

3

How many languages are there in the world?

M. Paul Lewis

How many languages are there? Who counts them?

Where are they spoken? Which have the most speakers?

How many languages are there? That's one of those 'it all depends' questions: how you answer it depends on what you call a language, and deciding what is and what isn't a language is not as easy as you'd think.

Suppose your favorite breakfast food is thin round cakes of grilled batter with butter and syrup. You call them 'pancakes'. Your neighbor, who likes the same meal, might call them 'griddlecakes'. If either of you travelled to a restaurant in a nearby town you might find that you have to ask for 'flapiacks'. Now imagine that chain of contacts stretching out further. After a few hundred miles, even tiny differences along the way could add up to make it hard to understand people. They might even say something like 'Wassup?' to mean 'Hello!' Where do you draw the line between a dialect and a language? Where does one language leave off and another begin?

Sometimes it's not hard to figure out. People in Iraq speak Arabic; their neighbors in Iran speak Farsi, a completely unrelated language. At other times, though, the linguistic differences are small, and the answer becomes a matter of politics and sociology.

Swedes and Norwegians can understand each other easily. But they have different histories, customs, and governments, and they see themselves as two nations, speaking two languages, not one. The same thing, more or less, goes for Malaysians and Indonesians; or Macedonians and Bulgarians. Some groups go to great lengths to distinguish themselves from their linguistic cousins across a border: Serbs and Croats understand each other's speech perfectly well, but they use two different writing systems. Other groups do just the opposite: a billion people live in China, with at least seven mutually unintelligible forms of regional speech. But they're reluctant to see themselves as separate nations, so they've clung to a unique ancient writing system that can be used anywhere in the country and lets them think of themselves as united by a single language.

Another dimension that has to be considered is socioeconomic. Generally people with more education and economic opportunity adopt certain ways of talking. Those with less education and fewer opportunities often aren't able to learn the standard or prestigious forms. Frequently, educated folk have a difficult time understanding those who aren't part of their social class. While this occurs much more often in societies where social differences are clearly marked and strongly enforced, a certain amount of it goes on in any society. What you call a 'flapiack' someone else in your society might prefer to call a 'crepe'. Sometimes the differences in speech are so marked that people have difficulty understanding one another, and one or the other group in the society won't accept the variety they don't speak as being 'their language'. Also, sometimes the less prestigious group reacts by adopting their linguistic variety as a marker of their identity, and they may insist with great pride that they speak a different language.

So it's not easy to define what is or isn't a language, and counting is a matter of definitions. How many languages there are also depends on *when* you count them. Languages, like people, are born, they change and grow and sometimes have offspring, and they eventually dwindle and die. We'll never have an exact answer to the slippery question of how many languages there are,

but among the most dedicated counters are the researchers at *Ethnologue*, a comprehensive directory of the world's languages that released its sixteenth edition in 2009. Their estimate, based largely on how well speakers can understand each other, is that a total of 6,909 languages are actively spoken or signed in the world today.

Some of those are just about extinct, with only a handful of speakers left. In fact, about a *quarter* of the world's languages have fewer than a thousand speakers and nearly 60 percent of the languages in the world have fewer than ten thousand. The median population size is 7,560. Many of those small groups are using their languages quite vigorously, but keeping a language alive is harder when there are fewer people to speak it with and when another language is being picked up and used for communication with outsiders.

At the other end of the scale is a group of very dominant languages. Over the next century they'll probably drive hundreds, or even thousands, of the smaller languages to extinction, just as supermarkets drive small shopkeepers out of business. The largest by far is Mandarin—nearly nine hundred million people in China speak it as a native tongue. Hindi, English, and Spanish each have over three hundred million native speakers. The other leading languages—all of which have between one and two hundred million native speakers—are Bengali, Portuguese, Russian, Indonesian, Arabic, Japanese, German, and French.

In addition to simply counting how many languages there are, it is interesting to observe how those languages are distributed around the world. We often think of Europe as a very multilingual place. And it does have 234 different languages. But it doesn't compare with Asia, which has 2,322—nearly a third of the world's languages—or with Africa, which has almost as many (2,110).

The number of languages in the world is constantly changing. And the *diversity* of the world's languages is amazing. Some people yearn for the days before Babel, when it's said that everyone spoke the same tongue, but that seems like a short-sighted view: *every*

language is a window on the culture (and environmental setting) in which it's spoken and a window on the human mind as well. In addition, and perhaps more importantly, every distinct language represents an identity—a group of people who see themselves as belonging together and as having a shared history. There are good reasons not only to study them but to preserve what we can of *all* of them.

About the author

M. Paul Lewis is a sociolinguistics consultant with SIL International (a nonprofit faith-based language development organization) and the editor of *Ethnologue: Languages of the World*. Born in the U.S., he is the child of a multilingual, multicultural family with parents and grandparents born in England, Wales, and Argentina. He lived for nearly twenty years in Guatemala, where he studied and did research on the K'iche' language. He holds a Ph.D. in linguistics from Georgetown University. He currently resides in Malaysia with his wife. They have four grown children.

Suggestions for further reading

In this book

Various languages of the world are described in Chapters 52 (Native American languages), 53 (Latin), 54 (Italian), 55 (languages of Spain and Portugal), 56 (Russian), 57 (Icelandic), 58 (Hebrew and Yiddish), 59 (Arabic), 60 (languages of Africa), 61 (Chinese), 62 (Japanese), 63 (languages of India), 64 (Esperanto), and 65 (other invented languages). Also relevant are Chapters 27 (language death), 28 (language rescue), 38 (languages of the U.S.), and 50 (the Museum of Languages).

Elsewhere

Comrie, Bernard, ed. *The Atlas of Languages: The Origin and Development of Languages Throughout the World* (New Burlington Books, 1996).

Crystal, David. *Language Death* (Cambridge University Press, 2000). A good overview of the dynamics of language maintenance and death and related issues.

5 What was the original language?

Barry Hilton

When did language begin, and how? What language did the earliest humans speak?

Questions like these were easier to answer back when supernatural explanations were in fashion. You could just say that language was a gift granted to humans when they first appeared in the world, like their senses and their limbs. The answer to ‘when?’ was ‘when Adam and Eve lived in the Garden of Eden’; and as for identifying that first language, Chapter 7 of this book describes some of the theories advocated during that era.

Beginning in the eighteenth and nineteenth centuries, though, a different way of thinking about the history of languages began to develop: the science of historical linguistics. As described in more detail in Chapter 6, its practitioners have identified relationships among existing languages and shown how they fit into ‘family trees’ reflecting thousands of years of changing and splitting from previously existing languages.

In many cases, these language genealogies point back to ancestral languages that no longer exist. Historical linguists have developed a method of reconstructing those long-dead languages from clues surviving in their descendants, and almost all of them believe it allows valid deductions about languages whose descendants have been separated for up to about seven thousand years.

Some think that it’s possible to look, cautiously, even further into the past. Most, though, believe that languages separated for ten thousand or more years have changed too much for the method to be reliable; and modern humans have been around five or ten times that long.

This leaves a large gap to bridge: what happened to change wordless early humans, or near-humans, into the talkers that we’ve since become? And what was their speech like? There’s been no shortage of speculation on these subjects, beginning in the late nineteenth century. Maybe, it was suggested, early people invented speech by imitating animal calls or other natural sounds and, over time, attaching meaning to them; or by attaching meaning to their own inarticulate grunts of emotion or exertion. Guesses like these are a legitimate step in scientific inquiry if they generate hypotheses that can be verified, but there didn’t seem to be any way of finding relevant concrete evidence. Critics, even friendly ones, applied mocking names like the ‘bow-wow’ theory, the ‘ding-dong’ theory, the ‘pooh-pooh’ theory and the ‘yo-he-ho’ theory. For several decades, the Origin of Language was an unfashionable field of study.

Beginning around the last quarter of the twentieth century, though, increasing amounts of brainpower—and more and more *kinds* of brainpower—have been devoted to the question, and interest is picking up.

Paleontologists studying fossils and ancient artifacts have improved our chronology of humanity’s early past, sharpening debate over *when* language is most likely to have emerged: with the first tool-using members of genus *Homo* some two million years ago? Or perhaps with the artistic flourishing that roughly coincided with the appearance of anatomically modern humans some fifty thousand years ago?

Other researchers have looked for modern analogs to the earliest human language origins: psychologists have intensively studied how infants make the transition from wordless creatures into talking children; primatologists have devised ingenious experiments to determine how much or how little human-like linguistic behavior apes can learn; and neurologists and anatomists

are making clearer to us just how extensively human language is enabled and limited by the human body and brain.

The anatomists, in particular, have suggested that language was impossible until humans had both the right kind of vocal tract to produce speech sounds and the right kind of nervous system to control them. One physical distinction between modern humans and all other animals, even chimpanzees and earlier humans, appears to be critical: a lowered larynx. Your dog can eat his food in a few quick gulps, but he can't talk. You can talk, but you can also choke from food lodged in your larynx. The human ability to make speech sounds is not a bonus provided by the body systems designed for breathing, chewing, and swallowing—it's just the opposite: the lowered larynx (and associated changes in the pharynx and mouth) is a handicap to the usual animal uses of mouth and throat, but on balance this handicap is far outweighed by the great survival value in speech. You can talk—and participate in civilization—because you *can't* wolf your food.

It's unlikely that these multidisciplinary efforts will allow us to reconstruct what words our prehistoric ancestors said, or what their speech sounded like. But some interesting late twentieth-century research suggests that we may be able to know something about the *grammar* of the earliest languages—how words came together to form sentences. Within the past few centuries several new languages of a special kind have been born. European colonists arriving in the third world communicated with their local laborers using *pidgin* languages, a kind of adult babytalk using a hodgepodge of words from different languages, strung together with a rudimentary grammar. When children are raised speaking a pidgin as their native language, they turn it into a full-fledged language called a *creole*, with a broader vocabulary and a more elaborate grammar. Now here's the fascinating part: unrelated creole languages in places as far apart as Suriname, Haiti, Hawai'i, and Papua New Guinea have radically different vocabularies, but some researchers find their grammars very similar, suggesting that the human brain may be hardwired to create particular patterns of speech. Could this be a clue to how the earliest languages worked?

About the author

Barry Hilton is the Associate Editor of this book and was a member of the review board of the radio series from which it was adapted. He is a freelance writer/editor and independent scholar living in Maine and working as a marketing specialist for a small publishing company. He is an honors graduate of Harvard College who, after graduate studies at Cornell, Yale, and George Washington Universities and the Foreign Service Institute, has travelled extensively and lived in both Europe and Asia. In a variety of U.S. government assignments he has made professional use of Vietnamese, Chinese, Japanese, French and German. He describes himself as an armchair philologist and recovering polyglot.

Suggestions for further reading

In this book

The origins and history of languages are also discussed in Chapters 6 (language relationships), 7 (the language of Adam and Eve), 8 (language change), 10 (pidgins and creoles), 51 (origins of English), and 57 (Icelandic).

Elsewhere

Kennally, Christine. *The First Word* (Viking, 2007). A comprehensive and highly readable overview of the multidisciplinary investigations that are revitalizing the study of how language began.

Bickerton, Derek. *Roots of Language* (Karoma, 1981). A readable, serious presentation of the by-no-means-mainstream theory that evidence relevant to the prehistoric origins of language can be found in creolization processes observable today.

6 Do all languages come from the same source?

Allan R. Bomhard

What does it mean to say that two languages are related?

Are all languages related?

Have you ever studied German, or Spanish, or French? If you have, you were probably grateful for cognates: foreign words that sound and look like English words with related meanings. In German, your parents are your *Mutter* and your *Vater*. In Spanish, they are your *madre* and *padre*. In French, they are your *mère* and *père*.

These resemblances not only make language learning easier, they tell us something about the history of languages. English and German share some similar vocabulary because they are both descendants of a language called Proto-West-Germanic, spoken by tribes in northern Europe well over two thousand years ago. Over time, migrations split that language into dialects, and some of the tribes moved across the North Sea into the British Isles. Fifteen centuries of separate development turned the speech of the British Isles into varieties of English, while the language of the mainlanders turned into varieties of German. So we have two languages, obviously different, but also so alike that they are clearly part of the same language family.

Language families, like families of people, can be connected into larger and larger groupings, spreading outward in territory and backward in time, as our relatives do on a genealogy chart.

The Germanic family that English, German, and several other languages belong to has a cousin, the Romance family, which includes not only French, Spanish, Portuguese, Italian, and Romanian, but also several other languages that have Latin as their common ancestor; and there are other cousins as well.

Now let's look back a step further in time. The Germanic and Romance language families share a common ancestor called Proto-Indo-European. It was spoken by tribes living some six thousand to seven thousand years ago, probably in the steppes north and east of the Black Sea. From there, the tribes spread westward across Europe, and eastward and southward into Iran and northern India. As they spread and lost contact with each other, their language changed into Germanic, Romance, Celtic, Greek, Armenian, Albanian, Baltic, Slavic, Indic languages, Iranian languages, and several other languages that are now extinct. Taken together, they make up the Indo-European language family, the most widely spoken group of languages in the world today.

As different as the Indo-European languages were from one another, they all preserved bits of ancient vocabulary and grammar. Linguists have used these bits to figure out relationships and actually reconstruct the older languages. Sir William Jones opened the way at the end of the eighteenth century through a remarkable analysis of the classical Indic language Sanskrit, showing that Sanskrit was related to languages in Europe such as Latin and Greek. And now, even though no one has seen or spoken the original Indo-European parent language for thousands of years, we have a fairly good idea of what it may have sounded like. Moreover, by cracking the code of Indo-European, we have taken a big step toward answering the question, can all languages be linked in a super family tree that begins with a single ancestral language?

To find out, linguists have increasingly studied and compared non-Indo-European languages, asking: what families do they belong to? How far back can those families be traced? Clearly, many non-Indo-European languages can be grouped together. For example, Finnish, Estonian, and Hungarian, which are surrounded

by Indo-European in the heartland of Europe, are not in the Indo-European language family. But they do group together with several other languages to form a non-Indo-European language family called Uralic. Similarly, there is a family called Turkic, which takes in Turkish, Azerbaijani, Uzbek, Uighur, Kazakh, and several other languages in Central Asia. In East Asia, the Sino-Tibetan family includes over two hundred and fifty languages, the largest of which is Mandarin Chinese. Linguists think that at least two hundred language families exist; the obvious next question is, are any of these families related to each other?

There are theorists who believe that we can lump all of the languages of the world—including even language isolates like Basque, which seems to fit nowhere—into a handful of giant families, called ‘macrofamilies’, such as Nostratic, Dene-Caucasian, Amerind, and the like.

But maybe we can’t go that far. The fact that the word for ‘dog’ in an Australian native language called Mbabaram is ‘dog’ does not mean that Mbabaram is related to English; it is just a random resemblance. The fact that Chinese calls coffee *kāfēi* does not mean that Chinese is related to English, either; the origin is a Turkish word that happens to have been borrowed by both Chinese and English.

Furthermore, we know that languages change continuously; new words join the vocabulary, while older words, including cognates, disappear; and the same has happened to grammar. After tens of thousands of years of change, can we reliably find a common ancestor? Do all languages come from the same source? The answer is: Maybe ... and maybe not. It’s too soon to know.

About the author

Allan R. Bomhard is a linguist living in Charleston, South Carolina. His main areas of interest are distant linguistic relationship and Indo-European comparative linguistics. He has published over fifty articles and seven books.

Suggestions for further reading

In this book

Chapters on the history and origins of language include 5 (earliest language), 7 (the language of Adam and Eve), 8 (language change), 10 (pidgins and creoles), 51 (origins of English), 57 (Icelandic), and 58 (Hebrew and Yiddish). Chapters focusing on language families include 52 (Native American languages), 60 (African languages), and 63 (languages of India).

Elsewhere

Baldi, Philip. *An Introduction to the Indo-European Languages* (Southern Illinois University Press, 1983). An excellent overview of the Indo-European language family. Both beginners and knowledgeable readers will find much of interest here.

Bomhard, Allan R. *Reconstructing Proto-Nostratic: Comparative Phonology, Morphology, and Vocabulary* (E. J. Brill, 2008). Though fairly technical for beginners, this book is the most comprehensive treatment of the subject that has been published to date.

Comrie, Bernard, ed. *The World’s Major Languages* (Oxford University Press, 1987). A comprehensive survey of the major languages spoken in the world today.

Fortson, Benjamin W., IV. *Indo-European Language and Culture: An Introduction* (Wiley-Blackwell, second edition 2010). An exhaustive, up-to-date, and accessible overview of Indo-European comparative-historical linguistics. Beginners should start by reading Baldi’s book listed above before tackling this work.

Pedersen, Holger, translated by John Webster Spargo. *The Discovery of Language: Linguistic Science in the Nineteenth Century* (Indiana University Press, 1931). Though dated and lacking information about more recent scholarship, this remains the most comprehensive introduction to the history of the study of languages.

Ruhlen, Merritt. *A Guide to the World’s Languages, Vol. 1: Classification* (Stanford University Press, 1991). Though this work is a comprehensive and reliable guide to the classification of nearly all known languages, some of the proposals regarding larger groupings remain controversial.

McWhorter, John. *The Story of Human Language* (The Teaching Company, 2005). An outstanding series of 36 lectures (30 minutes each, available on DVD, videotape, or audio CD) on the history and development of language.

Wade, Nicholas, ed. *The Science Times Book of Language and Linguistics* (The Lyons Press, 2000). A collection of very readable essays. The topics, organization and style correspond to the approach of the National Museum of Language.

51

Where did English come from?

John Algeo

Was English originally a German dialect? If so, how did it get to be English? How did the Vikings and the French get involved? What can dictionaries tell us about the history of English?

English did come from the same ancestor as German, but there's a lot more to the story. In the fifth century, Celts lived in the British Isles. But warfare among them got so fierce that one local king asked for help from Germanic tribes living in southern Denmark and northern Germany. He got more than he bargained for: the tribes came as allies, but they liked the island so much they decided to take it over.

Two of the main tribes in this group came from regions called Angeln and Saxony, which is why we call the language they brought to Britain 'Anglo-Saxon.' The speech of the tribes who stayed on the European Continent eventually became modern German, Dutch, and Scandinavian languages; and Anglo-Saxon, also known as 'Old English,' grew into the English we speak today.

On their new turf, these Germanic Anglo-Saxons started to talk in new ways. The tribes they drove to the fringes of Britain left them some Celtic place names. But more important, the newcomers were converted to Christianity, so a good deal of Latin crept into their language. Another influence showed up in the ninth and

tenth centuries, when Britain—which by then was called Angleland, or England—was invaded again, this time by Scandinavian cousins of the Anglo-Saxons: Viking raiders, who ruled all of England for a couple of decades. Their contact with the Anglo-Saxons was so close that they've given us some of our everyday words—like *sister*, *sky*, *law*, *take*, *window*, and the pronouns *they*, *them*, and *their*.

The greatest additions to English resulted from another invasion we all know about: '1066 and all that.' In that year, England was conquered by descendants of a different group of Vikings—the Normans, men of the North, who had settled in the tenth century along the coast of France and there learned French. The region of France they ruled still bears their name—Normandy. When they took over England, they made French the government language. So England became a trilingual country: officials used Norman French, the church used Latin, and the common people spoke a version of English we call 'Middle English.'

The common people were by far the majority, and by the late fourteenth century their English reasserted itself over French as the language of Britain. But it was a different English from the Anglo-Saxon spoken before the Conquest. Over the years it had absorbed an enormous number of French words for legal, governmental, military, and cultural matters—words like *judge*, *royal*, *soldier*, and a host of food terms like *fruit* and *beef*. And its grammar had changed dramatically, losing many of its inflectional endings.

At the end of the fifteenth century, printing was introduced in England, which helped standardize the language. And in the sixteenth century, Englishmen began to explore the globe. They encountered new things that needed to be talked about with new words. They settled in North America, the Caribbean, Africa, South Asia, Australia, and the South Seas.

As it became a global language, English influenced other languages—and was influenced by them. Most of our core vocabulary comes directly from Old English: words like *mother*, *earth*, *love*, *hate*, *cow*, *man*, and *glad*. But we have borrowed words from many other languages: Greek (*pathos*), Welsh (*penguin*), Irish

Gaelic (*galore*), Scots Gaelic (*slogan*), Icelandic (*geyser*), Swedish (*ombudsman*), Norwegian (*ski*), Danish (*skoad*), Spanish (*ranch*), Portuguese (*molasses*), Italian (*balcony*), Dutch (*boss*), German (*semester*), Yiddish (*bagel*), Arabic (*harem*), Hebrew (*sibboleth*), Persian (*bazaar*), Sanskrit (*yoga*), Hindi (*shampoo*), Roman or Gypsy (*pal*), Tamil (*curry*), Chinese (*gung-ho*), Japanese (*karracke*), Malay (*gingham*), Tahitian (*tattoo*), Tongan (*taboo*), Hawaiian (*ukulele*), Australian Dharuk (*boomerang*), Australian Guugu Yimidhiir (*kangaroo*), Bantu (*goober*), Wolof (*jigger* or *chigger*), Russian (*mammoth*), Hungarian (*paprika*), Turkish (*jackal*), Algonquian (*possum*), Dakota (*tepee*), and Navajo (*hogan*). Most of the words in a large dictionary—perhaps as many as 85 or 90 percent—either are loanwords from other languages or have been invented in English using elements borrowed from other languages.

By now the language has expanded far beyond its tribal beginnings. It's a first language in countries settled by the English. It's a second language in countries like India and the Philippines, which were part of the British Empire or under American influence. And it's a foreign language used around the globe for business, science, technology, and commerce. A Scandinavian pilot landing his plane in Greece talks with the air controller in English. It's also the main language of the worldwide Internet.

So did English come from German? No—it's closely related to German, but what began as the tongue of a small Germanic tribe in northwestern Europe morphed over time into something very different—a blend of dozens of languages that came to be spoken in virtually every country in the world.

About the author

John Algeo is Professor Emeritus at the University of Georgia. He is the author of *British or American English? A Handbook of Word and Grammar Patterns* (2006) and of *The Origins and Development of the English Language* (sixth edition 2010), and editor of *The Cambridge History of the English Language: Volume 6, English in North America* (2002) and of *Fifty Years Among the New Words* (1993). He is past president of the Dictionary Society of North America, the American Dialect Society, and