of literary inspiration can be traced, associated with distinctive language varieties: the Old Javanese of the *kakawin*, the Middle Javanese of the *kidung*, poetry of local forms and

inspiration, and the Modern Javanese that is dated from the 16th century onwards, after Islam was established.

Full of Sanskrit loanwords, Javanese has also been influenced for many centuries by Arabic and Malay. It has in its turn influenced Balinese, Sundanese, Madurese, Malay, Banjar and numerous other languages of the Malay archipelago. *Kawi*, a literary form of Old Javanese, is still the language of court and ritual in Bali.

As many as seven million Javanese speakers have settled outside Java, many of them resettled under the Indonesian government's *transmigrasi* programme. There are about 60,000 Javanese speakers, descendants of migrant workers, in the former Dutch territory of Suriname.

Sundanese, Javanese and Madurese on the map

The dialects of these three languages cover Java and the adjacent smaller islands – except Greater Jakarta, where the predominant language is Indonesian (MALAY).

**Indonesia:
Sundanese, Javanese and Madurese**
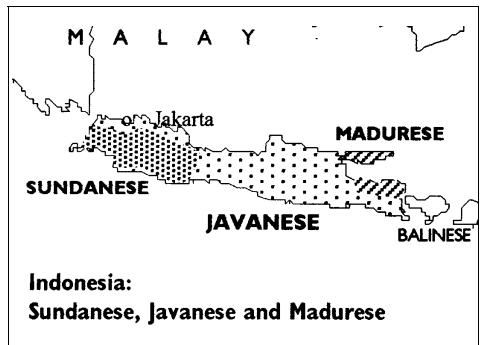
Javanese has western, central and eastern dialect groups. The dialect of the former government centres of Surakarta and Yogyakarta sets the standard. *Osing*, spoken around Banyuwangi, is particularly distinctive. Tengger has 500,000 speakers on the slopes of Mt Bromo in eastern Java.

SUNDANESE dialect groupings follow traditional administrative subdivisions: Banten, Priangan, Bogor and Krawang, and Cirebon.

MADURESE is usually divided into western, central and eastern dialect groups, though communications on the island are easy and sharp dialect differences have not emerged. The Madurese of the Kangean islands is so different from that of Madura that some would class it as a separate language. That spoken on the north-east coast of Java is apparently the result of relatively recent migration: separate mainland dialects have not been recorded.

The speech levels of Javanese

Javanese speech takes account, in every sentence that is spoken, of the relative status of speaker and hearer. The practice is sometimes said to have been learnt from the Hindu caste system – but it has perhaps nowhere been built into spoken language so thoroughly as in Java. It affects tone of voice (which varies between *alus* 'refined' and *kasar* 'crude') and, most especially, vocabulary. Choice of words ranges between *ngoko* 'low' and *krama* 'high'. *Ngoko* is a complete language, in which everything can be said that one ever needs to – but it cannot be used in formal situations. In talking of social superiors, even more formality is needed – *krama inggil*, 'high *krama*', referring to the honoured person's actions; *krama andhap*, 'humble *krama*', referring to one's own.

For a table of numerals in Ngoko and Krama see SUNDANESE.

Ngoko, high *krama* and humble *krama*

Ngoko	*Aku* **negekeki** *kancaku buku*	I **gave** my friend a book
Krama andhap	*Aku* **njaosi** *bapakku buku*	I **gave** my father a book
Krama inggil	*Bapak* **maringi** *aku buku*	My father **gave** me a book

English speakers can and do make such distinctions – 'I **was privileged to offer** . . .', 'The Director **was kind enough to give** . . .' but in Javanese they are far more systematic. There are dialect distinctions in the *krama* language, which has a larger vocabulary in eastern Java than in the west of the island.

Example from Elinor Clark Horne, 'Javanese' in *International encyclopedia of linguistics* (New York: Oxford University Press, 1992) vol. 2 pp. 254–9

The Javanese alphabet

Alphabetical order, in this script of Indian origin, does not follow the otherwise universal Indian rule. Instead, the following doggerel verse acts as a mnemonic:

hana càraka,	ꦲꦤꦕꦫꦏ	There were two envoys,
data sawala;	ꦝꦠꦱꦮꦭ	They began to fight;
padha jayanya,	ꦥꦝꦗꦪꦚ	Their prowess was equal,
maga bathanga!	ꦩꦒꦨꦛꦔ	Both fell down dead!

After J. C. Kuipers and R. McDermott in *The world's writing systems* ed. P. T. Daniels, W. Bright (New York: Oxford University Press, 1996) p. 478

JINGHPAW

700,000 SPEAKERS

Burma, China, India

One of the SINO-TIBETAN LANGUAGES, Jinghpaw is a major language of Kachin state, Burma. It has been described as the 'linguistic cross-roads' of the Sino-Tibetan family, showing similarities with Tibetan to the north, with Burmese to the south and with Yi and its relatives to the east.

Edmund Leach, in *Political systems of highland Burma* (London: Bell, 1954), has shown that the cultural and linguistic position of its speakers is equally fluid. There is a continuing, complex pattern of change in which Kachins may adopt a new language, Jinghpaw, Shan or Burmese, which tends to go with a new social orientation.

Jinghpaw is the local form of the name. *Singpho*, *Chingpho* and *Jingpo* are spellings adopted in various foreign sources; the Burmese form is *Theinbaw*. Jinghpaw has sometimes been called 'the *Kachin* language' (*Kakhyen* is an older form of this name): the majority of inhabitants of Kachin state speak either Jinghpaw or Burmese or both.

Kachin country is mountainous, with deep valleys. The two major towns are Myitkyina and Bhamo, both of them traditional nodes of long distance land trade. But little is known of the history of Jinghpaw and its speakers before the 19th century. The border between China and British Burma, cutting through Jinghpaw-speaking country, was fixed in the 1890s.

Jinghpaw has borrowed vocabulary from Shan and, rather less liberally, from Chinese and Burmese. It has also borrowed from English, particularly for religious and cultural concepts, as a result of the work of American Baptist missionaries who were active in Bhamo by 1900. They devised the Latin orthography in which Jinghpaw is normally printed.

The first ten numerals are: *ləngai*, *ləhkawng*, *məsum*, *məli*, *mənga kru*, *sənit*, *məcat*, *jəhku*, *shi*.

### Three uses of traditional Jinghpaw literature

'Traditional Kachin culture tolerates premarital sexual relations. The front apartment of a house is the *nla dap* where adolescents gather in the evening for singing, recitation of poetry, and, finally, lovemaking.

'The priest's chief task is to offer sacrifices to the spirits with appropriate invocations. Most of his training consists in learning prayers and formulas, some of which are quite long, in a special, obsolescent style of speech. The priest is recompensed with a part of the sacrifice. An additional specialized status is that of *jaiwa*. Although older sources regard this as a religious status, their descriptions support the term 'saga-teller' used by Leach. Certain rites are accompanied by lengthy recitations of traditional cosmogony and history, of which the saga-teller is the repository.'

Frank M. LeBar and others, *Ethnic groups of mainland southeast Asia* (New Haven: Human Relations Area Files Press, 1964) pp. 15, 17

Jinghpaw on the map

In Kachin state, Burma, where over 500,000 Jinghpaw speakers live, the language serves as lingua franca for speakers of the minority hill languages of the Burmese-Lolo group, Atsi (Tsai-wa), Lashi and Maru. In China there are about 100,000 speakers: Jinghpaw (*Jingpo*) is an official nationality in China. For Jinghpaw speakers in the northern Shan state no figures are available. The number of speakers in Assam is relatively small.

South-East Asia: Jinghpaw
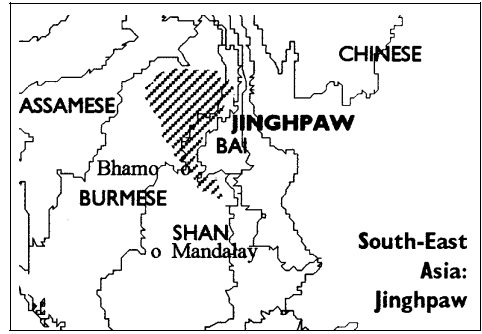
JUDEZMO

150,000 SPEAKERS

Israel and other countries

Judezmo is the SPANISH-like language spoken by descendants of the Jews who were exiled from Spain and Portugal at the end of the 15th century. Most of them settled in the western parts of what was then the Ottoman Turkish Empire – but in this century the majority have migrated again, this time mainly to Israel.

Its most obvious difference from Spanish is that Judezmo is written in the Hebrew alphabet.

The Iberian peninsula had been divided politically and religiously, and there is some evidence that Jews of the peninsula, in the centuries that preceded their exile, spoke both Spanish-related and Arabic-related dialects. Although Spanish (including MOZARABIC), Hebrew and Aramaic are the principal constituents of Judezmo, Arabic elements can also be traced in it, such as *alxá* 'Sunday' from Arabic *al-akhad*.

In 1492 most of the Jews of Spain fled the persecution of the Inquisition and found refuge in the Ottoman Empire; in 1497 the Jews of Portugal, equally threatened, followed them. They settled in Istanbul, the empire's new capital, as well as in Saloniki (where for a while they formed a majority of the population) and in many other towns of the Balkans and Anatolia.

Engaged in business and trade, gradually prospering, these Sephardic Jews became the dominant group of the very diverse Jewish communities of the empire. Their language has continued to differentiate gradually from Spanish and Portuguese, and eventually became the lingua franca of nearly all Ottoman Jews.

By 1948 there were probably over 150,000 speakers of Judaeo-Spanish or Judezmo, including large communities in the United States and Latin America and about 70,000 in Turkey. Half of these moved to Israel in that year, and the Turkish community has continued to shrink: there are now only 15,000 or fewer in Istanbul, where they are known as *Spañoles*.

Sepharad is the classical Jewish name for the Iberian peninsula: thus the members of this ancient exile community are the *Sephardim*. Their language was named *Judezmo* 'the Jewish language' from the Iberian point of view, *Ladino* 'the Romance language' from the Jewish point of view. But in practice these two names are used for two different forms of the written language.

Judezmo emerged as a widely used written language, as well as a spoken language, in the later 19th century, in the western cities of the Ottoman Empire. There was a lively press in Sarajevo and Constantinople and places between. There was considerable variation in written forms – from a Westernised Judezmo, *Lingua Franqueada*, typical of a French-educated intelligentsia, to styles that were much closer to everyday popular speech.

Ladino was a somewhat artificial development – a language used specifically for the translation of Hebrew and Aramaic liturgical texts. In Ladino texts, the Hebrew and Aramaic elements of Judezmo are consciously avoided.

There is a Judezmo newspaper in Istanbul, *Şalom*: it nowadays publishes articles in Turkish as well as Judezmo. In general Judezmo is now in steep decline both as a written and as a spoken language: in Israel, modern Hebrew tends to take its place.

KABYLE

2,500,000 SPEAKERS

Algeria, France

Kabyle is one of the BERBER LANGUAGES (see map at TAMAZIGHT) with over two million speakers (one recent estimate is 7,000,000) in north-eastern Algeria. There are half a million Kabyle-speaking migrants in French and Belgian cities. Its speakers call themselves *Qbaili*, from the Arabic *qbaila* 'tribe'.

Although some information on Kabyle and other Berber languages is recorded by medieval Arabic authors, serious study of Kabyle followed European interest in North Africa in the 18th and 19th centuries. Adolphe Hanoteau's grammar of Kabyle, aimed at French administrators, was published in 1858.

Male speakers of Kabyle are generally bilingual in Arabic, the language of education; many also speak French. Women are more likely to be monolingual. The active Kabyle resistance movement urges the adoption not of Kabyle but of the Algerian dialect of Arabic as the national language in place of the official classical Arabic. But the Algerian constitutional amendment of 1996, declaring Arabic the sole language of business, education and administration, poses a new threat to the survival of Kabyle culture.

Kabyle has numerous Arabic loanwords (sometimes difficult to recognise after phonetic reshaping) as a result of the centuries-long co-existence of the two languages: *ṭeffaḥ* 'apple' from Arabic *tuffāh*. Latin loanwords, dating from the Roman Empire, include *tifirest* 'pear tree', from Latin *pirus* with the Berber feminine affixes *t- -t*.

The first ten numerals in Kabyle are: *yiwen, sin, tlata, rewâa, xemsa, setta, set̯ța, tmanya, tesâa, âas̯ṛa*.

The Kabyle alphabet

a â b c č d ḍ e f g gw ǧ h ḥ i j k kw l m n q ɣ r ṛ s ṣ t ṭ ț, u w x y z ẓ ž

Kabyle script

Among Kabyle speakers, knowledge of the ancient Libyco-Berber alphabet, still used by the Tuareg, has long disappeared. Although nowhere an official language, Kabyle is beginning to find its way into print, in a newly standardised Latin alphabet of 37 letters.

After M. Malherbe, *Les langages de l'humanité* (Paris: Laffont, 1995) p. 1008

KADAI LANGUAGES

Kadai is the name usually applied to a language family that consists of the TAI LANGUAGES, the Kam-Sui languages (see DONG) and some other distantly related, little-known languages of Hainan and southern China. For more information on these, and a map, see LI. Paul Benedict demonstrated the relationship in 1942 in a paper entitled 'Thai, Kadai and Indonesian; a new alignment in southeastern Asia' in the *American anthropologist*. He argued that the Kadai family was linked to a wider grouping of AUSTRO-TAI LANGUAGES, and a growing number of linguists agree.

The family called *Kadai* or *Tai-Kadai* by western linguists is also recognised in China, where it is known as the *Zhuang-Dong* family, ZHUANG being the name of the largest Tai-speaking minority in China, while Dong is the name of the most important language in the Kam-Sui group.