

FOR AS FAR back as we can trace his history, Man has always spoken many different languages. If at one time he spoke a single language, from which all other languages subsequently descended, linguistic science is unlikely to uncover any hard evidence to confirm such a fact.

In the 19th century scholars made a concerted effort to reconstruct what was then assumed to be Man's original language. Major contemporary languages were exhaustively analyzed in the hope of discovering some common elements that might point to a single primeval source. Languages of isolated primitive peoples were examined in the hope of finding a revealing "fossil" tongue. But the search was in vain. Today linguists realize that a clear picture simply cannot be obtained of events that occurred perhaps a million years ago. We are faced with a complete lack of data about the beginnings of language, and any study of its subsequent evolution must be confined to the more recent historical period. Yet even in this limited aspect of the inquiry, we find ourselves confronted with a myriad of languages.

At present the languages of the world number in the thousands. To establish an exact or even an approximate number is out of the question, for many are scarcely known and it is impossible to draw a clear-cut distinction between language and dialect. In many cases, as one travels across a region the language gradually merges into a neighboring one and it becomes impossible to state for certain just what language is being spoken. But although exact numbers are unavailable, we do know that the American Indian languages number more than a thousand, the languages of Africa close to a thousand, and the single island of New Guinea contributes some 700 more. India has over 150, the Soviet Union 130, while China has several dozen, as do a number of other countries. Even in the United States more than fifty different Indian languages are spoken.

It is important, however, to view these figures in their proper perspective. A single statistic tells a great deal: of the several thousand languages of the world, fewer than 100 are spoken by over 95 percent of the earth's population. One language, Chinese, accounts for 20 percent all by itself, and if we add English, Spanish, Russian, and Hindi, the figure rises to about 45 percent. German, Japanese, Arabic, Bengali, Portuguese, French, and Italian bring the figure to 60 percent, while the next dozen most important languages raise it to 75 percent. When we realize that the last 5 percent speak thousands of different languages, it is clear that the great majority of these languages are spoken by tiny numbers of people—a few thousand in some cases, a few hundred in others, many in only a single village, some by only a few families, some even by only one or two people. As the number of speakers diminishes in each case, a fateful decision must inevitably be made: the members of the rising generation must abandon their mother tongue and adopt instead a more widely-spoken neighboring

language that will be of greater use to them. By such a decision the lesser language is literally condemned to death, its ultimate passing awaiting only the death of the last surviving speaker. This process is going on at present more rapidly than is generally realized, having been greatly accelerated by the advent of mass communications which propagate major languages in areas where hitherto only minor languages were spoken. The days of most of the Indian languages of the United States are probably numbered, and in Black Africa, with the emphasis on English, French, Swahili, and other major languages, many hundreds of lesser languages will probably also gradually disappear.

The chart on the following pages lists some 500 different languages. The number is arbitrary and could easily have been larger or smaller, but it attempts to cover the entire world without burdening the reader with a multitude of unfamiliar and unimportant names. For names that are unfamiliar, the index at the back will direct the reader to information about where each language is spoken. Sample texts and translations have been provided for nearly 200 of the languages.

The chart has been arranged to show the major language families of the world. While it is impossible to establish a single protolanguage as the ancestor of all others, it is clear that many modern languages do descend from a common ancestor, as evidenced by basic similarities in their vocabulary and grammatical structure. Those exhibiting such similarities are combined into familiar groupings such as the Romance, Germanic, and Slavic languages. These in turn are grouped into families whose various branches, though having diverged farther back in time and thus evincing less obvious similarities, are nonetheless also of common ancestry. The largest family of languages by far is the Indo-European, whose speakers now embrace approximately half the world's population. It was the discovery of the existence of the Indo-European family that led linguists to seek more tenuous links to other families and thus to prove the existence of an original common language for all of mankind. But here the effort broke down, and at this stage we must content ourselves with the existence of some twenty important language families and perhaps fifty or more lesser ones. Again the figure is imprecise, for the question of whether certain families are related to each other is still open to debate. And to complicate the picture further, a few languages such as Basque seem to defy all linguistic classification and show no similarity whatever to any other existing language.

The chart is followed by a brief description of each language family. This in turn is followed by individual descriptions of the more important languages, beginning with a sample text and translation. The final section of the book contains a country-by-country breakdown.

Indo-European Languages

THE INDO-EUROPEAN FAMILY of languages is the world's largest, embracing most of the languages of Europe, America, and much of Asia. It includes the two great classical languages of antiquity, Latin and Greek; the Germanic languages such as English, German, Dutch, and Swedish; the Romance languages such as Italian, French, Spanish, and Portuguese; the Celtic languages such as Welsh and Gaelic; the Slavic languages such as Russian, Polish, Czech, and Serbo-Croatian; the Baltic languages, Lithuanian and Latvian; the Iranian languages such as Persian and Pashto; the Indic languages such as Sanskrit and Hindi; and other miscellaneous languages such as Albanian and Armenian. In Europe only Basque, Finnish, Estonian, Hungarian, Turkish, and a few languages of Russia are not of this family; the others have apparently all descended from an original parent tongue.

Who were the original Indo-Europeans and when and where did they live? Since they left no written documents, which are, after all, the basis of history, the answers to these questions can be best obtained by attempting to reconstruct their language. If we may assume that a word that is similar in most of the Indo-European languages designates a concept that existed in the original Indo-European society and that, on the other hand, a word that varies in most Indo-European languages designates a concept not discovered until later, we may then draw certain tentative conclusions. It would appear that the Indo-Europeans lived in a cold northern region, that it was not near the water, but among forests, that they raised such domestic animals as the sheep, the dog, the cow, and the horse, that among wild animals they knew the bear and the wolf, and that among metals they probably knew only copper. Many believe that it was the use of the horse and chariot that enabled them to overrun such an enormous expanse of territory.

The general consensus is that the original Indo-European civilization developed somewhere in eastern Europe about 3000 B.C. About 2500 B.C. it broke up; the people left their homeland and migrated in many different directions. Some moved into Greece, others made their way into Italy, others moved through Central Europe until they ultimately reached the British Isles. Another division headed northward into Russia, while still another branch crossed Iran and Afghanistan and eventually reached India. Wherever they settled, the Indo-Europeans appear to have overcome the existing population and imposed their language upon them. One must conclude that they were a most remarkable people.

The possibility of so many languages having descended from a common ancestor was first suggested in 1786, though the similarity of Sanskrit and Italian was noted as early as the 16th century. By 1818 more than fifty separate languages were established as Indo-European; Albanian was added to the list in 1854 and Armenian in 1875. The total number of Indo-European speakers is about 2¼ billion people, approximately half the earth's total population.

The table below, giving the equivalents of six English words in numerous languages, will serve to illustrate the basic interrelation of the Indo-European languages, as contrasted with the languages of other families.

INDO-EUROPEAN LANGUAGES

English	month	mother	new	night	nose	three
Welsh	mis	mam	newydd	nos	trwyn	tri
Gaelic	mí	máthair	nua	oíche	srón	trí
French	mois	mère	nouveau	nuit	nez	trois
Spanish	mes	madre	nuevo	noche	nariz	tres
Portuguese	mês	mãe	novo	noite	nariz	três
Italian	mese	madre	nuovo	notte	naso	tre
Latin	mensis	mater	novus	nox	nasus	tres
German	Monat	Mutter	neu	Nacht	Nase	drei
Dutch	maand	moeder	nieuw	nacht	neus	drie
Icelandic	mánuður	móðir	nýr	nótt	nef	þrjú
Swedish	månad	moder	ny	natt	näsa	tre
Polish	miesiąc	matka	nowy	noc	nos	trzy
Czech	měsíc	matka	nový	noc	nos	tři
Rumanian	lună	mamă	nou	noapte	nas	trei
Albanian	muaj	nënë	i ri	natë	hundë	tre, tri
Greek	men	meter	neos	nux	rhís	treis
Russian	mesyats	mat'	novy	noch'	nos	tri
Lithuanian	mėnuo	motina	naujas	naktis	nosis	trys
Armenian	amis	mayr	nor	kisher	kit	yerek
Persian	māh	mādar	nau	shab	bini	se
Sanskrit	mās	matar	nava	nakt	nās	trayas

NON-INDO-EUROPEAN LANGUAGES

	(month)	(mother)	(new)	(night)	(nose)	(three)
Basque	hilabete	ama	berri	gai	sūdūr	hirur
Finnish	kuukausi	äiti	uusi	yö	nenä	kolme
Hungarian	hónap	anya	új	éjszaka	orr	három
Turkish	ay	anne	yeni	gece	burun	üç

The various branches of the Indo-European family are of sufficient importance to merit a brief discussion in their own right. We shall therefore touch upon the Germanic, Romance, Celtic, Slavic, Baltic, Iranian, and Indic languages.

GERMANIC LANGUAGES. The Germanic languages include English, German, Dutch (or Flemish), and the Scandinavian languages—Swedish, Danish, Norwegian, and Icelandic. Yiddish and Luxembourgian are offshoots of German, and Afrikaans is based on Dutch. Frisian, spoken in northern Holland, developed independently, as did Faroese, a Scandinavian language spoken in the Faroe Islands.

It is generally assumed that by the first century B.C. Germanic peoples speaking a fairly uniform language were living on both sides of the North and Baltic seas. In time there developed the so-called West, East, and North Germanic dialects. The West Germanic tribes settled in the lands between the Elbe and Oder rivers, and it is here that the German language gradually evolved. The East Germanic tribes settled east of the Oder River, but their languages have long since become extinct. In Scandinavia the North Germanic tribes spoke a language we now call Old Norse, the ancestor of the modern Scandinavian languages. In the 5th century A.D. three West Germanic tribes, the Angles, Saxons, and Jutes, crossed the North Sea into Britain, bringing with them a language that would later be known as English. And in the 9th century Old Norse was carried far westward to Iceland.

In the development of any language or language family, certain mutations inevitably occur that set it off from other languages or language families with which it shares a common origin. One such example is the sound shift that gradually occurred in the Germanic languages in the first millennium B.C. A number of Indo-European consonants acquired different values in the Germanic languages, as may be shown by a comparison between Latin, which retained the Indo-European consonants, and English, a Germanic language which did not. The Indo-European consonant *d* became *t* in the Germanic

languages (e.g., Latin *duo*, English *two*), *k* or *c* became *h* (*collis*/hill), *t* became *th* (*tonitrus*/thunder), *p* became *f* (*piscis*/fish), and *g* became *g* or *c* (*ager*/acre). This phenomenon was first described in detail in the 19th century by the German philologist Jacob Grimm (perhaps better known as the author, together with his brother Wilhelm, of *Grimm's Fairy Tales*). Known as Grimm's Law, it was a landmark in the development of modern philology.

ROMANCE LANGUAGES. The Romance languages are the modern descendants of Latin, the language of the Roman Empire. Of the modern Romance languages, Italian, French, Spanish, Portuguese, and Rumanian are each the language of an entire nation, while Catalan, Provençal, Rhaeto-Romanic, Sardinian, and Moldavian are confined to smaller areas within individual countries.

As the armies of Rome extended the boundaries of the Empire into much of the continent of Europe, Latin was introduced everywhere as the new language of administration. Spoken Latin remained fairly uniform in the beginning, though it already differed markedly from the Latin of classical literature. But as the Empire began to crumble, and Roman administrators began to disappear, the Latin of each region began to develop in its own individual way. Separated from each other by great distances, and naturally influenced by the speech of surrounding peoples, each developed its own distinctive characteristics to the point where it became a separate language.

Since we are dealing here with a slow and imperceptible process, it is impossible to say when spoken Latin ends and Romance begins. But the divergence was certainly under way by the 5th century, and by the 8th century we can detect unmistakable differences in the basic vocabulary and grammar of the various Romance dialects. The oldest text in a Romance language is the Oaths of Strasbourg, a document in Old French dating from A.D. 842.

The evolution of the Romance languages continued into modern times, each continually influenced by new geographic and ethnic factors. Each language has borrowed heavily from various non-Romance languages—French from Germanic and Celtic, Spanish and Portuguese from Arabic, and Rumanian from Slavic, Hungarian, Albanian, and Turkish. Many words exhibit remarkable uniformity throughout—e.g., bread: *pane* (Italian), *pain* (French), *pan* (Spanish), *pão* (Portuguese), *pline* (Rumanian)—while others clearly show the effects of isolation and borrowing—e.g., child: *bambino* (Italian), *enfant* (French), *niño* (Spanish), *criança* (Portuguese), *copil* (Rumanian).

12

CELTIC LANGUAGES. The Celtic languages (the initial *c* may be pronounced as either *s* or *k*) are the indigenous languages of Ireland, Scotland, and Wales. They include Gaelic (known as Irish in Ireland), Welsh, and Breton, the latter spoken in northwestern France.

The Celts were once a powerful people who dominated the area of southern Germany and the northern Alps in the first millennium B.C. About the beginning of the 5th century B.C. they began to migrate in all directions, reaching the remotest parts of Europe in a number of successive waves. The date of their arrival in the British Isles is unknown, but we are certain that when the Anglo-Saxons arrived in the 5th century A.D. they were met by a people speaking a Celtic language. In time the Celts were pushed back by the English into the west and north, leaving only Wales and the Scottish Highlands Celtic-speaking. In the 6th century one large group of Celts emigrated from Cornwall and southern Wales to Brittany, in northwestern France, where today they still speak the Celtic language known as Breton.

The Celtic languages are, sad to say, the one branch of the Indo-European family whose very survival is seriously endangered. Cornish, another Celtic language, became extinct in the 18th century, while Manx, spoken for centuries on the Isle of Man, died out only recently. While Welsh and Irish still show signs of vitality, both are continually losing ground to the all-pervasive influence of English. Attempts to encourage the teaching and speaking of the Celtic languages have met with some success, but with each passing generation the number of speakers diminishes.

SLAVIC LANGUAGES. The Slavic languages, spoken in the Soviet Union, Poland, Czechoslovakia, Yugoslavia, Bulgaria, and a small part of East Germany, form another major division of Indo-European. The modern Slavic languages number eleven: Russian, Ukrainian, Belorussian, Polish, Czech, Slovak, Bulgarian, Serbo-Croatian, Slovenian, Macedonian, and Sorbian (Lusatian).

The origin of the Slavic people is clouded in obscurity. Their homeland appears to have been the area between the Vistula and Dnieper rivers, in present-day Poland and Russia. Perhaps by the 7th century B.C. they could be identified as a distinct ethnic group. In later centuries they began a slow and steady migration in different directions, eventually dividing into the three distinct groups evident in the Slavic languages of today. The western Slavs (ancestors of the Poles, Czechs, and Slovaks) migrated toward the Elbe and Oder rivers in Germany and Poland, where they eventually adopted the Roman Catholic faith. The southern Slavs (ancestors of the Yugoslavs and Bulgarians) moved into

the Balkans, where some (the Serbs and the Bulgarians) adopted the Greek Orthodox faith, while others (the Croats and the Slovenes) adopted Roman Catholicism. The eastern Slavs made their way into Russia, where they too, in the 10th century, adopted Greek Orthodoxy.

The first Slavic language used for literary purposes was Old Church Slavonic, written in the Cyrillic alphabet devised by the Greek missionaries Cyril and Methodius in the 9th century. As individual alphabets were later developed for the various Slavic languages, the choice was made entirely by religion. The Cyrillic alphabet was adopted by the Orthodox Russians, Ukrainians, Belorussians, Bulgarians, Serbs, and Macedonians, while the Latin script was adopted by the Roman Catholic Poles, Czechs, Slovaks, Croats, Slovenes, and Sorbs. Only the use of two alphabets distinguishes Serbian from Croatian, which are otherwise, for all practical purposes, one and the same language: Serbo-Croatian.

BALTIC LANGUAGES. The Baltic languages presently number only two—Lithuanian and Latvian—several others having died out centuries ago. They are the most conservative of the Indo-European languages, retaining a number of archaic features of Indo-European that vanished from the others long before they were committed to writing. The Baltic languages share a number of common features with the Slavic languages, leading some scholars to suggest a Balto-Slavic subgroup within the Indo-European family.

The original Baltic peoples are believed to have moved into western Russia about 2000 B.C. during the great migrations of the Indo-European tribes. For centuries they occupied a large area extending from the Oka River, near present-day Moscow, westward as far as the Baltic Sea. About the 6th century A.D. the eastern Balts were forced to move westward by the more numerous Slavs, and soon afterward settled in their present homeland. By the 10th century Lithuanian and Latvian were clearly distinct languages. Today the two are not mutually intelligible, but even a cursory comparison of their vocabulary is sufficient to show their common origin.

IRANIAN LANGUAGES. The Iranian languages are dominated by Persian, a major language of antiquity and spoken today by over 35 million people. Others include Pashto, spoken in Afghanistan and Pakistan; Kurdish, the language of the Kurds; Tadjik and Ossetian, spoken in the Soviet Union; and Baluchi, spoken in Pakistan and Iran.

The Iranian languages and the Indic languages, described below,

14

15

together form what is known as the Indo-Iranian subgroup of the Indo-European family. While the other Indo-European migrations appear to have been toward the west, the Indo-Iranians headed south-east, toward the Caspian Sea and on to Iran and Afghanistan. After traversing arid deserts and great mountains, they finally reached India at a date estimated at about 1500 B.C. By perhaps 1000 B.C. the dialects of India and Iran were sufficiently different to be considered separate languages.

The 7th century B.C. witnessed the rise of the great religion of Zoroastrianism, whose sacred texts, the *Zend Avesta*, were written in an ancient Iranian language called Avestan. By the following century the might of the Persian Empire had made Persian the dominant language of the ancient world.

Since the conquests of Islam in the 7th century A.D. the Iranian languages have been written in the Arabic script. Prior to World War II the Soviet government created Cyrillic-based alphabets for Tadzhik and Ossetian, as well as for Kurdish as spoken in the Soviet Union.

INDIC LANGUAGES. The Indic languages, also known as Indo-Aryan languages, are spoken over a vast area embracing the northern two-thirds of India, as well as most of Pakistan, Bangladesh, Sri Lanka (Ceylon), and Nepal. In India the most important Indic languages are Hindi, Urdu (which closely resembles Hindi), Bengali, Punjabi, Marathi, Gujarati, Oriya, and Assamese. Urdu and Punjabi, as well as Sindhi, are also spoken in Pakistan, while Bengali is also the dominant language of Bangladesh. Another Indic language, Sinhalese, is the principal language of Sri Lanka, while Nepali is spoken in Nepal, and Kashmiri in Kashmir. Romany, the language of the Gypsies, is also of this family.

The Indic languages are the modern descendants of Sanskrit, brought to India by Indo-European settlers about 1500 B.C. The language of the sacred Hindu scriptures, it gradually gave way to the Prakrit, or Middle Indic, languages. It is out of these that the modern Indian languages eventually evolved. The date of their appearance cannot be fixed precisely but the first literary documents began to be available about 1200 A.D.

The majority of the Indic languages are written in variations of a script known as Devanagari, which appeared in India about the 7th century A.D. The Indic languages of Pakistan—Urdu and Sindhi—are written in the Arabic script.