Status of Language

Jonathan Sheppard

STATUS OF
LANGUAGE

STATUS OF LANGUAGE

Jonathan Sheppard

Status of Language
by Jonathan Sheppard

Contents

Chapter 1

Official Language

An **official language** is a language given a special status in a particular country, state, or other jurisdiction. Typically the term "official language" does not refer to the language used by a people or country, but by its government (e.g. judiciary, legislature, and/or administration).

178 countries recognize an official language, 101 of them recognizing more than one. The government of Italy made Italian official only in 1999, and some nations (such as the United States) have never declared official languages at the national level. Other nations have declared non-indigenous official languages.

Many of the world's constitutions mention one or more official or national languages. Some countries use the official language designation to empower indigenous groups by giving them access to the government in their native languages. In countries that do not formally designate an official language, a *de facto* national language usually evolves. English is the most common official language, with recognized status in 51 countries. Arabic, French, and Spanish are also widely recognized.

An official language that is also an indigenous language is called *endoglossic*, one that is not indigenous is *exoglossic*. An instance is Nigeria which has three endoglossic official languages. By this the country aims to protect the indigenous languages although at the same time recognising the English

language as its lingua franca. In spatial terms, indigenous (endoglossic) languages are mostly employed in the function of official (state) languages in Eurasia, while mainly non-indigenous (exoglossic) imperial (European) languages fulfill this function in most of the "Rest of the World" (that is, in Africa, the Americas, Australia and Oceania). Ethiopia, Somalia, North Africa, Central African Republic, Greenland, and Paraguay are among the exceptions to this tendency.

History

Around 500 BC, when Darius the Great annexed Mesopotamia to the Persian Empire, he chose a form of the Aramaic language (the so-called Official Aramaic or Imperial Aramaic) as the vehicle for written communication between the different regions of the vast empire with its different peoples and languages. Aramaic script was widely employed from Egypt in the southwest to Bactria and Sogdiana in the northeast. Texts were dictated in the native dialects and written down in Aramaic, and then read out again in the native language at the places they were received.

The First Emperor of Qin standardized the written language of China after unifying the country in 221 BC. Classical Chinese would remain the standard written language for the next 2000 years. Standardization of the spoken language received less political attention, and Mandarin developed on an *ad hoc* basis from the dialects of the various imperial capitals until being officially standardized in the early twentieth century.

Statistics

According to an undated chart by the American pro-English-only organization known as U.S. English, 178 countries have an official language at the national level. Among those, English is the most common with 67 nations giving it official status. French is second with 29 countries, Arabic is third with 26 countries and Spanish is fourth with 21 countries, Portuguese is the official language of 10 countries and German is official in 6.

Some countries—like Australia, United Kingdom and the United States—have no official language recognized as such at national level. On the other extreme, Bolivia officially recognizes 37 languages, the most by any country in the world. Second to Bolivia is India with 23 official languages. South Africa is the country with the greatest number (11) of official languages that all have equal status; Bolivia gives primacy to Spanish, and India gives primacy to English and Hindi.

Political alternatives

The selection of an official language (or no official language) is often contentious. An alternative to having a single official language is "official multilingualism", where a government recognizes multiple official languages. Under this system, all government services are available in all official languages. Each citizen may choose their preferred language when conducting business. Most countries are multilingual and many are officially multilingual. Taiwan, Canada, Philippines,

Belgium, Switzerland, and the European Union are examples of official multilingualism. This has been described as controversial and, in some other areas where it has been proposed, the idea has been rejected. It has also been described as necessary for the recognition of different groups or as an advantage for the country in presenting itself to outsiders.

In specific countries/territories

Afghanistan

In accordance with *Chapter 1, Article 16* of the Constitution of Afghanistan, the Afghan government gives equal status to Pashto and Dari as official languages.

Bangladesh

After the independence of Bangladesh in 1971, the then Head of the State Sheikh Mujibur Rahman adopted the policy of 'one state one language'. The *de facto* national language, Bengali, is the sole official language of Bangladesh according to the third article of the Constitution of Bangladesh. The government of Bangladesh introduced the Bengali Language Implementation Act, 1987 to ensure the mandatory use of Bengali in all government affairs.

Bulgaria

Bulgarian is the sole official language in Bulgaria.

Belarus

Belarusian and Russian have official status in the Republic of Belarus.

Belgium

Belgium has three official languages: Dutch, French and German.

Canada

In accordance with the *Constitution Act, 1982* the (federal) Government of Canada gives equal status to English and French as official languages. The Province of New Brunswick is also officially bilingual, as is the Yukon. Nunavut has four official languages. The Northwest Territories has eleven official languages. All provinces, however, offer some necessary services in both English and French.

Canadian advocates of a single official language say it promotes national identity. In Canada, debate has focused on whether the local majority language should be made the exclusive language of public business.

Ethiopia

Ethiopia has five official languages (Amharic alone until 2020) Amharic, Oromo, Somali, Tigrinya, and Afar.

Finland

According to the Finnish constitution, Finnish and Swedish are the official languages of the republic. Citizens have the right to communicate in either language with government agencies.

Germany

German is the official language of Germany. However, its minority languages include Sorbian (Upper Sorbian and Lower Sorbian), Romani, Danish and North Frisian, which are officially recognised. Migrant languages like Turkish, Russian and Spanish are widespread, but are not officially recognised languages.

Hong Kong

According to the Basic Law of Hong Kong and the Official Languages Ordinance, both Chinese and English are the official languages of Hong Kong with equal status. The variety of Chinese is not stipulated; however, Cantonese, being the language most commonly used by the majority of Hongkongers, forms the *de facto* standard. Similarly, Traditional Chinese characters are most commonly used in Hong Kong and form the *de facto* standard for written Chinese, however there is an increasing presence of Simplified Chinese characters particularly in areas related to tourism. In government use, documents written using Traditional Chinese characters are authoritative over ones written with Simplified Chinese characters.

India

The Constitution of India (part 17) designates the official language of the Government of India as English as well as Standard Hindi written in the Devanagari script.

The Eighth Schedule of the Indian Constitution lists have 22 languages, which have been referred to as scheduled languages and given recognition, status and official encouragement. In addition, the Government of India has awarded the distinction of classical language to Tamil, Sanskrit, Kannada, Telugu, Malayalam and Odia.

Israel

On 19 July 2018, the Knesset passed a basic law under the title Israel as the Nation-State of the Jewish People, which defines Hebrew as "the State's language" and Arabic as a language with "a special status in the State" (article 4). The law further says that it should not be interpreted as compromising the status of the Arabic language in practice prior to the enactment of the basic law, namely, it preserves the status quo and changes the status of Hebrew and Arabic only nominally.

Before the enactment of the aforementioned basic law, the status of official language in Israel was determined by the 82nd paragraph of the "Palestine Order in Council" issued on 14 August 1922, for the British Mandate of Palestine, as amended in 1939:

All Ordinances, official notices and official forms of the Government and all official notices of local authorities and municipalities in areas to be prescribed by order of the High Commissioner, shall be published in English, Arabic, and Hebrew."

This law, like most other laws of the British Mandate, was adopted in the State of Israel, subject to certain amendments published by the provisional legislative branch on 19 May 1948. The amendment states that:

- "Any provision in the law requiring the use of the English language is repealed."

In most public schools, the main teaching language is Hebrew, English is taught as a second language, and most students learn a third language, usually Arabic but not necessarily. Other public schools have Arabic as their main teaching language, and they teach Hebrew as a second language and English as a third one. There are also bilingual schools which aim to teach in both Hebrew and Arabic equally.

Some languages other than Hebrew and Arabic, such as English, Russian, Amharic, Yiddish and Ladino enjoy a somewhat special status, but are not considered to be official languages. For instance, at least 5% of the broadcasting time of privately owned TV-channels must be translated into Russian (a similar privilege is granted to Arabic), warnings must be translated to several languages, signs are mostly trilingual (Hebrew, Arabic and English), and the government supports Yiddish and Ladino culture (alongside Hebrew culture and Arabic culture).

Latvia

The Constitution of Latvia (or Satversme) designated Latvian as the state language. In 2012 there was initiative to hold a referendum on constitutional amendments, elevating Russian as a state language. Kristīne Jarinovska in her analysis describes the proposal in the following way:

It proposed several constitutional amendments for introducing Russian as Latvia's second official language—i.e., amendments to the Satversme's Articles 4 (on Latvian as the state language), 18 (on the solemn promise of a member of Parliament to strengthen the Latvian language), 21 (on Latvian as the working language of the Parliament), 101 (on Latvian as the working language of local governments), and 104 (on the right to receive a reply to a petition in Latvian). Obviously, the proposed amendments would have influenced other constitutional norms as well. Moreover, since Article 4 of the Satversme alike norms of independence, democracy, sovereignty, territorial wholeness, and basic principles of elections that form the core of the Satversme (according to Article 77 of the Satversme), the initiative, in fact, proposed discontinuing an existing state and establishing a new one that is no longer a nation-state wherein Latvians exercise their rights to self-determination, enjoying and maintaining their cultural uniqueness

Netherlands

Dutch is the official language of the Netherlands. In the province of Friesland, Frisian is the official second language.

Dutch is also the official language of the Caribbean Netherlands (the islands Bonaire, Saba and Sint Eustatius), but is not the main spoken language. Papiamento is most often spoken on Bonaire and English on Saba and Sint Eustatius, and these languages can be used in official documents.

Low Saxon and Limburgish are languages acknowledged by the European Charter, and are spoken in specific regions of the Netherlands.

New Zealand

New Zealand has three official languages. English is the *de facto* and principal official language, accepted in all situations. The Māori language and New Zealand Sign Language both have limited *de jure* official status under the Māori Language Act 1987 and New Zealand Sign Language Act 2006

Nigeria

The official language of Nigeria is English, which was chosen to facilitate the cultural and linguistic unity of the country. British colonisation ended in 1960.

Pakistan

Urdu is the national language of Pakistan. Urdu and English both are official languages in Pakistan. Pakistan has more than 60 other languages.

Russia

Russian is the official language of the Russian Federation and in all federal subjects, however many minority languages have official status in the areas where they are indigenous. One type of federal subject in Russia, republics, are allowed to adopt additional official languages alongside Russian in their own constitutions. Republics are often based around particular native ethnic groups, and are often areas where ethnic Russians and native Russian-language speakers are a minority.

South Africa

South Africa has eleven official languages that are mostly indigenous. Due to limited funding, however, the government rarely produces documents in most of the languages. Accusations of mismanagement and corruption have been leveled against the Pan South African Language Board, established to promote multilingualism, to develop the 11 official languages, and to protect language rights in the country.

Switzerland

The four national languages of Switzerland are German, French, Italian and Romansh. At the federal level German, French and Italian are official languages, the official languages of individual cantons depend on the languages spoken in them.

Taiwan

Mandarin is the most common language used in government. After World War II the mainland Chinese-run government made Mandarin the official language, and it was used in the schools and in government. Under the *National languages development act*, political participation can be conducted in any national language, which is defined as a "natural language used by an original people group of Taiwan", which also includes Formosan languages, Hakka and Taiwanese Hokkien. According to Taiwan's Legislative Yuan, amendments were made to the Hakka Basic Act to make Hakka an official language of Taiwan.

Ukraine

The official language of Ukraine is Ukrainian. The status of Russian as a regional language caused significant political controversy.

United Kingdom

The de facto official language of the United Kingdom is English. In Wales, the Welsh language, spoken by approximately 20% of the population, has limited de jure official status.

United States

English is the de facto national language of the United States. While there is no official language at the federal level, 32 of the 50 U.S. states and all five inhabited U.S. territories have

designated English as one, or the only, official language, while courts have found that residents in the 50 states do not have a right to government services in their preferred language. Public debate in the last few decades has focused on whether Spanish should be recognized by the government, or whether all business should be done in English.

California allows people to take their driving test in the following 32 languages: Amharic, Arabic, Armenian, Chinese, Croatian, English, French, German, Greek, Hebrew, Hindi, Hmong, Hungarian, Indonesian, Italian, Japanese, Khmer, Korean, Laotian, Persian, Polish, Portuguese, Punjabi, Romanian, Russian, Samoan, Spanish, Tagalog/Filipino, Thai, Tongan, Turkish, and Vietnamese.

New York state provides voter-registration forms in the following five languages: Bengali, Chinese, English, Korean and Spanish. The same languages are also on ballot papers in certain parts of the state (namely, New York City).

The pro-English-only website U.S. English sees a multilingual government as one in which its "services actually encourage the growth of linguistic enclaves...[and] contributes to racial and ethnic conflicts". Opponents of an official language policy in the United States argue that it would hamper "the government's ability to reach out, communicate, and warn people in the event of a natural or man-made disaster such as a hurricane, pandemic, or...another terrorist attack". Professor of politics Alan Patten argues that disengagement (officially ignoring the issue) works well in religious issues but that it is not possible with language issues because it must offer public services in some language. Even if it makes a conscious effort

not to establish an official language, a *de facto* official language, or the "national language", will nevertheless emerge. Indeed, two-thirds of Americans believe that English is the United States' official language.

Yugoslavia

Sometimes an official language definition can be motivated more by national identity than by linguistic concerns. When Yugoslavia dissolved in 1991, the country had four official languages—Serbo-Croatian, Slovene, Albanian and Macedonian. Serbo-Croatian was used as a *lingua franca* for mutual understanding and was also the language of the military.

When Croatia declared independence (1991) it defined its official language as Croatian, and Serbia likewise defined its official language as Serbian. Bosnia-Herzegovina defined three official languages: Bosnian, Croatian, and Serbian. From the linguistic point of view, the different names refer to national varieties of the same language, which is known under the appellation of Serbo-Croatian. It is said by some that the Bosnian government chose to define three languages to reinforce ethnic differences and keep the country divided. The language used in Montenegro, traditionally considered a dialect of Serbian, became standardized as the Montenegrin language upon Montenegro's declaration (2006) of independence.

Languages of the United States

Although the United States does not have an official language, the most commonly used language is English (specifically, American English), which is the de facto national language, and the only one spoken at home by approximately 78% of the U.S. population. Many other languages are also spoken at home, especially Spanish (13.4% of the population), according to the American Community Survey (ACS) of the U.S. Census Bureau; these include indigenous languages and languages brought to the U.S. by people from Europe, Africa, and Asia. However, the majority of speakers of these languages are bilingual and also speak English. Although 21.6% of U.S. residents report that they speak a language other than English at home, only 8.4% speak English less than "very well." Several other languages, notably creoles and sign languages, have developed in the United States. Approximately 430 languages are spoken or signed by the population, of which 176 are indigenous to the area. Fifty-two languages formerly spoken in the country's territory are now extinct.

Most common languages

Based on annual data from the American Community Survey, the United States Census Bureau regularly publishes information on the most common languages spoken at home. It also reports the English speaking ability of people who speak a language other than English at home. In 2017, the U.S. Census

Bureau published information on the number of speakers of over 350 languages as surveyed by the ACS from 2009 to 2013, but it does not regularly tabulate and report data for that many languages.

According to the ACS in 2017, the most common languages spoken at home by people aged five years of age or older are as follows:

- English only – 239 million
- Spanish – 41 million
- Chinese (including Mandarin, Cantonese, Hokkien and all other varieties) – 3.5 million
- Tagalog (including Filipino) – 1.7 million
- Vietnamese – 1.5 million
- Arabic – 1.2 million
- French – 1.2 million
- Korean – 1.1 million
- Russian – 0.94 million
- German – 0.92 million
- Haitian Creole – 0.87 million
- Hindi – 0.86 million
- Portuguese – 0.79 million
- Italian – 0.58 million
- Polish – 0.52 million
- Yiddish – 0.51 million
- Japanese – 0.46 million
- Persian (including Farsi, Dari and Tajik) – 0.42 million
- Gujarati – 0.41 million
- Telugu – 0.37 million
- Bengali – 0.32 million

- Tai–Kadai (including Thai and Lao) – 0.31 million
- Urdu – 0.3 million
- Greek – 0.27 million
- Punjabi – 0.29 million
- Tamil – 0.27 million
- Armenian – 0.24 million
- Serbo-Croatian (including Bosnian, Croatian, Montenegrin, and Serbian) – 0.24 million
- Hebrew – 0.23 million
- Hmong – 0.22 million
- Bantu languages (including Swahili) – 0.22 million
- Khmer – 0.20 million
- Navajo – 0.16 million
- other Indo-European languages – 578,492
- other Afro-Asiatic languages – 521,932
- other Niger–Congo languages – 515,629
- other West Germanic languages – 487,675
- other Austronesian languages – 467,718
- other Indic languages – 409,631
- other languages of Asia – 384,154
- other Slavic languages – 338,644
- other Dravidian languages – 241,678
- other languages of North America – 195,550
- other and unspecified languages – 258,257

The ACS is not a full census but an annual sample-based survey conducted by the U.S. Census Bureau. The language statistics are based on responses to a three-part question asked about all members of a target U.S. household who are at least five years old. The first part asks if they "speak a language other than English at home." If so, the head of the household or main respondent is asked to report which

language each member speaks in the home, and how well each individual speaks English. It does not ask how well individuals speak any other language of the household. Thus, some respondents might have only a limited speaking ability of that language. In addition, it is difficult to make historical comparisons of the numbers of speakers because language questions used by the U.S. Census changed numerous times before 1980.

The ACS does not tabulate the number of people who report the use of American Sign Language at home, so such data must come from other sources. While modern estimates indicate that American Sign Language was signed by as many as 500,000 Americans in 1972 (the last official survey of sign language), estimates as recently as 2011 were closer to 100,000. Various cultural factors, such as the passage of the Americans with Disabilities Act, have resulted in far greater educational opportunities for hearing-impaired children, which could double or triple the number of current users of American Sign Language.

English is the most common language spoken in the United States with approximately 239 million speakers. Spanish is spoken by approximately 35 million people. The United States has the world's fifth largest Spanish-speaking population, outnumbered only by Mexico, Colombia, Spain, and Argentina; other estimates put the United States at over 50 million, second only to Mexico. Throughout the Southwestern United States and Puerto Rico, long-established Spanish-speaking communities coexist with large numbers of more recent Hispanophone immigrants. Although many new Latin American immigrants are less than fluent in English, nearly all second-

generation Hispanic Americans speak English fluently, while only about half still speak Spanish.

According to the 2000 U.S. Census, people of German ancestry made up the largest single ethnic group in the United States, but German language ranked fifth. Italian, Polish, and French are still widely spoken among populations descending from immigrants from those countries in the early 20th century, but the use of these languages is dwindling as the older generations die. Russian is also spoken by immigrant populations.

Tagalog and Vietnamese have over one million speakers each in the United States, almost entirely within recent immigrant populations. Both languages, along with the varieties of Chinese (mostly Cantonese, Taishanese, and Standard Mandarin), Japanese, and Korean, are now used in elections in Alaska, California, Hawaii, Illinois, New York, Texas, and Washington.

Native American languages are spoken in smaller pockets of the country, but these populations are decreasing, and the languages are almost never widely used outside of reservations. Besides English, Spanish, French, German, Navajo and other Native American languages, all other languages are usually learned from immigrant ancestors that came after the time of independence or learned through some form of education.

American Sign Language is the most common sign language in the United States although there are unrelated sign languages which have been developed in the States and territories—

mostly in the Pacific. No concrete numbers exist for signers but something upwards of 250,000 is common.

Official languages

The United States has never had an official language at the federal level, but English is typically used at the federal level and in states without an official language. Outside of Puerto Rico, English is the primary language used for legislation, regulations, executive orders, treaties, federal court rulings, and all other official pronouncements. Nonetheless, laws require documents such as ballots to be printed in multiple languages when there are large numbers of non-English speakers in an area.

32 states, in some cases as part of what has been called the English-only movement, have adopted legislation granting official status to English. Hawaiian, although having few native speakers, is an official language along with English of the state of Hawaii. Alaska has made some 20 native languages official, along with English; for example, Alaska provides voting information in English, Iñupiaq, Central Yup'ik, Gwich'in, Siberian Yupik, Koyukon, and Tagalog. On July 1, 2019, a law went into effect making Lakota, Dakota, and Nakota the official indigenous languages of South Dakota. Typically only "English" is specified, not a particular dialect, but from 1923 to 1969, the state of Illinois recognized its official language as "American".

French is a *de facto*, but unofficial, language in Maine and Louisiana, while New Mexico law grants Spanish a special

status. The government of Louisiana offers services and most documents in both English and French, and New Mexico does so in English and Spanish.

English is at least one of the official languages all five permanently inhabited U.S. territories. In Puerto Rico both English and Spanish are official, although Spanish has been declared the principal official language. The school system and the government operate almost entirely in Spanish, but federal law requires the United States District Court for the District of Puerto Rico to use English, like the rest of the federal court system. Guam recognizes English and Chamorro. In the U.S. Virgin Islands, English is the only official language. In American Samoa, both English and Samoan are officially recognized; English is common but Samoan is also seen in some official communications. In the Northern Mariana Islands, English, Chamorro, and Carolinian are official.

In New Mexico, although the state constitution does not specify an official language, laws are published in English and Spanish, and government materials and services are legally required (by Act) to be made accessible to speakers of both languages as well as Navajo and various Pueblo languages. New Mexico also has its own dialect of Spanish, which differs from Spanish spoken in the rest of Latin America.

Algonquian, Cherokee, and Sioux are among many other Native American languages which are official or co-official on many U.S. Indian reservations and Pueblos. In Oklahoma before statehood in 1907, territory officials debated whether or not to have Cherokee, Choctaw, and Muscogee languages as co-official, but the idea never gained ground. Cherokee is officially

recognized by the Cherokee Nation within the Cherokee tribal jurisdiction area in eastern Oklahoma.

After New Amsterdam (formerly a Dutch colony) was transferred to English administration (becoming the Province of New York) in the late 17th century, English supplanted Dutch as the official language. However, "Dutch remained the primary language for many civil and ecclesiastical functions and most private affairs for the next century." The Jersey Dutch dialect is now extinct.

California has agreed to allow the publication of state documents in other languages to represent minority groups and immigrant communities. Languages such as Spanish, Chinese, Korean, Tagalog, Persian, Russian, Vietnamese, and Thai appear in official state documents, and the Department of Motor Vehicles publishes in nine languages.

The issue of multilingualism also applies in the states of Arizona and Texas. While the constitution of Texas has no official language policy, Arizona passed a proposition in 2006 declaring English as the official language. Nonetheless, Arizona law requires the distribution of voting ballots in Spanish, as well as indigenous languages such as Navajo, O'odham and Hopi, in counties where they are spoken.

A popular urban legend called the Muhlenberg legend claims that German was almost made an official language of the United States but lost by one vote. In reality, it was a request by a group of German immigrants to have an official translation of laws into German. House speaker Frederick Muhlenberg has since become associated with the legend.

Education

Billingual education in the United States is an area of political controversy. The medium of instruction at almost all schools at all levels is English, outside of language classes and Puerto Rico, where Spanish is standard. English is also the language of instruction in American Samoa despite most students speaking the Samoan language natively. There are hundreds of language immersion schools across the United States teaching in a variety of languages, including Spanish, Hawaiian, Chamorro language#Revitalization efforts, and Mandarin Chinese

Indigenous languages

Native American languages

Native American languages predate European settlement of the New World. In a few parts of the U.S. (mostly on Indian reservations), they continue to be spoken fluently. Most of these languages are endangered, although there are efforts to revive them. Normally the fewer the speakers of a language the greater the degree of endangerment, but there are many small Native American language communities in the Southwest (Arizona and New Mexico) which continue to thrive despite their small size. In 1929, speaking of indigenous Native American languages, linguist Edward Sapir observed:

Few people realize that within the confines of the United States there is spoken today a far greater variety of languages ... than

in the whole of Europe. We may go further. We may say, quite literally and safely, that in the state of California alone there are greater and more numerous linguistic extremes than can be illustrated in all the length and breadth of Europe.

Navajo

According to the 2000 Census and other language surveys, the largest Native American language-speaking community by far is the Navajo. Navajo is an Athabaskan language of the Na-Dené family, with 178,000 speakers, primarily in the states of Arizona, New Mexico, and Utah. Altogether, Navajo speakers make up more than 50% of all Native American language speakers in the United States. Western Apache, with 12,500 speakers, also mostly in Arizona, is closely related to Navajo but not mutually intelligible with it. Navajo and other Athabaskan languages in the Southwest are relative outliers; most other Athabascan languages are spoken in the Pacific Northwest and Alaska. Navajo has struggled to keep a healthy speaker base, although this problem has been alleviated to some extent by extensive education programs in the Navajo Nation, including a Navajo language immersion school in Fort Defiance, Arizona.

Cherokee

Cherokee is the Iroquoian language spoken by the Cherokee people, and the official language of the Cherokee Nation. Significant numbers of Cherokee speakers of all ages still populate the Qualla Boundary in Cherokee, North Carolina and

several counties within the Cherokee Nation of Oklahoma, significantly Cherokee, Sequoyah, Mayes, Adair, and Delaware. Increasing numbers of Cherokee youth are renewing interest in the traditions, history, and language of their ancestors. Cherokee-speaking communities stand at the forefront of language preservation, and at local schools, all lessons are taught in Cherokee and thus it serves as the medium of instruction from pre-school on up. Also, church services and traditional ceremonial "stomp" dances are held in the language in Oklahoma and on the Qualla Boundary in North Carolina.

Cherokee is one of the few, or perhaps the only, Native American language with an increasing population of speakers, and along with Navajo it is the only indigenous American language with more than 50,000 speakers, a figure most likely achieved through the tribe's 10-year long language preservation plan involving growing new speakers through immersion schools for children, developing new words for modern phrases, teaching the language to non-Cherokee in schools and universities, fostering the language among young adults so their children can use that language at home, developing iPhone and iPad apps for language education, the development of Cherokee language radio stations including Cherokee Voices, Cherokee Sounds, and promoting the writing system through public signage, products like the Apple iPhone, internet use through Google including Gmail, and others so the language remains relevant in the 21st century.

Other Native American languages

Dakota is a Siouan language with 18,000 speakers in the US alone (22,000 including speakers in Canada), not counting 6,000 speakers of the closely related Lakota. Most speakers live in the states of North Dakota and South Dakota. Other Siouan languages include the closely related Winnebago, and the more distant Crow, among others.

Central Alaskan Yup'ik is an Eskimo-Aleut language with 16,000 speakers, most of whom live in Alaska. The term "Yupik" is applied to its relatives, which are not necessarily mutually intelligible with Central Alaskan, including Naukan and Central Siberian, among others.

The O'odham language, spoken by the Pima and the Tohono O'odham, is a Uto-Aztecan language with more than 12,000 speakers, most of whom live in central and southern Arizona and northern Sonora. Other Uto-Aztecan languages include Hopi, Shoshone, and the Pai-Ute languages.

Choctaw has 11,000 speakers. Choctaw is part of the Muskogean family, like Seminole and Alabama.

The Algonquian language family includes languages like Chippewa/Ojibwe, Cheyenne, and Cree.

Keres has 11,000 speakers in New Mexico and is a language isolate. The Keres pueblo people are the largest of the Pueblo nations. The Keres pueblo of Acoma is the oldest continually inhabited community in the United States. Zuni, another

isolate, has around 10,000 speakers, most of whom reside within the Zuni pueblo.

Because of immigration from Mexico, there are Mexican native American languages speakers in the US. There are thousands of Nahuatl, Mixtec, Zapotec and Trique speakers in communities stablished mainly in the southern states.

Although the languages of the Americas have a history stretching back about 17,000 to 12,000 years, current knowledge of them is limited. There are doubtlessly a number of undocumented languages that were once spoken in the United States that are missing from historical record.

List of Native American languages

- Below is an estimate of Native American languages "spoken at home" in the United States (American Community Survey 2006–2008). This is not an exhaustive list of Native American languages in the US. Because the distinction between dialect and language isn't always clear, multiple dialects of varying mutual intelligibility may be classified as a single language, while a group of effectively identical dialects may be classified separately for historical or cultural reasons. Languages included here may be classified as "extinct" (having no living native speakers), but many extinct or moribund Native American languages are the subjects of ongoing language revitalization efforts; other extinct

languages undergoing revitalization might not be listed here.

Native American sign languages

A sign-language trade pidgin, known as Plains Indian Sign Language, Plains Standard or Plains Sign Talk, arose among the Native Americans of the plains. Each signing nation had a separate signed version of their oral language, that was used by the hearing, and these were not mutually intelligible. Plains Standard was used to communicate between these nations. It seems to have started in Texas and then spread north, through the Great Plains, as far as British Columbia. There are still a few users today, especially among the Crow, Cheyenne, and Arapaho. Unlike other sign languages developed by hearing people, it shares the spatial grammar of deaf sign languages. Through intergenerational transmission, Plains Sign Talk became a working language still in use today in some Deaf First Nations or Native American communities.

As Plains Sign Talk was so widespread and was a spectrum of dialects and accents, it probably hosted several languages under its umbrella. One is potentially Navajo Sign Language which is in use by a sole Navajo clan.

Additionally, Plateau Sign Language existed alongside Plains Sign Talk as either a trade pidgin or another language around the Columbia Plateau and surrounding regions.

Austronesian languages

Hawaiian

Hawaiian is an official state language of Hawaii as prescribed in the Constitution of Hawaii. Hawaiian has 1,000 native speakers. Formerly considered critically endangered, Hawaiian is showing signs of language renaissance. The recent trend is based on new Hawaiian language immersion programs of the Hawaii State Department of Education and the University of Hawaii, as well as efforts by the Hawaii State Legislature and county governments to preserve Hawaiian place names. In 1993, about 8,000 could speak and understand it; today estimates range up to 27,000. Hawaiian is related to the Māori language spoken by around 150,000 New Zealanders and Cook Islanders as well as the Tahitian language which is spoken by another 120,000 people of Tahiti.

Samoan

Samoan is an official territorial language of American Samoa. Samoans make up 90% of the population, and most people are bilingual.

Chamorro

Chamorro is co-official in the Mariana Islands, both in the territory of Guam and in the Commonwealth of the Northern

Mariana Islands. In Guam, the indigenous Chamorro people make up about 60% of the population.

Carolinian

Carolinian is also co-official in the Northern Marianas, where only 14% of people speak English at home.

African, Asian and European languages

Some of the first European languages to be spoken in the U.S. were English, Dutch, French, Spanish, and Swedish.

From the mid-19th century on, the nation had large numbers of immigrants who spoke little or no English, and throughout the country state laws, constitutions, and legislative proceedings appeared in the languages of politically important immigrant groups. There have been bilingual schools and local newspapers in such languages as German, Ukrainian, Hungarian, Irish, Italian, Norwegian, Greek, Polish, Swedish, Romanian, Czech, Japanese, Yiddish, Hebrew, Lithuanian, Welsh, Cantonese, Bulgarian, Dutch, Portuguese and others, despite opposing English-only laws that, for example, illegalized church services, telephone conversations, and even conversations in the street or on railway platforms in any language other than English, until the first of these laws was ruled unconstitutional in 1923 (*Meyer v. Nebraska*).

Currently, Asian languages account for the majority of languages spoken in immigrant communities: Korean, the varieties of Chinese, Arabic, Vietnamese, Tagalog, Persian. South Asian languages like Hindi, Urdu, Tamil, Kannada, Gujarati, Marathi, Bengali, Telugu, Punjabi, Malayalam and Sylheti language.

Typically, immigrant languages tend to be lost through assimilation within two or three generations, though there are some groups such as the Louisiana Creoles (French), Pennsylvania Dutch (German) in a state where large numbers of people were heard to speak it before the 1950s, and the original settlers of the Southwest (Spanish) who have maintained their languages for centuries.

English

English was inherited from British colonization, and it is spoken by the majority of the population. English has become increasingly common; when the US was founded, 40% of Americans spoke English as a first language. In 2002, 87% of Americans spoke English as their first language. It serves as the de facto national language, the language in which government business is carried out. According to the U.S. Census Bureau, 80% spoke only English at home and all but approximately 13,600,000 U.S. residents age 5 and over speak English "well" or "very well".

American English is different from British English in terms of spelling (one example being the dropped "u" in words such as color/colour), grammar, vocabulary, pronunciation, and slang usage. The differences are not usually a barrier to effective

communication between an American English and a British English speaker.

Some states, like California, have amended their constitutions to make English the only official language, but in practice, this only means that official government documents must *at least* be in English, and *does not* mean that they should be exclusively available only in English. For example, the standard California Class C driver's license examination is available in 32 different languages.

Spanish

- Spanish was also inherited from colonization and is sanctioned as official in the commonwealth of Puerto Rico. Spanish is also taught in various regions as a second language, especially in areas with large Hispanic populations such as the Southwestern United States along the border with Mexico, as well as Florida, parts of California, the District of Columbia, Illinois, New Jersey, and New York. In Hispanic communities across the country, bilingual signs in both Spanish and English may be quite common. Furthermore, numerous neighborhoods exist (such as Washington Heights in New York City or Little Havana in Miami) in which entire city blocks will have only Spanish language signs and Spanish-speaking people.

In addition to Spanish-speaking Hispanic populations, younger generations of non-Hispanics in the United States seem to be learning Spanish in larger numbers due to the growing

Hispanic population and increasing popularity of Latin American movies and music performed in the Spanish language. A 2009 American Community Survey (ACS) conducted by the United States Census Bureau, showed that Spanish was spoken at home by over 35 million people aged 5 or older, making the United States the world's fifth-largest Spanish-speaking community, outnumbered only by Mexico, Colombia, Spain, and Argentina. Since then, the number of persons reported on the ACS to speak Spanish at home has increased (see table).

New Mexican Spanish

In northern New Mexico and southern Colorado, Spanish speakers have been isolated for centuries in the southern Rockies, and developed a distinct dialect of Spanish spoken nowhere else: New Mexican Spanish. The dialect features a mix of Castilian, Galician and, more recently, Mexican Spanish, as well as Pueblo loan words. New Mexican Spanish also contains a large proportion of English loan words, particularly for technological words (e.g. *bos*, *troca*, and *telefón*).

Speakers of New Mexican Spanish are mainly descendants of Spanish colonists who arrived in New Mexico in the sixteenth through the eighteenth centuries. During this time, contact with the rest of Spanish America was limited, and New Mexican Spanish developed on its own course. In the meantime, Spanish colonists coexisted with and intermarried with Puebloan peoples and Navajos. After the Mexican–American War, New Mexico and all its inhabitants came under the

governance of the English-speaking United States, and for the next hundred years, English-speakers increased in number.

Puerto Rican Spanish

Puerto Rican Spanish is the main language and dialect of the people of Puerto Rico, as well as many people descended from Puerto Ricans elsewhere throughout the United States.

Spanglish

Spanglish is a code-switching variant of Spanish and English and is spoken in areas with large bilingual populations of Spanish and English speakers, such as along the Mexico–United States border (California, Arizona, New Mexico, and Texas), Florida, and New York City.

Chinese

The population of Chinese speakers was increasing rapidly in the 20th century because the number of Chinese immigrants increased at a rate more than 50% since 1940. 2.8 million Americans speak some variety of Chinese, which combined are counted by the federal census as the third most-spoken language in the country. Until the late 20th century, Yue dialects including Taishanese and Cantonese were the most common among immigrants or the descendants of immigrants, especially in California. Since the opening of the People's Republic of China, Mandarin, the official language in PRC and Republic of China (Taiwan) has become increasingly prevalent.

Many Americans of all ethnic backgrounds are also learning Mandarin and to a far lesser extent Cantonese.

In New York City in 2002, Mandarin was spoken as a native language among only 10% of Chinese speakers but was predicted to replace Cantonese as the lingua franca among Chinese speakers.

Tagalog

Tagalog speakers were already present in the United States as early as the late sixteenth century as sailors contracted by the Spanish colonial government. In the eighteenth century, they established settlements in Louisiana, such as Saint Malo. After the American annexation of the Philippines, the number of Tagalog speakers steadily increased, as Filipinos began to migrate to the U.S. as students or contract laborers. Their numbers, however, decreased upon Philippine independence, as some Filipinos were repatriated.

Today, Tagalog, together with its standardized form Filipino, is spoken by over a million and a half Filipino Americans and is promoted by Filipino American civic organizations and Philippine consulates. As Filipinos are the second largest Asian ethnic group in the United States, Tagalog is the second most spoken Asian language in the country. Taglish, a form of code-switching between Tagalog and English, is also spoken by a number of Filipino Americans.

Tagalog is also taught at some universities where a significant number of Filipinos exist. As it is the national and most spoken language of the Philippines, most Filipinos in the

United States are proficient in Tagalog in addition to their local regional language.

Vietnamese

According to the 2010 Census, there are over 1.5 million Americans who identify themselves as Vietnamese in origin, ranking fourth among the Asian American groups and forming the largest Overseas Vietnamese population.

Orange County, California, is home to the largest concentration of ethnic Vietnamese outside Vietnam, especially in its Little Saigon area. Other significant Vietnamese communities are found in the metropolitan areas of San Jose, Houston, Seattle, Northern Virginia, and New Orleans. Similarly to other overseas Vietnamese communities in Western countries (except France), the Vietnamese population in the United States was established following the Fall of Saigon in 1975 and communist takeover of South Vietnam following the Vietnam War.

Korean

In 2011 over 1.1 million Americans spoke Korean at home. This number increased greatly at the end of the 20th century, increasing 327% from the 300,000 speakers in 1980. The greatest concentration of these speakers was in the Los Angeles, New York, and Washington D.C. metro areas. Speakers of Korean are found in the Koreatowns.

French

French, the fourth-most-common language (when all varieties of French are combined and separate yet related languages such as Haitian Creole are counted as French), is spoken mainly by the Louisiana Creole, native French, Cajun, Haitian, and French-Canadian populations. It is widely spoken in Maine, New Hampshire, Vermont, and in Louisiana, with notable Francophone enclaves in St. Clair County, Michigan, many rural areas of the Upper Peninsula of Michigan and the northern San Francisco Bay area. Because of its legacy in Louisiana, that state is served by the Council for the Development of French in Louisiana (CODOFIL), the only state agency in the United States whose mission is to serve a linguistic population, which registered Louisiana as an observer in the economic and cultural *Organisation internationale de la Francophonie* in 2018.

Three varieties of French developed within what is now the United States in colonial times: Louisiana French, Missouri French, and New England French (essentially a variant of Canadian French). French is the second language in the states of Louisiana and Maine. The largest French-speaking communities in the United States reside in Northeast Maine; Hollywood and Miami, Florida; New York City; certain areas of rural Louisiana; and small minorities in Vermont and New Hampshire. Many of the New England communities are connected to the dialect found across the border in Quebec or New Brunswick. More than 13 million Americans possess primary French heritage, but only 2 million speak French or a French Creole language at home.

German

German was a widely spoken language in some of the colonies, especially Pennsylvania, where a number of German-speaking Protestants and other religious minorities settled to escape persecution in Europe. Another wave of settlement occurred when Germans fleeing the failure of 19th Century German revolutions emigrated to the United States. A large number of these German immigrants settled in the urban areas, with neighborhoods in many cities being German-speaking and numerous local German language newspapers and periodicals established. German farmers also took up farming around the country, including the Texas Hill Country, at this time. The language was widely spoken until the United States entered World War I.

In the early twentieth century, German was the most widely studied foreign language in the United States, and prior to World War I, more than 6% of American schoolchildren received their primary education exclusively in German, though some of these Germans came from areas outside of Germany proper. Currently, more than 49 million Americans claim German ancestry, the largest self-described ethnic group in the U.S., but less than 4% of them speak a language other than English at home, according to the 2005 American Community Survey. The Amish speak a dialect of German known as Pennsylvania German. One reason for this decline of German language was the perception during both World Wars that speaking the language of the enemy was unpatriotic; foreign language instruction was banned in places during the First World War. Unlike earlier waves, they were more concentrated in cities and integrated quickly. Another reason

for the decline in German was the lack of first-generation immigrants, as immigration to the United States from Western Europe slowed following the World Wars.

There is a myth (known as the Muhlenberg Vote) that German was to be the official language of the U.S., but this is inaccurate and based on a failed early attempt to have government documents translated into German. The myth also extends to German being the second official language of Pennsylvania; however, Pennsylvania has no official language. Although more than 49 million Americans claim they have German ancestors, only 1.24 million Americans speak German at home. Many of these people are either Amish and Mennonites or Germans having newly immigrated (e.g. for professional reasons).

Pennsylvania Dutch/German

Pennsylvania Dutch or Pennsylvania German is a Palatine Rhine Franconian language that traditionally was spoken mainly in Pennsylvania, but that since the 19th century has spread to the Midwest (Ohio, Indiana, Iowa and other states), where the majority of speakers live today. It evolved from the German dialect of the Palatinate brought over to America by the Pennsylvania Dutch people before 1800. Originally spoken by adherents of different Christian denominations (Lutherans, Mennonites, Amish, German Baptist Brethren, Catholics) today it is mainly spoken by Amish and Old Order Mennonites.

Texas German

Texas German is a group of High German dialects spoken by German Texans, descendants of German immigrants who settled in Texas in the mid-19th century.

Hindi and Urdu

Hindi and Urdu are the two standard registers of the Hindustani language, an Indo-Aryan language native to North India, Central India, and Pakistan. While the formal registers draw vocabulary from Sanskrit and Arabic & Persian respectively, the colloquial forms are indistinguishable. Hindi and Urdu are widely spoken among the Indian who hails from Northern India and Pakistani communities in the United States as a first or second language. Speakers are concentrated in states with large South Asian populations, including California, Illinois, New Jersey, New York, Texas and Virginia.

Additionally, Hindi-Urdu (Hindustani) along with some other prominent Indian languages is an important cultural language for many South Asians who have different mother tongues and dialects. Bollywood in particular, as well as film music, is an important cultural product that influences many South Asian youth. Some Pakistanis, Indian Bengalis and Bangladeshis learn the language or its dialects through films.

Arabic

Varieties of Arabic are spoken by immigrants from the Middle East as well as many Muslim Americans. The highest concentrations of native Arabic speakers reside in heavily urban areas like Chicago, New York City, and Los Angeles. Detroit and the surrounding areas of Michigan boast a significant Arabic-speaking population including many Arab Christians of Lebanese, Syrian, and Palestinian descent.

Arabic is used for religious purposes by Muslim Americans and by some Arab Christians (notably Catholics of the Melkite and Maronite Churches as well as Rum Orthodox, i.e. Antiochian Orthodox Christians and Coptic churches.). A significant number of educated Arab professionals who immigrate often already know English quite well, as it is widely used in the Middle East. Lebanese immigrants also have a broader understanding of French as do many Arabic-speaking immigrants from North Africa.

Italian

The Italian language and its various dialects have been widely spoken in the United States for more than one hundred years, primarily due to large-scale immigration from the late 19th century to the mid 20th century.

In addition to Standard Italian learned by most people today, there has been a strong representation of the dialects and languages of Southern Italy amongst the immigrant population (Sicilian and Neapolitan in particular). As of 2009, though

15,638,348 American citizens report themselves as Italian-Americans, only 753,992 of these report speaking the Italian language at home (0.3264% of the US population).

Dutch

There has been a Dutch presence in America since 1602 when the government of the Republic of the Seven United Netherlands chartered the Dutch East India Company (*Vereenigde Oostindische Compagnie*, or VOC) with the mission of exploring for a passage to the Indies and claiming any uncharted territories for the Dutch republic. In 1664, English troops under the command of the Duke of York (later James II of England) attacked the New Netherland colony. Being greatly outnumbered, director general Peter Stuyvesant surrendered New Amsterdam, with Fort Orange following soon. New Amsterdam was renamed New York, Fort Orange was renamed Fort Albany. Dutch city names can still be found in New York's neighbourhoods. Harlem is named after Haarlem, Staten Island is named after Staten Eiland, and Brooklyn is named after Breukelen.

Dutch was still spoken in many parts of New York at the time of the Revolution. For example, Alexander Hamilton's wife Eliza Hamilton attended a Dutch-language church during their marriage.

African-American abolitionist and women's rights activist Sojourner Truth (born "Isabella Baumfree") was a native speaker of Dutch.

Martin Van Buren, the first President born in the United States following its independence, spoke Dutch as his native language, making him the only President whose first language was not English.

In a 1990 demographic consensus, 3% of surveyed citizens claimed descent from Dutch settlers. Modern estimates place the Dutch American population at 5 million, lagging just a bit behind Scottish Americans and Swedish Americans.

Notable Dutch Americans include the Roosevelts (Theodore Roosevelt, Franklin Delano Roosevelt, and Eleanor Roosevelt), Marlon Brando, Thomas Alva Edison, Martin Van Buren and the Vanderbilts. The Roosevelts are direct descendants of Dutch settlers of the New Netherland colony in the 17th century.

Around 142,000 people in the United States still speak the Dutch language at home as of 2013, in addition to 23,000 speakers of the closely related Afrikaans. They are concentrated mainly in California (23,500), Florida (10,900), Pennsylvania (9,900), Ohio (9,600), New York (8,700) and Michigan (6,600; i.e. almost entirely in the city of Holland).

A vernacular dialect of Dutch known as Jersey Dutch was spoken by a significant number of people in the New Jersey area between the start of the 17th century to the mid-20th century. With the beginning of the 20th century, usage of the language became restricted to internal family circles, with an ever-growing number of people abandoning the language in favor of English. It suffered gradual decline throughout the 20th century, and it ultimately dissipated from casual usage.

Telugu

In the second half of the 20th century, Telugu people from India (especially from Andhra Pradesh, Telangana, Karnataka and Tamil Nadu) migrated as professionals to the United States, Canada, Europe, and Southeast Asia. The Telugu American population enumerates over 886,988 individuals, Many big organisations such as Telugu Association of North America, American Telugu Association, etc are working and serving for the growth of the community.

Central New Jersey is home to the largest population concentration of Telugu people. Telugu people have also settled in New York City and the DC metropolitan area, as well as on the West Coast in Silicon Valley. The New York City and Los Angeles metropolitan areas are home to the largest concentrations of Telugu-speakers.

Vast majority of the Telugu people are Hindu with significant minorities like Muslims, Christians and even Buddhists.

Due to the rapidly increasing number of students and workers from Andhra Pradesh and Telangana in America, Telugu Americans have suffered from hate crimes in America. The most notable of these incidents was the 2017 Olathe, Kansas shooting, in which a white supremacist, Adam Purinton, harassed two Telugu immigrants, Srinivas Kuchibhotla and Alok Madasani, under the pretense that Kuchibhotla and Madasani were Iranians or illegal immigrants; later shooting them, killing Kuchibhotla and wounding Madasani as well as Ian Grillot, a white American who had come to the defense of Kuchibhotla and Madasani.

Notable people like Satya Nadella, Arvind Krishna, Shantanu Narayen, Nina Davuluri, Raja Kumari, Balamurali Ambati, Hari Kondabolu and several others were widely recognised and are successful in various fields in the United States.

Finnish

The first Finnish settlers in America were amongst the settlers who came from Sweden and Finland to the New Sweden colony. Most colonists were Finnish. However, the Finnish language was not preserved as well among subsequent generations as Swedish.

Between the 1890s and the outbreak of the first World War, an estimated quarter million Finnish citizens immigrated to the United States, mainly in rural areas of the Midwest and more specifically in the mining regions of Northeastern Minnesota, Northern Wisconsin and Michigan's Upper Peninsula. Hancock, Michigan, as of 2005, still incorporates bi-lingual street signs written in both English and Finnish. Americans of Finnish origin yield at 800,000 individuals, though only 26,000 speak the language at home. There is a distinctive dialect of English to be found in the Upper Peninsula, known as Yooper. Yooper often has a Finnish cadence and uses Finnish sentence structure with modified English, German, Swedish, Norwegian, and Finnish vocabulary. Notable Finnish Americans include U.S. Communist Party leader Gus Hall, film director Renny Harlin, and the Canadian-born actress Pamela Anderson.

Northern Clark County, Washington (encompassing Yacolt, Amboy, Battle Ground and Chelatchie) contains a large exclave of Old Apostolic Lutherans who originally immigrated from

Finland. Many families in this portion of the county speak fluent Finnish at home before learning English. Another noteworthy Finnish community in the United States is found in Lake Worth, Florida, north of Miami.

Russian

The Russian language is spoken in areas of some states, including New York, California, Washington, New Jersey, Illinois, Massachusetts, Pennsylvania, and Alaska. It is especially spoken in immigrant neighborhoods of some cities: New York City, Boston, Los Angeles, San Francisco, Philadelphia, Chicago, Seattle, Sacramento, Spokane, Miami, Vancouver, Washington, Portland, Oregon, and Woodburn, Oregon. The Russian-American Company used to own Alaska Territory until selling it after the Crimean War. Russian had always been limited, especially after the assassination of the Romanov dynasty of tsars. Starting in the 1970s and continuing until the mid-1990s, many Russian-speaking people from the Soviet Union and later its constituent republics such as Russia, Moldova, Ukraine, Belarus, and Uzbekistan have immigrated to the United States, increasing the language's usage in America.

The largest Russian-speaking neighborhoods in the United States are found in Brooklyn, Queens, and Staten Island in New York City (specifically the Brighton Beach area of Brooklyn); parts of Los Angeles, particularly West Los Angeles and West Hollywood; parts of Philadelphia, particularly the Far Northeast; and parts of Miami like Sunny Isles Beach.

In Nikolaevsk, Alaska, 66.57% of the population speaks Russian at home.

Slavic Voice of America media group serves Russian-speaking Americans out of Dallas.

Hebrew

Modern Hebrew is spoken by Israeli immigrants. Liturgical Hebrew is used as a religious or liturgical language by many of the United States' approximately 7 million Jews.

Ilocano

Like the Tagalogs, the Ilocanos are an Austronesian stock which came from the Philippines. They were the first Filipinos to migrate en masse to the United States. They first entered the State of Hawaii and worked there in the vast plantations.

As they did in the Philippine provinces of Northern Luzon and Mindanao, they quickly gained importance in the areas where they settled. Thus, the state of Hawaii became no less different from the Philippines in terms of percentage of Ilocano speakers.

Like Tagalog, Ilocano is also being taught in universities where most of the Filipinos reside.

South Asian languages

There are many South Asians in the United States. These include Indians, Pakistanis, and Bangladeshis, who speak various South Asian languages. Major South Asian languages spoken in the US include Telugu (see "Telugu" above), Malayalam, Kannada, Tamil (see "Tamil" below), Gujarati, Hindi and Urdu (see 'Hindi-Urdu" above), Bengali, Punjabi, Sinhala, Nepali (see "Nepali" below), and Marathi.

Tamil

The Tamil community in the United States is largely bilingual. Tamil is taught in weekly classes in many Hindu temples and by associations such as the American Tamil Academy in South Brunswick, Tamil Jersey School in Jersey City, New Jersey,

The written form of the language is highly formal and quite distinct from the spoken form. A few universities, such as the University of Chicago and the University of California Berkeley, have graduate programs in the language.

In the second half of the 20th century, Tamils from India migrated as skilled professionals to the United States, Canada, Europe, and Southeast Asia. The Tamil American population numbers over 195,685 individuals, and the Federation of Tamil Sangams of North America functions as an umbrella organization for the growing community.

The New York City and Los Angeles metropolitan areas are home to the largest concentrations of Tamil-speaking Sri

Lankan Americans. New York City's Staten Island alone is estimated to be home to more than 5,000 Sri Lankan Americans, one of the largest Sri Lankan populations outside Sri Lanka itself, and a significant proportion of whom speak Tamil.

Central New Jersey is home to the largest population concentration of Tamils. New Jersey houses its own Tamil Sangam. Sizeable populations of Indian American Tamils have also settled in the New York City and Washington metropolitan areas, as well as on the West Coast in Silicon Valley, where there are Tamil associations such as the Bay Area Tamil Mandram. The New York City and Los Angeles metropolitan areas are home to the largest concentrations of Tamil-speaking Sri Lankan Americans, with New York City's Staten Island alone estimated to be home to more than 5,000 Sri Lankan Americans, one of the largest Sri Lankan populations outside Sri Lanka itself, and a significant proportion of whom speak Tamil.

Irish

About 40 million Americans have Irish ancestry, many of whose ancestors would have spoken Irish. In 2013, around 20,600 Americans spoke Irish at home. An additional 1,600 spoke Scottish Gaelic. As of 2008 it was the 76th most spoken language in the United States.

Khmer (Cambodian)

Between 1981 and 1985 about 150,000 Cambodians resettled in the United States. Before 1975 very few Cambodians came to the United States. Those who did were children of upper-class families sent abroad to attend school. After the fall of Phnom Penh to the communist Khmer Rouge in 1975, some Cambodians managed to escape. In 2007 the American Community Survey reported that there were approximately 200,000 Cambodians living in the United States, making up about 2% percent of the Asian population. This population is, however, heavily concentrated in two areas: the Los Angeles metropolitan area in California, especially the city of Long Beach; and Greater Boston in New England, especially Lowell, Massachusetts. These two areas hold a majority of the Cambodians living in the US.

Polish and Silesian

As of 2013, around 580,000 Americans spoke Polish at home. The Polish language is very common in the Chicago metropolitan area. Chicago's third largest white ethnic groups are those of Polish descent, after German and Irish. The Polish people and the Polish language in Chicago were very prevalent in the early years of the city, and today the 650,000 Poles in Chicago make up one of the largest ethnically Polish populations in the world, comparable to the city of Wrocław, the fourth largest city in Poland. That makes it one of the most important centers of Polonia and the Polish language in the United States, a fact that the city celebrates every Labor Day weekend at the Taste of Polonia Festival in Jefferson Park.

Texas Silesian, a dialect of the Silesian language (itself controversially considered a branch of Polish by some linguists), has been used by Texas Silesians in American settlements from 1852 to the present.

Portuguese

The first Portuguese speakers in America were Portuguese Jews who had fled the Portuguese Inquisition. They spoke Judeo-Portuguese and founded the earliest Jewish communities in the Thirteen Colonies, two of which still exist: Congregation Shearith Israel in New York and Congregation Mikveh Israel in Philadelphia. However, by the end of the 18th century, their use of Portuguese had been replaced by English.

In the late 19th century, many Portuguese, mainly Azoreans, Madeirans and Cape Verdeans (who prior to independence in 1975 were Portuguese citizens), immigrated to the United States, settling in cities like Providence, Rhode Island, New Bedford, Massachusetts, and Santa Cruz, California. There was also a substantial Portuguese immigration to Hawaii, which at the time was not yet part of the United States.

In the mid-late 20th century there was another wave of Portuguese immigration to the US, mainly the Northeast (New Jersey, New York, Connecticut, Massachusetts), and for a time Portuguese became a major language in Newark, New Jersey. Many Portuguese Americans may include descendants of Portuguese settlers born in Portuguese Africa (known as Portuguese Africans, or, in Portugal, as *retornados*) and Asia (mostly Macau). There were around 1 million Portuguese Americans in the United States by the year 2000. Portuguese

(European Portuguese) has been spoken in the United States by small communities of immigrants, mainly in the metropolitan New York City area, like Newark, New Jersey.

The Portuguese language is also spoken widely by Brazilian Americans, concentrated in Miami, New York City, and Boston.

Swedish

There has been a Swedish presence in America since the New Sweden colony came into existence in March 1638.

Widespread diaspora of Swedish immigration did not occur until the latter half of the 19th century, bringing in a total of a million Swedes. No other country had a higher percentage of its people leave for the United States except Ireland and Norway. At the beginning of the 20th century, Minnesota had the highest ethnic Swedish population in the world after the city of Stockholm.

3.7% of US residents claim descent from Scandinavian ancestors, amounting to roughly 11–12 million people. According to SIL's Ethnologue, over half a million ethnic Swedes still speak the language, though according to the 2007 American Community Survey only 56,715 speak it at home. Cultural assimilation has contributed to the gradual and steady decline of the language in the US. After the independence of the US from the Kingdom of Great Britain, the government encouraged colonists to adopt the English language as a common medium of communication, and in some cases, imposed it upon them. Subsequent generations of Swedish Americans received education in English and spoke it

as their first language. Lutheran churches scattered across the Midwest started abandoning Swedish in favor of English as their language of worship. Swedish newspapers and publications alike slowly faded away.

There are sizable Swedish communities in Minnesota, Ohio, Maryland, Philadelphia, and Delaware, along with small isolated pockets in Pennsylvania, San Francisco, Fort Lauderdale, and New York. Chicago once contained a large Swedish enclave called Andersonville on the city's north side.

John Morton, the person who cast the decisive vote leading to Pennsylvania's support for the United States Declaration of Independence, was of Finnish descent. Finland was part of the Kingdom of Sweden in the 18th century.

Welsh

Up to two million Americans are thought to have Welsh ancestry. However, there is very little Welsh being used commonly in the United States. According to the 2007 American Community Survey, 2,285 people speak Welsh at home; primarily spoken in California (415), Florida (225), New York (204), Ohio (135), and New Jersey (130). Some place names, such as Bryn Mawr in Chicago and Bryn Mawr, Pennsylvania (English: Big Hill) are Welsh. Several towns in Pennsylvania, mostly in the Welsh Tract, have Welsh namesakes, including Uwchlan, Bala Cynwyd, Gwynedd, and Tredyffrin.

Yiddish

Yiddish has a much longer history in the United States than Hebrew. It has been present since at least the late 19th century and continues to have roughly 148,000 speakers as of the 2009 American Community Survey. Though they came from varying geographic backgrounds and nuanced approaches to worship, immigrant Jews of Eastern Europe and Russia were often united under a common understanding of the Yiddish language once they settled in America, and at one point dozens of publications were available in most East Coast cities. Though it has declined by quite a bit since the end of WWII, it has by no means disappeared. Many Israeli immigrants and expatriates have at least some understanding of the language in addition to Hebrew, and many of the descendants of the great migration of Ashkenazi Jews of the past century pepper their mostly English vocabulary with some loan words. Furthermore, it is a lingua franca among Orthodox Jewry (particularly Hasidic Jewry), concentrated in Los Angeles, Miami, and New York. A significant diffusion of Yiddish loan words into the non-Jewish population continues to be a distinguishing feature of New York City English.

Nepali

The first Nepalese to enter the United States were classified as "other Asian". Immigration records show that between 1881 and 1890, 1,910 "other Asians" were admitted to the United States. However, Nepal did not open its borders until 1950, and most Nepalis who left the country during that time went primarily to India to study. Nepalese Americans were first

classified as a separate ethnic group in 1974 when 56 Nepalese immigrated to the United States. New York City, Boston, Chicago, Denver, Gainesville, Florida, Portland, Oregon and Saint Paul, Minnesota have the largest number of Nepalese. There are some Nepalese community or cultural events in every American state, including Dashain, Tihar, Losar, Holi, Teej Special, and Nepali New Year.

New American languages

Several distinct natural languages and pidgins have developed on American soil, after the founding of the United States as a nation, including full languages like creole and sign languages.

Spoken languages

Angloromani

Angloromani is an English creole or mixed language spoken by Romani Americans.

Chinuk Wawa or Chinook Jargon

A pidgin of 700–800 words of French, English, Cree and other Native origins is the old trade language of the Pacific Northwest. It was used extensively among both European and Native peoples of the Oregon Territory, even used instead of English at home by many pioneer families. It is estimated that around 100,000 people spoke it at its peak, between 1858 and

1900, and it was last widely used in Seattle just before World War II.

Gullah

An English-African creole language spoken on the Sea Islands of South Carolina and Georgia retains strong influences of West African languages. The language is sometimes referred to as "Geechee".

Hawai'i Pidgin

The Hawai'ian English creole language, locally known as Hawaiian Pidgin, is used by locals and is considered an unofficial language of the state of Hawaii.

Louisiana Creole French

A French Creole language spoken by the Louisiana Creole people of the state of Louisiana, close to Haitian Creole, Colonial French, and Cajun French (language of Acadians deported from New France after 1755 and the Grand Dérangement). French Creole languages are spoken by millions of people worldwide, mainly in the United States, Caribbean, and Indian Ocean areas.

Sign languages

Alongside the numerous and varied oral languages, the United States also boasts several sign languages. Historically, the US was home to some six or more sign languages (that number rising with the probability that Plains Sign Talk is actually a language family with several languages under its umbrella) which has fallen with the death of several of these.

As with all sign languages around the world that developed organically, these are full languages distinct from any oral language. American Sign Language (unlike Signed English) is not a derivation of English. Some languages present here were trade pidgins which were used first as a system of communication across national and linguistic boundaries of the Native Americans, however, they have since developed into mature languages as children learned them as a first language.

American Sign Language

American Sign Language (ASL) is the native language of a number of deaf and hearing people in America (roughly 100,000 to 500,000). While some sources have stated that ASL is the third most frequently used language in the United States, after English and Spanish, recent scholarship has pointed out that most of these estimates are based on numbers conflating deafness with ASL use, and that the last actual study of this (in 1972) seems to indicate an upper bound of 500,000 ASL speakers at the time.

- Black American Sign Language (BASL) developed in the southeastern US, where separate residential schools were maintained for white and black deaf children. BASL shares much of the same vocabulary and grammatical structure as ASL and is generally considered one of its dialects.

Hawai'i Sign Language

Hawaii Sign Language is moribund with only a handful of speakers on O'ahu, Lana'i, Kaua'i and possibly Ni'ihau. Some of these speakers may actually be speaking a creolized version of HSL and ASL, however, research is slow-going. The language was once called Hawai'i Pidgin Sign Language as many people thought it was a derivation of ASL which was discovered to be false and to be a separate language altogether.

Plains Sign Talk

Once a trade pidgin and the most far-reaching sign language in North America, **Plains Sign Talk** or **Plains Sign Language** is now critically endangered with an unknown number of speakers.

- Navajo Sign Language has been found to be in use in one clan of Navajo; however, whether it is a dialect of Plains Sign Talk or a separate language remains unknown.
- Plateau Sign Language is another trade pidgin that may have become a separate language, **Plateau Sign**

Language replaced Plains Sign Talk in the Columbia Plateau and surrounding regions of British Columbia, Washington, Oregon, and Idaho. It is now extinct.

Martha's Vineyard Sign Language

Martha's Vineyard Sign Language is now extinct. Along with French Sign Language, it was one of several main contributors to American Sign Language.

Henniker Sign Language

Henniker Sign Language is now extinct but was once found around the Henniker region of New Hampshire and formed a basis for American Sign Language.

Sandy River Valley Sign Language

Sandy River Valley Sign Language is now extinct but once could be found around the Sandy River Valley in Maine. It was one of several main contributors to American Sign Language.

Chapter 3

Arabic

Arabic is a Semitic language that first emerged in the 1st to 4th centuries CE. It is now the lingua franca of the Arab world. It is named after the Arabs, a term initially used to describe peoples living in the Arabian Peninsula bounded by eastern Egypt in the west, Mesopotamia in the east, and the Anti-Lebanon mountains and Northern Syria in the north, as perceived by ancient Greek geographers. The ISO assigns language codes to 32 varieties of Arabic, including its standard form, Modern Standard Arabic, also referred to as Literary Arabic, which is modernized Classical Arabic. This distinction exists primarily among Western linguists; Arabic speakers themselves generally do not distinguish between Modern Standard Arabic and Classical Arabic, but rather refer to both as *al-ʿarabiyyatu l-fuṣḥā* (ٱلْعَرَبِيَّة ٱلْفُصْحَىٰ "the eloquent Arabic") or simply *al-fuṣḥā* (ٱلْفُصْحَىٰ). Modern Standard Arabic is an official language of 26 states and 1 disputed territory, the third most after English and French.

Arabic is widely taught in schools and universities around the world and is used to varying degrees in workplaces, governments and the media. Arabic, in its standard form, is the official language of 26 states, as well as the liturgical language of the religion of Islam, since the Quran and the Hadiths were written in Arabic.

During the Middle Ages, Arabic was a major vehicle of culture in Europe, especially in science, mathematics and philosophy.

As a result, many European languages have also borrowed many words from it. Arabic influence, mainly in vocabulary, is seen in European languages—mainly Spanish and to a lesser extent Portuguese, Catalan, and Sicilian—owing to both the proximity of Christian European and Muslim Arab civilizations and the long-lasting Arabic culture and language presence, mainly in Southern Iberia, during the Al-Andalus era. The Maltese language is a Semitic language developed from a dialect of Arabic and written in the Latin alphabet. The Balkan languages, including Greek and Bulgarian, have also acquired a significant number of words of Arabic origin through contact with Ottoman Turkish.

Arabic has influenced many other languages around the globe throughout its history. Some of the most influenced languages are Persian, Turkish, Hindustani (Hindi and Urdu), Kashmiri, Kurdish, Bosnian, Kazakh, Bengali, Malay (Indonesian and Malaysian), Maldivian, Pashto, Punjabi, Albanian, Armenian, Azerbaijani, Sicilian, Spanish, Greek, Bulgarian, Tagalog, Sindhi, Odia and Hausa and some languages in parts of Africa. Conversely, Arabic has borrowed words from other languages, including Hebrew, Greek, Aramaic, and Persian in medieval times and languages such as English and French in modern times.

Arabic is the religious language of 1.8 billion Muslims, and Arabic is one of six official languages of the United Nations. All varieties of Arabic combined are spoken by perhaps as many as 422 million speakers (native and non-native) in the Arab world, making it the fifth most spoken language in the world. Arabic is written with the Arabic alphabet, which is an abjad script and is written from right to left, although the spoken varieties

are sometimes written in ASCII Latin from left to right with no standardized orthography.

Classification

Arabic is usually, but not universally, classified as a Central Semitic language. It is related to languages in other subgroups of the Semitic language group (Northwest Semitic, South Semitic, East Semitic, West Semitic), such as Aramaic, Syriac, Hebrew, Ugaritic, Phoenician, Canaanite, Amorite, Ammonite, Eblaite, epigraphic Ancient North Arabian, epigraphic Ancient South Arabian, Ethiopic, Modern South Arabian, and numerous other dead and modern languages. Linguists still differ as to the best classification of Semitic language sub-groups. The Semitic languages changed a great deal between Proto-Semitic and the emergence of the Central Semitic languages, particularly in grammar. Innovations of the Central Semitic languages—all maintained in Arabic—include:

- The conversion of the suffix-conjugated stative formation (*jalas-*) into a past tense.
- The conversion of the prefix-conjugated preterite-tense formation (*yajlis-*) into a present tense.
- The elimination of other prefix-conjugated mood/aspect forms (e.g., a present tense formed by doubling the middle root, a perfect formed by infixing a /t/ after the first root consonant, probably a jussive formed by a stress shift) in favor of new moods formed by endings attached to the prefix-conjugation forms (e.g., *-u* for indicative, *-a* for

subjunctive, no ending for jussive, *-an* or *-anna* for energetic).

- The development of an internal passive.

There are several features which Classical Arabic, the modern Arabic varieties, as well as the Safaitic and Hismaic inscriptions share which are unattested in any other Central Semitic language variety, including the Dadanitic and Taymanitic languages of the northern Hejaz. These features are evidence of common descent from a hypothetical ancestor, Proto-Arabic. The following features can be reconstructed with confidence for Proto-Arabic:

- negative particles *m* * /mā/; *l'n* */lā-ʾ an/ to Classical Arabic lan
- mafˁ ūl G-passive participle
- prepositions and adverbs *f*, *ˁn*, *ˁnd*, *ḥt*, *ˁkdy*
- a subjunctive in *-a*
- *t*-demonstratives
- leveling of the *-at* allomorph of the feminine ending
- *ʾn* complementizer and subordinator
- the use of *f-* to introduce modal clauses
- independent object pronoun in *(ʾ)y*
- vestiges of *nunation*

History

Old Arabic

Arabia boasted a wide variety of Semitic languages in antiquity. In the southwest, various Central Semitic languages

both belonging to and outside of the Ancient South Arabian family (e.g. Southern Thamudic) were spoken. It is also believed that the ancestors of the Modern South Arabian languages (non-Central Semitic languages) were also spoken in southern Arabia at this time. To the north, in the oases of northern Hejaz, Dadanitic and Taymanitic held some prestige as inscriptional languages. In Najd and parts of western Arabia, a language known to scholars as Thamudic C is attested. In eastern Arabia, inscriptions in a script derived from ASA attest to a language known as Hasaitic. Finally, on the northwestern frontier of Arabia, various languages known to scholars as Thamudic B, Thamudic D, Safaitic, and Hismaic are attested. The last two share important isoglosses with later forms of Arabic, leading scholars to theorize that Safaitic and Hismaic are in fact early forms of Arabic and that they should be considered Old Arabic.

Linguists generally believe that "Old Arabic" (a collection of related dialects that constitute the precursor of Arabic) first emerged around the 1st century CE. Previously, the earliest attestation of Old Arabic was thought to be a single 1st century CE inscription in Sabaic script at Qaryat Al-Faw, in southern present-day Saudi Arabia. However, this inscription does not participate in several of the key innovations of the Arabic language group, such as the conversion of Semitic mimation to nunation in the singular. It is best reassessed as a separate language on the Central Semitic dialect continuum.

It was also thought that Old Arabic coexisted alongside—and then gradually displaced--epigraphic Ancient North Arabian (ANA), which was theorized to have been the regional tongue for many centuries. ANA, despite its name, was considered a

very distinct language, and mutually unintelligible, from "Arabic". Scholars named its variant dialects after the towns where the inscriptions were discovered (Dadanitic, Taymanitic, Hismaic, Safaitic). However, most arguments for a single ANA language or language family were based on the shape of the definite article, a prefixed h-. It has been argued that the h- is an archaism and not a shared innovation, and thus unsuitable for language classification, rendering the hypothesis of an ANA language family untenable. Safaitic and Hismaic, previously considered ANA, should be considered Old Arabic due to the fact that they participate in the innovations common to all forms of Arabic.

The earliest attestation of continuous Arabic text in an ancestor of the modern Arabic script are three lines of poetry by a man named Garm(')allāhe found in En Avdat, Israel, and dated to around 125 CE. This is followed by the Namara inscription, an epitaph of the Lakhmid king Mar 'al-Qays bar 'Amro, dating to 328 CE, found at Namaraa, Syria. From the 4th to the 6th centuries, the Nabataean script evolves into the Arabic script recognizable from the early Islamic era. There are inscriptions in an undotted, 17-letter Arabic script dating to the 6th century CE, found at four locations in Syria (Zabad, Jabal 'Usays, Harran, Umm al-Jimaal). The oldest surviving papyrus in Arabic dates to 643 CE, and it uses dots to produce the modern 28-letter Arabic alphabet. The language of that papyrus and of the Qur'an are referred to by linguists as "Quranic Arabic", as distinct from its codification soon thereafter into "Classical Arabic".

Old Hejazi and Classical Arabic

- In late pre-Islamic times, a transdialectal and transcommunal variety of Arabic emerged in the Hejaz which continued living its parallel life after literary Arabic had been institutionally standardized in the 2nd and 3rd century of the Hijra, most strongly in Judeo-Christian texts, keeping alive ancient features eliminated from the "learned" tradition (Classical Arabic). This variety and both its classicizing and "lay" iterations have been termed Middle Arabic in the past, but they are thought to continue an Old Higazi register. It is clear that the orthography of the Qur'an was not developed for the standardized form of Classical Arabic; rather, it shows the attempt on the part of writers to record an archaic form of Old Higazi.

In the late 6th century AD, a relatively uniform intertribal "poetic koine" distinct from the spoken vernaculars developed based on the Bedouin dialects of Najd, probably in connection with the court of al-Ḥīra. During the first Islamic century, the majority of Arabic poets and Arabic-writing persons spoke Arabic as their mother tongue. Their texts, although mainly preserved in far later manuscripts, contain traces of non-standardized Classical Arabic elements in morphology and syntax.

Standardization

Abu al-Aswad al-Du'ali (c. 603–689) is credited with standardizing Arabic grammar, or *an-naḥw* (النّحو "the way"), and pioneering a system of diacritics to differentiate consonants (نقط الإعجام *nuqat l-iʾjām* "pointing for non-Arabs") and indicate vocalization (التشكيل *at-tashkil*). Al-Khalil ibn Ahmad al-Farahidi (718 – 786) compiled the first Arabic dictionary, *Kitāb al-ʾAyn* (كتاب العين "The Book of the Letter ع"), and is credited with establishing the rules of Arabic prosody. Al-Jahiz (776-868) proposed to Al-Akhfash al-Akbar an overhaul of the grammar of Arabic, but it would not come to pass two centuries. The standardization of Arabic reached completion around the end of the 8th century. The first comprehensive description of the ʿarabiyya "Arabic", Sībawayhi's *al-Kitāb*, is based first of all upon a corpus of poetic texts, in addition to Qur'an usage and Bedouin informants whom he considered to be reliable speakers of the ʿarabiyya.

Spread

Arabic spread with the spread of Islam. Following the early Muslim conquests, Arabic gained vocabulary from Middle Persian and Turkish. In the early Abbasid period, many Classical Greek terms entered Arabic through translations carried out at Baghdad's House of Wisdom.

By the 8th century, knowledge of Classical Arabic had become an essential prerequisite for rising into the higher classes throughout the Islamic world, both for Muslims and non-Muslims. For example, Maimonides, the Andalusi Jewish

philosopher, authored works in Judeo-Arabic—Arabic written in Hebrew script—including his famous *The Guide for the Perplexed* (دلالة الحائرين*Dalālat al-☐āʾirīn*).

Development

Ibn Jinni of Mosul, a pioneer in phonology, wrote prolifically in the 10th century on Arabic morphology and phonology in works such as *Kitāb Al-Munṣif, Kitāb Al-Muḥtasab, andKitāb Al-Khaṣāʾiṣ* [ar].

Ibn Mada' of Cordoba (1116–1196) realized the overhaul of Arabic grammar first proposed by Al-Jahiz 200 years prior.

The Maghrebi lexicographer Ibn Manzur compiled *Lisān al-ʿArab* [ar] (لســــانالعرب, "Tongue of Arabs"), a major reference dictionary of Arabic, in 1290.

Neo-Arabic

Charles Ferguson's koine theory (Ferguson 1959) claims that the modern Arabic dialects collectively descend from a single military koine that sprang up during the Islamic conquests; this view has been challenged in recent times. Ahmad al-Jallad proposes that there were at least two considerably distinct types of Arabic on the eve of the conquests: Northern and Central (Al-Jallad 2009). The modern dialects emerged from a new contact situation produced following the conquests. Instead of the emergence of a single or multiple koines, the dialects contain several sedimentary layers of borrowed and areal features, which they absorbed at different points in their

linguistic histories. According to Veersteegh and Bickerton, colloquial Arabic dialects arose from pidginized Arabic formed from contact between Arabs and conquered peoples. Pidginization and subsequent creolization among Arabs and arabized peoples could explain relative morphological and phonological simplicity of vernacular Arabic compared to Classical and MSA.

In around the 11th and 12th centuries in al-Andalus, the *zajal* and *muwashah* poetry forms developed in the dialectical Arabic of Cordoba and the Maghreb.

Nahda

In the wake of the industrial revolution and European hegemony and colonialism, pioneering Arabic presses, such as the Amiri Press established by Muhammad Ali (1819), dramatically changed the diffusion and consumption of Arabic literature and publications.

The *Nahda* cultural renaissance saw the creation of a number of Arabic academies modeled after the *Académie française* that aimed to develop the Arabic lexicon to suit these transformations, first in Damascus (1919), then in Cairo (1932), Baghdad (1948), Rabat (1960), Amman (1977), Khartum [ar] (1993), and Tunis (1993). In 1997, a bureau of Arabization standardization was added to the Educational, Cultural, and Scientific Organization of the Arab League. These academies and organizations have worked toward the Arabization of the sciences, creating terms in Arabic to describe new concepts, toward the standardization of these new terms throughout the Arabic-speaking world, and toward

the development of Arabic as a world language. This gave rise to what Western scholars call Modern Standard Arabic.

From the 1950s, Arabization became a postcolonial nationalist policy in countries such as Tunisia, Algeria, Morocco, and Sudan.

Classical, Modern Standard and spoken Arabic

Arabic usually refers to Standard Arabic, which Western linguists divide into Classical Arabic and Modern Standard Arabic. It could also refer to any of a variety of regional vernacular Arabic dialects, which are not necessarily mutually intelligible.

Classical Arabic is the language found in the Quran, used from the period of Pre-Islamic Arabia to that of the Abbasid Caliphate. Classical Arabic is prescriptive, according to the syntactic and grammatical norms laid down by classical grammarians (such as Sibawayh) and the vocabulary defined in classical dictionaries (such as the *Lisān al-ʻArab*).

Modern Standard Arabic (MSA) largely follows the grammatical standards of Classical Arabic and uses much of the same vocabulary. However, it has discarded some grammatical constructions and vocabulary that no longer have any counterpart in the spoken varieties and has adopted certain new constructions and vocabulary from the spoken varieties. Much of the new vocabulary is used to denote concepts that

have arisen in the industrial and post-industrial era, especially in modern times. Due to its grounding in Classical Arabic, Modern Standard Arabic is removed over a millennium from everyday speech, which is construed as a multitude of dialects of this language. These dialects and Modern Standard Arabic are described by some scholars as not mutually comprehensible. The former are usually acquired in families, while the latter is taught in formal education settings. However, there have been studies reporting some degree of comprehension of stories told in the standard variety among preschool-aged children. The relation between Modern Standard Arabic and these dialects is sometimes compared to that of Classical Latin and Vulgar Latin vernaculars (which became Romance languages) in medieval and early modern Europe. This view though does not take into account the widespread use of Modern Standard Arabic as a medium of audiovisual communication in today's mass media—a function Latin has never performed.

MSA is the variety used in most current, printed Arabic publications, spoken by some of the Arabic media across North Africa and the Middle East, and understood by most educated Arabic speakers. "Literary Arabic" and "Standard Arabic" (فُصْحَى *fuṣḥá*) are less strictly defined terms that may refer to Modern Standard Arabic or Classical Arabic.

Some of the differences between Classical Arabic (CA) and Modern Standard Arabic (MSA) are as follows:

- Certain grammatical constructions of CA that have no counterpart in any modern vernacular dialect

(e.g., the energetic mood) are almost never used in Modern Standard Arabic.

- Case distinctions are very rare in Arabic vernaculars. As a result, MSA is generally composed without case distinctions in mind, and the proper cases are added after the fact, when necessary. Because most case endings are noted using final short vowels, which are normally left unwritten in the Arabic script, it is unnecessary to determine the proper case of most words. The practical result of this is that MSA, like English and Standard Chinese, is written in a strongly determined word order and alternative orders that were used in CA for emphasis are rare. In addition, because of the lack of case marking in the spoken varieties, most speakers cannot consistently use the correct endings in extemporaneous speech. As a result, spoken MSA tends to drop or regularize the endings except when reading from a prepared text.
- The numeral system in CA is complex and heavily tied in with the case system. This system is never used in MSA, even in the most formal of circumstances; instead, a significantly simplified system is used, approximating the system of the conservative spoken varieties.

MSA uses much Classical vocabulary (e.g., *dhahaba* 'to go') that is not present in the spoken varieties, but deletes Classical words that sound obsolete in MSA. In addition, MSA has borrowed or coined many terms for concepts that did not exist in Quranic times, and MSA continues to evolve. Some words have been borrowed from other languages—notice that

transliteration mainly indicates spelling and not real pronunciation (e.g., فِلْم *film* 'film' or ديمقراطية *dīmuqrāṭiyyah* 'democracy').

However, the current preference is to avoid direct borrowings, preferring to either use loan translations (e.g., فرع *farʻ* 'branch', also used for the branch of a company or organization; جناح *janāḥ* 'wing', is also used for the wing of an airplane, building, air force, etc.), or to coin new words using forms within existing roots (استماتة *istimātah* 'apoptosis', using the root موت *m/w/t* 'death' put into the Xth form, or جامعة *jāmiʻah* 'university', based on جمع *jamaʻa* 'to gather, unite'; جمهورية *jumhūriyyah* 'republic', based on جمهور *jumhūr* 'multitude'). An earlier tendency was to redefine an older word although this has fallen into disuse (e.g., هاتف *hātif* 'telephone' < 'invisible caller (in Sufism)'; جريدة *jarīdah* 'newspaper' < 'palm-leaf stalk').

Colloquial or *dialectal* Arabic refers to the many national or regional varieties which constitute the everyday spoken language and evolved from Classical Arabic. Colloquial Arabic has many regional variants; geographically distant varieties usually differ enough to be mutually unintelligible, and some linguists consider them distinct languages. However, research indicates a high degree of mutual intelligibility between closely related Arabic variants for native speakers listening to words, sentences, and texts; and between more distantly related dialects in interactional situations.

The varieties are typically unwritten. They are often used in informal spoken media, such as soap operas and talk shows, as well as occasionally in certain forms of written media such as poetry and printed advertising.

The only variety of modern Arabic to have acquired official language status is Maltese, which is spoken in (predominantly Catholic) Malta and written with the Latin script. It is descended from Classical Arabic through Siculo-Arabic, but is not mutually intelligible with any other variety of Arabic. Most linguists list it as a separate language rather than as a dialect of Arabic.

Even during Muhammad's lifetime, there were dialects of spoken Arabic. Muhammad spoke in the dialect of Mecca, in the western Arabian peninsula, and it was in this dialect that the Quran was written down. However, the dialects of the eastern Arabian peninsula were considered the most prestigious at the time, so the language of the Quran was ultimately converted to follow the eastern phonology. It is this phonology that underlies the modern pronunciation of Classical Arabic. The phonological differences between these two dialects account for some of the complexities of Arabic writing, most notably the writing of the glottal stop or *hamzah* (which was preserved in the eastern dialects but lost in western speech) and the use of *alif maqṣūrah* (representing a sound preserved in the western dialects but merged with *ā* in eastern speech).

Language and dialect

The sociolinguistic situation of Arabic in modern times provides a prime example of the linguistic phenomenon of diglossia, which is the normal use of two separate varieties of the same language, usually in different social situations. *Tawleed* is the process of giving a new shade of meaning to an

old classical word. For example, *al-hatif* lexicographically, means the one whose sound is heard but whose person remains unseen. Now the term *al-hatif* is used for a telephone. Therefore, the process of *tawleed* can express the needs of modern civilization in a manner that would appear to be originally Arabic. In the case of Arabic, educated Arabs of any nationality can be assumed to speak both their school-taught Standard Arabic as well as their native, mutually unintelligible "dialects"; these dialects linguistically constitute separate languages which may have dialects of their own. When educated Arabs of different dialects engage in conversation (for example, a Moroccan speaking with a Lebanese), many speakers code-switch back and forth between the dialectal and standard varieties of the language, sometimes even within the same sentence. Arabic speakers often improve their familiarity with other dialects via music or film.

The issue of whether Arabic is one language or many languages is politically charged, in the same way it is for the varieties of Chinese, Hindi and Urdu, Serbian and Croatian, Scots and English, etc. In contrast to speakers of Hindi and Urdu who claim they cannot understand each other even when they can, speakers of the varieties of Arabic will claim they can all understand each other even when they cannot. The issue of diglossia between spoken and written language is a significant complicating factor: A single written form, significantly different from any of the spoken varieties learned natively, unites a number of sometimes divergent spoken forms. For political reasons, Arabs mostly assert that they all speak a single language, despite significant issues of mutual incomprehensibility among differing spoken versions.

From a linguistic standpoint, it is often said that the various spoken varieties of Arabic differ among each other collectively about as much as the Romance languages. This is an apt comparison in a number of ways. The period of divergence from a single spoken form is similar—perhaps 1500 years for Arabic, 2000 years for the Romance languages. Also, while it is comprehensible to people from the Maghreb, a linguistically innovative variety such as Moroccan Arabic is essentially incomprehensible to Arabs from the Mashriq, much as French is incomprehensible to Spanish or Italian speakers but relatively easily learned by them. This suggests that the spoken varieties may linguistically be considered separate languages.

Influence of Arabic on other languages

The influence of Arabic has been most important in Islamic countries, because it is the language of the Islamic sacred book, the Quran. Arabic is also an important source of vocabulary for languages such as Amharic, Azerbaijani, Baluchi, Bengali, Berber, Bosnian, Chaldean, Chechen, Chittagonian, Croatian, Dagestani, English, German, Gujarati, Hausa, Hindi, Kazakh, Kurdish, Kutchi, Kyrgyz, Malay (Malaysian and Indonesian), Pashto, Persian, Punjabi, Rohingya, Romance languages (French, Catalan, Italian, Portuguese, Sicilian, Spanish, etc.) Saraiki, Sindhi, Somali, Sylheti, Swahili, Tagalog, Tigrinya, Turkish, Turkmen, Urdu, Uyghur, Uzbek, Visayan and Wolof, as well as other languages in countries where these languages are spoken.

The Education Minister of France Jean-Michel Blanquer has emphasized the learning and usage of Arabic in French schools.

In addition, English has many Arabic loanwords, some directly, but most via other Mediterranean languages. Examples of such words include admiral, adobe, alchemy, alcohol, algebra, algorithm, alkaline, almanac, amber, arsenal, assassin, candy, carat, cipher, coffee, cotton, ghoul, hazard, jar, kismet, lemon, loofah, magazine, mattress, sherbet, sofa, sumac, tariff, and zenith. Other languages such as Maltese and Kinubi derive ultimately from Arabic, rather than merely borrowing vocabulary or grammatical rules.

Terms borrowed range from religious terminology (like Berber *taẓallit*, "prayer", from *salat* (صلاة*ṣalāh*)), academic terms (like Uyghur *mentiq*, "logic"), and economic items (like English *coffee*) to placeholders (like Spanish *fulano*, "so-and-so"), everyday terms (like Hindustani *lekin*, "but", or Spanish *taza* and French *tasse*, meaning "cup"), and expressions (like Catalan *a betzef*, "galore, in quantity"). Most Berber varieties (such as Kabyle), along with Swahili, borrow some numbers from Arabic. Most Islamic religious terms are direct borrowings from Arabic, such as صلاة (*salat*), "prayer", and إمام (*imam*), "prayer leader."

In languages not directly in contact with the Arab world, Arabic loanwords are often transferred indirectly via other languages rather than being transferred directly from Arabic. For example, most Arabic loanwords in Hindustani and Turkish entered through Persian. Older Arabic loanwords in Hausa were borrowed from Kanuri.

Arabic words also made their way into several West African languages as Islam spread across the Sahara. Variants of Arabic words such as كتاب *kitāb* ("book") have spread to the languages of African groups who had no direct contact with Arab traders.

Since throughout the Islamic world, Arabic occupied a position similar to that of Latin in Europe, many of the Arabic concepts in the fields of science, philosophy, commerce, etc. were coined from Arabic roots by non-native Arabic speakers, notably by Aramaic and Persian translators, and then found their way into other languages. This process of using Arabic roots, especially in Kurdish and Persian, to translate foreign concepts continued through to the 18th and 19th centuries, when swaths of Arab-inhabited lands were under Ottoman rule.

Influence of other languages on Arabic

The most important sources of borrowings into (pre-Islamic) Arabic are from the related (Semitic) languages Aramaic, which used to be the principal, international language of communication throughout the ancient Near and Middle East, and Ethiopic. In addition, many cultural, religious and political terms have entered Arabic from Iranian languages, notably Middle Persian, Parthian, and (Classical) Persian, and Hellenistic Greek (*kīmiyā'* has as origin the Greek *khymia*, meaning in that language the melting of metals; see Roger Dachez, *Histoire de la Médecine de l'Antiquité au XXe siècle*, Tallandier, 2008, p. 251), *alembic* (distiller) from *ambix* (cup),

almanac (climate) from *almenichiakon* (calendar). (For the origin of the last three borrowed words, see Alfred-Louis de Prémare, *Foundations of Islam*, Seuil, L'Univers Historique, 2002.) Some Arabic borrowings from Semitic or Persian languages are, as presented in De Prémare's above-cited book:

- *madīnah*/medina (مدينة, city or city square), a word of Aramaic origin (in which it means "a state")
- *jazīrah* (جزيرة), as in the well-known form الجـــزيرة "Al-Jazeera," means "island" and has its origin in the Syriac ܓܙܝܪܐ *gazīra*.
- *lāzaward* (لازورد) is taken from Persian لاژورد *lājvard*, the name of a blue stone, lapis lazuli. This word was borrowed in several European languages to mean (light) blue – azure in English, *azur* in French and *azul* in Portuguese and Spanish.

A comprehensive overview of the influence of other languages on Arabic is found in Lucas & Manfredi (2020).

Arabic alphabet and nationalism

There have been many instances of national movements to convert Arabic script into Latin script or to Romanize the language. Currently, the only language derived from Classical Arabic to use Latin script is Maltese.

Lebanon

The Beirut newspaper *La Syrie* pushed for the change from Arabic script to Latin letters in 1922. The major head of this

movement was Louis Massignon, a French Orientalist, who brought his concern before the Arabic Language Academy in Damascus in 1928. Massignon's attempt at Romanization failed as the Academy and population viewed the proposal as an attempt from the Western world to take over their country. Sa'id Afghani, a member of the Academy, mentioned that the movement to Romanize the script was a Zionist plan to dominate Lebanon.

Egypt

After the period of colonialism in Egypt, Egyptians were looking for a way to reclaim and re-emphasize Egyptian culture. As a result, some Egyptians pushed for an Egyptianization of the Arabic language in which the formal Arabic and the colloquial Arabic would be combined into one language and the Latin alphabet would be used. There was also the idea of finding a way to use Hieroglyphics instead of the Latin alphabet, but this was seen as too complicated to use. A scholar, Salama Musa agreed with the idea of applying a Latin alphabet to Arabic, as he believed that would allow Egypt to have a closer relationship with the West. He also believed that Latin script was key to the success of Egypt as it would allow for more advances in science and technology. This change in alphabet, he believed, would solve the problems inherent with Arabic, such as a lack of written vowels and difficulties writing foreign words that made it difficult for non-native speakers to learn. Ahmad Lutfi As Sayid and Muhammad Azmi, two Egyptian intellectuals, agreed with Musa and supported the push for Romanization. The idea that Romanization was necessary for modernization and growth in Egypt continued with Abd Al-Aziz Fahmi in 1944. He was the chairman for the

Writing and Grammar Committee for the Arabic Language Academy of Cairo. However, this effort failed as the Egyptian people felt a strong cultural tie to the Arabic alphabet. In particular, the older Egyptian generations believed that the Arabic alphabet had strong connections to Arab values and history, due to the long history of the Arabic alphabet (Shrivtiel, 189) in Muslim societies.

The language of the Quran and its influence on poetry

The Quran introduced a new way of writing to the world. People began studying and applying the unique styles they learned from the Quran to not only their own writing, but also their culture. Writers studied the unique structure and format of the Quran in order to identify and apply the figurative devices and their impact on the reader.

Quran's figurative devices

The Quran inspired musicality in poetry through the internal rhythm of the verses. The arrangement of words, how certain sounds create harmony, and the agreement of rhymes create the sense of rhythm within each verse. At times, the chapters of the Quran only have the rhythm in common.

The repetition in the Quran introduced the true power and impact repetition can have in poetry. The repetition of certain words and phrases made them appear more firm and explicit in

the Quran. The Quran uses constant metaphors of blindness and deafness to imply unbelief. Metaphors were not a new concept to poetry, however the strength of extended metaphors was. The explicit imagery in the Quran inspired many poets to include and focus on the feature in their own work. The poet ibn al-Mu'tazz wrote a book regarding the figures of speech inspired by his study of the Quran. Poets such as badr Shakir al sayyab expresses his political opinion in his work through imagery inspired by the forms of more harsher imagery used in the Quran. The Quran uses figurative devices in order to express the meaning in the most beautiful form possible. The study of the pauses in the Quran as well as other rhetoric allow it to be approached in a multiple ways.

Structure

Although the Quran is known for its fluency and harmony, the structure can be best described as not always being inherently chronological, but can also flow thematically instead (the chapters in the Quran have segments that flow in chronological order, however segments can transition into other segments not related in chronology, but could be related in topic). The suras, also known as chapters of the Quran, are not placed in chronological order. The only constant in their structure is that the longest are placed first and shorter ones follow. The topics discussed in the chapters can also have no direct relation to each other (as seen in many suras) and can share in their sense of rhyme. The Quran introduces to poetry the idea of abandoning order and scattering narratives throughout the text. Harmony is also present in the sound of the Quran. The elongations and accents present in the Quran create a harmonious flow within the writing. Unique sound of the

Quran recited, due to the accents, create a deeper level of understanding through a deeper emotional connection.

The Quran is written in a language that is simple and understandable by people. The simplicity of the writing inspired later poets to write in a more clear and clear-cut style. The words of the Quran, although unchanged, are to this day understandable and frequently used in both formal and informal Arabic. The simplicity of the language makes memorizing and reciting the Quran a slightly easier task.

Culture and the Quran

The writer al-Khattabi explains how culture is a required element to create a sense of art in work as well as understand it. He believes that the fluency and harmony which the Quran possess are not the only elements that make it beautiful and create a bond between the reader and the text. While a lot of poetry was deemed comparable to the Quran in that it is equal to or better than the composition of the Quran, a debate rose that such statements are not possible because humans are incapable of composing work comparable to the Quran. Because the structure of the Quran made it difficult for a clear timeline to be seen, Hadith were the main source of chronological order. The Hadith were passed down from generation to generation and this tradition became a large resource for understanding the context. Poetry after the Quran began possessing this element of tradition by including ambiguity and background information to be required to understand the meaning.

After the Quran came down to the people, the tradition of memorizing the verses became present. It is believed that the greater the amount of the Quran memorized, the greater the faith. As technology improved over time, hearing recitations of the Quran became more available as well as more tools to help memorize the verses. The tradition of Love Poetry served as a symbolic representation of a Muslim's desire for a closer contact with their Lord.

While the influence of the Quran on Arabic poetry is explained and defended by numerous writers, some writers such as Al-Baqillani believe that poetry and the Quran are in no conceivable way related due to the uniqueness of the Quran. Poetry's imperfections prove his points that they cannot be compared with the fluency the Quran holds.

Arabic and Islam

Classical Arabic is the language of poetry and literature (including news); it is also mainly the language of the Quran. Classical Arabic is closely associated with the religion of Islam because the Quran was written in it. Most of the world's Muslims do not speak Classical Arabic as their native language, but many can read the Quranic script and recite the Quran. Among non-Arab Muslims, translations of the Quran are most often accompanied by the original text. At present, Modern Standard Arabic (MSA) is also used in modernized versions of literary forms of the Quran.

Some Muslims present a monogenesis of languages and claim that the Arabic language was the language revealed by God for the benefit of mankind and the original language as a

prototype system of symbolic communication, based upon its system of triconsonantal roots, spoken by man from which all other languages were derived, having first been corrupted. Judaism has a similar account with the Tower of Babel.

Dialects and descendants

Colloquial Arabic is a collective term for the spoken dialects of Arabic used throughout the Arab world, which differ radically from the literary language. The main dialectal division is between the varieties within and outside of the Arabian peninsula, followed by that between sedentary varieties and the much more conservative Bedouin varieties. All the varieties outside of the Arabian peninsula (which include the large majority of speakers) have many features in common with each other that are not found in Classical Arabic. This has led researchers to postulate the existence of a prestige koine dialect in the one or two centuries immediately following the Arab conquest, whose features eventually spread to all newly conquered areas. These features are present to varying degrees inside the Arabian peninsula. Generally, the Arabian peninsula varieties have much more diversity than the non-peninsula varieties, but these have been understudied.

Within the non-peninsula varieties, the largest difference is between the non-Egyptian North African dialects (especially Moroccan Arabic) and the others. Moroccan Arabic in particular is hardly comprehensible to Arabic speakers east of Libya (although the converse is not true, in part due to the popularity of Egyptian films and other media).

One factor in the differentiation of the dialects is influence from the languages previously spoken in the areas, which have typically provided a significant number of new words and have sometimes also influenced pronunciation or word order; however, a much more significant factor for most dialects is, as among Romance languages, retention (or change of meaning) of different classical forms. Thus Iraqi *aku*, Levantine *fīh* and North African *kayən* all mean 'there is', and all come from Classical Arabic forms (*yakūn*, *fīhi*, *kā'in* respectively), but now sound very different.

Koiné

According to Charles A. Ferguson, the following are some of the characteristic features of the koiné that underlies all the modern dialects outside the Arabian peninsula. Although many other features are common to most or all of these varieties, Ferguson believes that these features in particular are unlikely to have evolved independently more than once or twice and together suggest the existence of the koine:

- Loss of the dual number except on nouns, with consistent plural agreement (cf. feminine singular agreement in plural inanimates).
- Change of *a* to *i* in many affixes (e.g., non-past-tense prefixes *ti- yi- ni-*; *wi-* 'and'; *il-* 'the'; feminine *-it* in the construct state).
- Loss of third-weak verbs ending in *w* (which merge with verbs ending in *y*).
- Reformation of geminate verbs, e.g., *ḥalaltu* 'I untied' →*ḥalēt(u)*.

- Conversion of separate words *lī* 'to me', *laka* 'to you', etc. into indirect-object clitic suffixes.
- Certain changes in the cardinal number system, e.g., *khamsat ayyām* 'five days' →*kham(a)s tiyyām*, where certain words have a special plural with prefixed *t*.
- Loss of the feminine elative (comparative).
- Adjective plurals of the form *kibār* 'big' →*kubār*.
- Change of nisba suffix *-iyy>i*.
- Certain lexical items, e.g., *jāb* 'bring' *<jāʼ a bi-* 'come with'; *shāf* 'see'; *ēsh* 'what' (or similar) *<ayyu shayʼ* 'which thing';*illi* (relative pronoun).
- Merger of /ʒˁ/ and /ðˁ/.

Dialect groups

- Egyptian Arabic is spoken by around 53 million people in Egypt (55 million worldwide). It is one of the most understood varieties of Arabic, due in large part to the widespread distribution of Egyptian films and television shows throughout the Arabic-speaking world
- Levantine Arabic includes North Levantine Arabic, South Levantine Arabic and Cypriot Arabic. It is spoken by about 21 million people in Lebanon, Syria, Jordan, Palestine, Israel, Cyprus and Turkey.
- Lebanese Arabic is a variety of Levantine Arabic spoken primarily in Lebanon.
- Jordanian Arabic is a continuum of mutually intelligible varieties of Levantine Arabic spoken by the population of the Kingdom of Jordan.

- Palestinian Arabic is a name of several dialects of the subgroup of Levantine Arabic spoken by the Palestinians in Palestine, by Arab citizens of Israel and in most Palestinian populations around the world.

- Samaritan Arabic, spoken by only several hundred in the Nablus region

- Cypriot Maronite Arabic, spoken in Cyprus

- Maghrebi Arabic, also called "Darija" spoken by about 70 million people in Morocco, Algeria, Tunisia and Libya. It also forms the basis of Maltese via the extinct Sicilian Arabic dialect. Maghrebi Arabic is very hard to understand for Arabic speakers from the Mashriq or Mesopotamia, the most comprehensible being Libyan Arabic and the most difficult Moroccan Arabic. The others such as Algerian Arabic can be considered in between the two in terms of difficulty.

- Libyan Arabic spoken in Libya and neighboring countries.

- Tunisian Arabic spoken in Tunisia and North-eastern Algeria

- Algerian Arabic spoken in Algeria

- Judeo-Algerian Arabic was spoken by Jews in Algeria until 1962

- Moroccan Arabic spoken in Morocco

- Hassaniya Arabic (3 million speakers), spoken in Mauritania, Western Sahara, some parts of the Azawad in northern Mali, southern Morocco and south-western Algeria.

- Andalusian Arabic, spoken in Spain until the 16th century.

- Siculo-Arabic (Sicilian Arabic), was spoken in Sicily and Malta between the end of the 9th century and the end of the 12th century and eventually evolved into the Maltese language.

- Maltese, spoken on the island of Malta, is the only fully separate standardized language to have originated from an Arabic dialect (the extinct Siculo-Arabic dialect), with independent literary norms. Maltese has evolved independently of Literary Arabic and its varieties into a standardized language over the past 800 years in a gradual process of Latinisation. Maltese is therefore considered an exceptional descendant of Arabic that has no diglossic relationship with Standard Arabic or Classical Arabic. Maltese is also different from Arabic and other Semitic languages since its morphology has been deeply influenced by Romance languages, Italian and Sicilian. It is also the only Semitic language written in the Latin script. In terms of basic everyday language, speakers of Maltese are reported to be able to understand less than a third of what is said to them in Tunisian Arabic, which is related to Siculo-Arabic, whereas speakers of Tunisian are able to understand about 40% of what is said to them in Maltese. This asymmetric intelligibility is considerably lower than the mutual intelligibility found between Maghrebi Arabic dialects. Maltese has its own dialects, with urban varieties of Maltese being closer to Standard Maltese than rural varieties.

- Mesopotamian Arabic, spoken by about 41.2 million people in Iraq (where it is called "Aamiyah"), eastern

Syria and southwestern Iran (Khuzestan) and in the southeastern of Turkey (in the eastern Mediterranean, Southeastern Anatolia Region)

- North Mesopotamian Arabic is a spoken north of the Hamrin Mountains in Iraq, in western Iran, northern Syria, and in southeastern Turkey (in the eastern Mediterranean Region, Southeastern Anatolia Region, and southern Eastern Anatolia Region).

- Judeo-Mesopotamian Arabic, also known as Iraqi Judeo Arabic and Yahudic, is a variety of Arabic spoken by Iraqi Jews of Mosul.

- Baghdad Arabic is the Arabic dialect spoken in Baghdad, and the surrounding cities and it is a subvariety of Mesopotamian Arabic.

- Baghdad Jewish Arabic is the dialect spoken by the Iraqi Jews of Baghdad.

- South Mesopotamian Arabic (Basrawi dialect) is the dialect spoken in southern Iraq, such as Basra, Dhi Qar and Najaf.

- Khuzestani Arabic is the dialect spoken in the Iranian province of Khuzestan. This dialect is a mix of Southen Mesopotamian Arabic and Gulf Arabic.

- Khorasani Arabic spoken in the Iranian province of Khorasan.

- Kuwaiti Arabic is a Gulf Arabic dialect spoken in Kuwait.

- Sudanese Arabic is spoken by 17 million people in Sudan and some parts of southern Egypt. Sudanese Arabic is quite distinct from the dialect of its neighbor to the north; rather, the Sudanese have a dialect similar to the Hejazi dialect.

- Juba Arabic spoken in South Sudan and southern Sudan
- Gulf Arabic, spoken by around four million people, predominantly in Kuwait, Bahrain, some parts of Oman, eastern Saudi Arabia coastal areas and some parts of UAE and Qatar. Also spoken in Iran's Bushehr and Hormozgan provinces. Although Gulf Arabic is spoken in Qatar, most Qatari citizens speak Najdi Arabic (Bedawi).
- Omani Arabic, distinct from the Gulf Arabic of Eastern Arabia and Bahrain, spoken in Central Oman. With recent oil wealth and mobility has spread over other parts of the Sultanate.
- Hadhrami Arabic, spoken by around 8 million people, predominantly in Hadhramaut, and in parts of the Arabian Peninsula, South and Southeast Asia, and East Africa by Hadhrami descendants.
- Yemeni Arabic spoken in Yemen, and southern Saudi Arabia by 15 million people. Similar to Gulf Arabic.
- Najdi Arabic, spoken by around 10 million people, mainly spoken in Najd, central and northern Saudi Arabia. Most Qatari citizens speak Najdi Arabic (Bedawi).
- Hejazi Arabic (6 million speakers), spoken in Hejaz, western Saudi Arabia
- Saharan Arabic spoken in some parts of Algeria, Niger and Mali
- Baharna Arabic (600,000 speakers), spoken by Bahrani Shi'ah in Bahrain and Qatif, the dialect exhibits many big differences from Gulf Arabic. It is also spoken to a lesser extent in Oman.

- Judeo-Arabic dialects – these are the dialects spoken by the Jews that had lived or continue to live in the Arab World. As Jewish migration to Israel took hold, the language did not thrive and is now considered endangered. So-called Qəltu Arabic.
- Chadian Arabic, spoken in Chad, Sudan, some parts of South Sudan, Central African Republic, Niger, Nigeria, Cameroon
- Central Asian Arabic, spoken in Uzbekistan, Tajikistan and Afghanistan, is highly endangered
- Shirvani Arabic, spoken in Azerbaijan and Dagestan until the 1930s, now extinct.

Phonology

History

Of the 29 Proto-Semitic consonants, only one has been lost: */ʃ/, which merged with /s/, while /ɬ/ became /ʃ/ (see Semitic languages). Various other consonants have changed their sound too, but have remained distinct. An original */p/ lenited to /f/, and */ɡ/ – consistently attested in pre-Islamic Greek transcription of Arabic languages – became palatalized to /ɡʲ/ or /ɟ/ by the time of the Quran and /dʒ/, /ɡ/, /ʒ/ or /ɟ/ after early Muslim conquests and in MSA (see Arabic phonology#Local variations for more detail). An original voiceless alveolar lateral fricative */ɬ/ became /ʃ/. Its emphatic counterpart /ɬˠ~ɮˤ/ was considered by Arabs to be the most unusual sound in Arabic (Hence the Classical Arabic's appellation ٱلضّاد لغة *lughat al-ḍād* or "language of the

ḍād"); for most modern dialects, it has become an emphatic stop /dˤ/ with loss of the laterality or with complete loss of any pharyngealization or velarization, /d/. (The classical *ḍād* pronunciation of pharyngealization /ɮˤ/ still occurs in the Mehri language, and the similar sound without velarization, /ɮ/, exists in other Modern South Arabian languages.)

Other changes may also have happened. Classical Arabic pronunciation is not thoroughly recorded and different reconstructions of the sound system of Proto-Semitic propose different phonetic values. One example is the emphatic consonants, which are pharyngealized in modern pronunciations but may have been velarized in the eighth century and glottalized in Proto-Semitic.

Reduction of /j/ and /w/ between vowels occurs in a number of circumstances and is responsible for much of the complexity of third-weak ("defective") verbs. Early Akkadian transcriptions of Arabic names shows that this reduction had not yet occurred as of the early part of the 1st millennium BC.

The Classical Arabic language as recorded was a poetic koine that reflected a consciously archaizing dialect, chosen based on the tribes of the western part of the Arabian Peninsula, who spoke the most conservative variants of Arabic. Even at the time of Muhammed and before, other dialects existed with many more changes, including the loss of most glottal stops, the loss of case endings, the reduction of the diphthongs /aj/ and /aw/ into monophthongs /eː , oː /, etc. Most of these changes are present in most or all modern varieties of Arabic.

An interesting feature of the writing system of the Quran (and hence of Classical Arabic) is that it contains certain features of Muhammad's native dialect of Mecca, corrected through diacritics into the forms of standard Classical Arabic. Among these features visible under the corrections are the loss of the glottal stop and a differing development of the reduction of certain final sequences containing /j/: Evidently, final /-awa/ became /aː/ as in the Classical language, but final /-aja/ became a different sound, possibly /eː/ (rather than again /aː/ in the Classical language). This is the apparent source of the *alif maqṣūrah* 'restricted alif' where a final /-aja/is reconstructed: a letter that would normally indicate /j/ or some similar high-vowel sound, but is taken in this context to be a logical variant of *alif* and represent the sound /aː/.

Although Classical Arabic was a unitary language and is now used in Quran, its pronunciation varies somewhat from country to country and from region to region within a country. It is influenced by colloquial dialects.

Literary Arabic

The "colloquial" spoken dialects of Arabic are learned at home and constitute the native languages of Arabic speakers. "Formal" Literary Arabic (usually specifically Modern Standard Arabic) is learned at school; although many speakers have a native-like command of the language, it is technically not the native language of any speakers. Both varieties can be both written and spoken, although the colloquial varieties are rarely written down and the formal variety is spoken mostly in formal circumstances, e.g., in radio and TV broadcasts, formal

lectures, parliamentary discussions and to some extent between speakers of different colloquial dialects. Even when the literary language is spoken, however, it is normally only spoken in its pure form when reading a prepared text out loud and communication between speakers of different colloquial dialects. When speaking extemporaneously (i.e. making up the language on the spot, as in a normal discussion among people), speakers tend to deviate somewhat from the strict literary language in the direction of the colloquial varieties. In fact, there is a continuous range of "in-between" spoken varieties: from nearly pure Modern Standard Arabic (MSA), to a form that still uses MSA grammar and vocabulary but with significant colloquial influence, to a form of the colloquial language that imports a number of words and grammatical constructions in MSA, to a form that is close to pure colloquial but with the "rough edges" (the most noticeably "vulgar" or non-Classical aspects) smoothed out, to pure colloquial. The particular variant (or *register*) used depends on the social class and education level of the speakers involved and the level of formality of the speech situation. Often it will vary within a single encounter, e.g., moving from nearly pure MSA to a more mixed language in the process of a radio interview, as the interviewee becomes more comfortable with the interviewer. This type of variation is characteristic of the diglossia that exists throughout the Arabic-speaking world.

Although Modern Standard Arabic (MSA) is a unitary language, its pronunciation varies somewhat from country to country and from region to region within a country. The variation in individual "accents" of MSA speakers tends to mirror corresponding variations in the colloquial speech of the speakers in question, but with the distinguishing

characteristics moderated somewhat. It is important in descriptions of "Arabic" phonology to distinguish between pronunciation of a given colloquial (spoken) dialect and the pronunciation of MSA by these same speakers. Although they are related, they are not the same. For example, the phoneme that derives from Classical Arabic /ɟ/ has many different pronunciations in the modern spoken varieties, e.g., [d͡ʒ ~ ʒ ~ j ~ gʲ ~ g] including the proposed original [ɟ]. Speakers whose native variety has either [d͡ʒ] or [ʒ] will use the same pronunciation when speaking MSA. Even speakers from Cairo, whose native Egyptian Arabic has [g], normally use [g] when speaking MSA. The [j] of Persian Gulf speakers is the only variant pronunciation which isn't found in MSA; [d͡ʒ~ʒ] is used instead, but may use [j] in MSA for comfortable pronunciation. Another reason of different pronunciations is influence of colloquial dialects. The differentiation of pronunciation of colloquial dialects is the influence from other languages previously spoken and some still presently spoken in the regions, such as Coptic in Egypt, Berber, Punic, or Phoenician in North Africa, Himyaritic, Modern South Arabian, and Old South Arabian in Yemen and Oman, and Aramaic and Canaanite languages (including Phoenician) in the Levant and Mesopotamia.

Another example: Many colloquial varieties are known for a type of vowel harmony in which the presence of an "emphatic consonant" triggers backed allophones of nearby vowels (especially of the low vowels /aː /, which are backed to [ɑ(ː)] in these circumstances and very often fronted to [æ(ː)] in all other circumstances). In many spoken varieties, the backed or "emphatic" vowel allophones spread a fair distance in both directions from the triggering consonant; in some varieties

(most notably Egyptian Arabic), the "emphatic" allophones spread throughout the entire word, usually including prefixes and suffixes, even at a distance of several syllables from the triggering consonant. Speakers of colloquial varieties with this vowel harmony tend to introduce it into their MSA pronunciation as well, but usually with a lesser degree of spreading than in the colloquial varieties. (For example, speakers of colloquial varieties with extremely long-distance harmony may allow a moderate, but not extreme, amount of spreading of the harmonic allophones in their MSA speech, while speakers of colloquial varieties with moderate-distance harmony may only harmonize immediately adjacent vowels in MSA.)

Vowels

Modern Standard Arabic has six pure vowels (while most modern dialects have eight pure vowels which includes the long vowels /eː oː /), with short /a i u/ and corresponding long vowels /aː iː uː /. There are also two diphthongs: /aj/ and /aw/.

The pronunciation of the vowels differs from speaker to speaker, in a way that tends to reflect the pronunciation of the corresponding colloquial variety. Nonetheless, there are some common trends. Most noticeable is the differing pronunciation of /a/ and /aː/, which tend towards fronted [æ(ː)], [a(ː)] or [ɛ(ː)] in most situations, but a back [ɑ(ː)] in the neighborhood of emphatic consonants. Some accents and dialects, such as those of the Hejaz region, have an open [a(ː)] or a central [ä(ː)] in all situations. The vowel /a/ varies towards [ə(ː)] too. Listen

to the final vowel in the recording of *al-ʿarabiyyah* at the beginning of this article, for example. The point is, Arabic has only three short vowel phonemes, so those phonemes can have a very wide range of allophones. The vowels /u/ and /ɪ/are often affected somewhat in emphatic neighborhoods as well, with generally more back or centralized allophones, but the differences are less great than for the low vowels. The pronunciation of short /u/ and /i/ tends towards [ʊ~o] and [i~e~ɨ], respectively, in many dialects.

The definition of both "emphatic" and "neighborhood" vary in ways that reflect (to some extent) corresponding variations in the spoken dialects. Generally, the consonants triggering "emphatic" allophones are the pharyngealized consonants /tˤ dˤ sˤðˤ/; /q/; and /r/, if not followed immediately by /i(ː)/. Frequently, the velar fricatives /x ɣ/ also trigger emphatic allophones; occasionally also the pharyngeal consonants /ʕħ/ (the former more than the latter). Many dialects have multiple emphatic allophones of each vowel, depending on the particular nearby consonants. In most MSA accents, emphatic coloring of vowels is limited to vowels immediately adjacent to a triggering consonant, although in some it spreads a bit farther: e.g., وقت *waqt*[wɑqt] 'time'; وطن *waṭan*[watˤɑn] 'homeland'; وسط المدينة *wasṭ al-madīnah*[wæstˤɑl mædiːnɐ] 'downtown' (sometimes [wɑstˤɑl mædiːnæ] or similar).

In a non-emphatic environment, the vowel /a/ in the diphthong /aj/ tends to be fronted even more than elsewhere, often pronounced[æj] or [ɛj]: hence سيف *sayf*[sajf ~ sæjf ~ sɛjf] 'sword' but صيف *ṣayf*[sˤɑjf] 'summer'. However, in accents with

no emphatic allophones of /a/ (e.g., in the Hejaz), the pronunciation [aj] or [äj] occurs in all situations.

The phoneme /ʤ/ is represented by the Arabic letter *jīm* (ج) and has many standard pronunciations. [ʤ] is characteristic of north Algeria, Iraq, and most of the Arabian peninsula but with an allophonic [ʒ] in some positions; [ʒ] occurs in most of the Levant and most of North Africa; and [g] is used in most of Egypt and some regions in Yemen and Oman. Generally this corresponds with the pronunciation in the colloquial dialects. In some regions in Sudan and Yemen, as well as in some Sudanese and Yemeni dialects, it may be either [gʲ] or [ɟ], representing the original pronunciation of Classical Arabic. Foreign words containing /g/may be transcribed with ج , غ , ك, ق , گ, ڭ or ڤ, mainly depending on the regional spoken variety of Arabic or the commonly diacriticized Arabic letter. In northern Egypt, where the Arabic letter *jīm* (ج) is normally pronounced [g], a separate phoneme /ʒ/, which may be transcribed with چ , occurs in a small number of mostly non-Arabic loanwords, e.g., /ʒakitta/ 'jacket'.

/θ/ (ث) can be pronounced as [s]. In some places of Maghreb it can be also pronounced as [ʦ].

/x/ and /ɣ/ (خ,غ) are velar, post-velar, or uvular.

In many varieties, /ħ, ʕ/ (ح,ع) are epiglottal[ʜ, ʢ] in Western Asia.

/l/is pronounced as velarized [ɫ] in الله /ʔallaːh/, the name of God, q.e. Allah, when the word follows *a*, *ā*, *u* or *ū* (after *i* or *ī* it

is unvelarized: بسم الله *bismi l–lāh*/bismillaːh/). Some speakers velarize other occurrences of /l/ in MSA, in imitation of their spoken dialects.

The emphatic consonant /dˤ/ was actually pronounced [ɮˤ], or possibly [dɮˤ]—either way, a highly unusual sound. The medieval Arabs actually termed their language *lughat al-ḍād* 'the language of the Ḍād' (the name of the letter used for this sound), since they thought the sound was unique to their language. (In fact, it also exists in a few other minority Semitic languages, e.g., Mehri.)

Arabic has consonants traditionally termed "emphatic" /tˤ, dˤ, sˤ, ðˤ/ (ظ , ص ,ط,ض), which exhibit simultaneous pharyngealization[tˤ, dˤ, sˤ, ðˤ] as well as varying degrees of velarization[tˠ, dˠ, sˠ, ðˠ] (depending on the region), so they may be written with the "Velarized or pharyngealized" diacritic (̴) as: /ł, đ, s̴, ð̴/. This simultaneous articulation is described as "Retracted Tongue Root" by phonologists. In some transcription systems, emphasis is shown by capitalizing the letter, for example, /dˤ/ is written ⟨D⟩; in others the letter is underlined or has a dot below it, for example, ⟨ḏ⟩.

Vowels and consonants can be phonologically short or long. Long (geminate) consonants are normally written doubled in Latin transcription (i.e. bb, dd, etc.), reflecting the presence of the Arabic diacritic mark *shaddah*, which indicates doubled consonants. In actual pronunciation, doubled consonants are held twice as long as short consonants. This consonant lengthening is phonemically contrastive: قبل *qabila* 'he accepted' vs. قبّل *qabbala* 'he kissed'.

Syllable structure

Arabic has two kinds of syllables: open syllables (CV) and (CVV)—and closed syllables (CVC), (CVVC) and (CVCC). The syllable types with two morae (units of time), i.e. CVC and CVV, are termed *heavy syllables*, while those with three morae, i.e. CVVC and CVCC, are *superheavy syllables*. Superheavy syllables in Classical Arabic occur in only two places: at the end of the sentence (due to pausal pronunciation) and in words such as حارّ *ḥārr* 'hot', مادّة *māddah* 'stuff, substance', تحاجّوا *taḥājjū* 'they disputed with each other', where a long *ā* occurs before two identical consonants (a former short vowel between the consonants has been lost). (In less formal pronunciations of Modern Standard Arabic, superheavy syllables are common at the end of words or before clitic suffixes such as *-nā* 'us, our', due to the deletion of final short vowels.)

In surface pronunciation, every vowel must be preceded by a consonant (which may include the glottal stop[ʔ]). There are no cases of hiatus within a word (where two vowels occur next to each other, without an intervening consonant). Some words do have an underlying vowel at the beginning, such as the definite article *al-* or words such as اشترا *ishtarā* 'he bought', اجتماع *ijtimāʿ* 'meeting'. When actually pronounced, one of three things happens:

- If the word occurs after another word ending in a consonant, there is a smooth transition from final consonant to initial vowel, e.g., الاجتماع *al-ijtimāʿ* 'meeting' /alidʒtimaːʕ/.

- If the word occurs after another word ending in a vowel, the initial vowel of the word is elided, e.g., بيت المدير *baytu (a)l-mudīr* 'house of the director' /bajtulmudiː r/.
- If the word occurs at the beginning of an utterance, a glottal stop [ʔ]is added onto the beginning, e.g., البيت هو *al-baytu huwa ...* 'The house is ...' /ʔalbajtuhuwa ... /.

Stress

Word stress is not phonemically contrastive in Standard Arabic. It bears a strong relationship to vowel length. The basic rules for Modern Standard Arabic are:

- A final vowel, long or short, may not be stressed.
- Only one of the last three syllables may be stressed.
- Given this restriction, the last heavy syllable (containing a long vowel or ending in a consonant) is stressed, if it is not the final syllable.
- If the final syllable is super heavy and closed (of the form CVVC or CVCC) it receives stress.
- If no syllable is heavy or super heavy, the first possible syllable (i.e. third from end) is stressed.
- As a special exception, in Form VII and VIII verb forms stress may not be on the first syllable, despite the above rules: Hence *inkatab(a)* 'he subscribed' (whether or not the final short vowel is pronounced), *yankatib(u)* 'he subscribes' (whether or not the final short vowel is pronounced), *yankatib* 'he should

subscribe (juss.)'. Likewise Form VIII *ishtarā* 'he bought', *yashtarī* 'he buys'.

Examples: *kitāb(un)* 'book', *kā-ti-b(un)* 'writer', *mak-ta-b(un)* 'desk', *ma-kā-ti-b(u)* 'desks', *mak-ta-ba-tun* 'library' (but *mak-ta-ba(-tun)* 'library' in short pronunciation), *ka-ta-bū* (Modern Standard Arabic) 'they wrote' = *ka-ta-bu* (dialect), *ka-ta-bū-h(u)* (Modern Standard Arabic) 'they wrote it' = *ka-ta-bū* (dialect), *ka-ta-ba-tā* (Modern Standard Arabic) 'they (dual, fem) wrote', *ka-tab-tu* (Modern Standard Arabic) 'I wrote' = *ka-tabt* (short form or dialect). Doubled consonants count as two consonants: *ma-jal-la-(tan)* 'magazine', *ma-ḥall(-un)* "place".

These rules may result in differently stressed syllables when final case endings are pronounced, vs. the normal situation where they are not pronounced, as in the above example of *mak-ta-ba-tun* 'library' in full pronunciation, but *mak-ta-ba(-tun)* 'library' in short pronunciation.

The restriction on final long vowels does not apply to the spoken dialects, where original final long vowels have been shortened and secondary final long vowels have arisen from loss of original final *-hu/hi*.

Some dialects have different stress rules. In the Cairo (Egyptian Arabic) dialect a heavy syllable may not carry stress more than two syllables from the end of a word, hence *mad-ra-sah* 'school', *qā-hi-rah* 'Cairo'. This also affects the way that Modern Standard Arabic is pronounced in Egypt. In the Arabic of Sanaa, stress is often retracted: *bay-tayn* 'two houses', *mā-sat-hum* 'their table', *ma-kā-tīb* 'desks', *zā-rat-ḥīn* 'sometimes', *mad-ra-sat-hum* 'their school'. (In this dialect, only syllables

with long vowels or diphthongs are considered heavy; in a two-syllable word, the final syllable can be stressed only if the preceding syllable is light; and in longer words, the final syllable cannot be stressed.)

Levels of pronunciation

The final short vowels (e.g., the case endings *-a -i -u* and mood endings *-u -a*) are often not pronounced in this language, despite forming part of the formal paradigm of nouns and verbs. The following levels of pronunciation exist:

Full pronunciation with pausa

This is the most formal level actually used in speech. All endings are pronounced as written, except at the end of an utterance, where the following changes occur:

- Final short vowels are not pronounced. (But possibly an exception is made for feminine plural *-na* and shortened vowels in the jussive/imperative of defective verbs, e.g., *irmi!* 'throw!'".)
- The entire indefinite noun endings *-in* and *-un* (with nunation) are left off. The ending *-an* is left off of nouns preceded by a *tāʾ marbūṭah* (i.e. the *-t* in the ending *-at-* that typically marks feminine nouns), but pronounced as *-ā* in other nouns (hence its writing in this fashion in the Arabic script).
- The *tāʾ marbūṭah* itself (typically of feminine nouns) is pronounced as *h*. (At least, this is the case in

104

extremely formal pronunciation, e.g., some Quranic recitations. In practice, this *h* is usually omitted.)

Formal short pronunciation

This is a formal level of pronunciation sometimes seen. It is somewhat like pronouncing all words as if they were in pausal position (with influence from the colloquial varieties). The following changes occur:

- Most final short vowels are not pronounced. However, the following short vowels *are* pronounced:
- feminine plural *-na*
- shortened vowels in the jussive/imperative of defective verbs, e.g., *irmi!* 'throw!'
- second-person singular feminine past-tense *-ti* and likewise *anti* 'you (fem. sg.)'
- sometimes, first-person singular past-tense *-tu*
- sometimes, second-person masculine past-tense *-ta* and likewise *anta* 'you (masc. sg.)'
- final *-a* in certain short words, e.g., *laysa* 'is not', *sawfa* (future-tense marker)
- The nunation endings *-an* *-in* *-un* are not pronounced. However, they *are* pronounced in adverbial accusative formations, e.g., *taqrīban*تَقْرِيبًا 'almost, approximately', *'ādatan*عَادَةً 'usually'.
- The *tāʾ marbūṭah* ending ة is unpronounced, *except* in construct state nouns, where it sounds as *t* (and in adverbial accusative constructions, e.g., *'ādatan*عَادَةً 'usually', where the entire *-tan* is pronounced).

- The masculine singular nisbah ending *-iyy* is actually pronounced *-ī* and is unstressed (but plural and feminine singular forms, i.e. when followed by a suffix, still sound as *-iyy-*).
- *Full endings* (including case endings) occur when a clitic object or possessive suffix is added (e.g., *-nā* 'us/our').

Informal short pronunciation

This is the pronunciation used by speakers of Modern Standard Arabic in extemporaneous speech, i.e. when producing new sentences rather than simply reading a prepared text. It is similar to formal short pronunciation except that the rules for dropping final vowels apply *even* when a clitic suffix is added. Basically, short-vowel case and mood endings are never pronounced and certain other changes occur that echo the corresponding colloquial pronunciations. Specifically:

- All the rules for formal short pronunciation apply, except as follows.
- The past tense singular endings written formally as *-tu -ta -ti* are pronounced *-t -t -ti.* But masculine *ʾanta* is pronounced in full.
- Unlike in formal short pronunciation, the rules for dropping or modifying final endings are also applied when a clitic object or possessive suffix is added (e.g., *-nā* 'us/our'). If this produces a sequence of three consonants, then one of the following happens, depending on the speaker's native colloquial variety:

- A short vowel (e.g., *-i-* or *-ⵣ-*) is consistently added, either between the second and third or the first and second consonants.

- Or, a short vowel is added only if an otherwise unpronounceable sequence occurs, typically due to a violation of the sonority hierarchy (e.g., *-rtn-* is pronounced as a three-consonant cluster, but *-trn-* needs to be broken up).

- Or, a short vowel is never added, but consonants like *r l m n* occurring between two other consonants will be pronounced as a syllabic consonant (as in the English words "butter bottle bottom button").

- When a doubled consonant occurs before another consonant (or finally), it is often shortened to a single consonant rather than a vowel added. (However, Moroccan Arabic never shortens doubled consonants or inserts short vowels to break up clusters, instead tolerating arbitrary-length series of arbitrary consonants and hence Moroccan Arabic speakers are likely to follow the same rules in their pronunciation of Modern Standard Arabic.)

- The clitic suffixes themselves tend also to be changed, in a way that avoids many possible occurrences of three-consonant clusters. In particular, *-ka -ki -hu* generally sound as *-ak -ik -uh*.

- Final long vowels are often shortened, merging with any short vowels that remain.

- Depending on the level of formality, the speaker's education level, etc., various grammatical changes may occur in ways that echo the colloquial variants:

- Any remaining case endings (e.g. masculine plural nominative *-ūn* vs. oblique *-īn*) will be leveled, with

the oblique form used everywhere. (However, in words like *ab* 'father' and *akh* 'brother' with special long-vowel case endings in the construct state, the nominative is used everywhere, hence *abū* 'father of', *akhū* 'brother of'.)

- Feminine plural endings in verbs and clitic suffixes will often drop out, with the masculine plural endings used instead. If the speaker's native variety has feminine plural endings, they may be preserved, but will often be modified in the direction of the forms used in the speaker's native variety, e.g. *-an* instead of *-na.*

- Dual endings will often drop out except on nouns and then used only for emphasis (similar to their use in the colloquial varieties); elsewhere, the plural endings are used (or feminine singular, if appropriate).

As mentioned above, many spoken dialects have a process of *emphasis spreading*, where the "emphasis" (pharyngealization) of emphatic consonants spreads forward and back through adjacent syllables, pharyngealizing all nearby consonants and triggering the back allophone [ɑ(ː)] in all nearby low vowels. The extent of emphasis spreading varies. For example, in Moroccan Arabic, it spreads as far as the first full vowel (i.e. sound derived from a long vowel or diphthong) on either side; in many Levantine dialects, it spreads indefinitely, but is blocked by any /j/ or /ʃ/; while in Egyptian Arabic, it usually spreads throughout the entire word, including prefixes and suffixes. In Moroccan Arabic, /i u/ also have emphatic allophones [e~ɛ] and [o~ɔ], respectively.

Unstressed short vowels, especially /i u/, are deleted in many contexts. Many sporadic examples of short vowel change have occurred (especially /a/→/i/ and interchange /i/↔/u/). Most Levantine dialects merge short /i u/ into /ə/ in most contexts (all except directly before a single final consonant). In Moroccan Arabic, on the other hand, short /u/ triggers labialization of nearby consonants (especially velar consonants and uvular consonants), and then short /a i u/ all merge into /ə/, which is deleted in many contexts. (The labialization plus /ə/is sometimes interpreted as an underlying phoneme /ŭ/.) This essentially causes the wholesale loss of the short-long vowel distinction, with the original long vowels /aː iː uː / remaining as half-long [aˑ iˑ uˑ], phonemically /a i u/, which are used to represent *both* short and long vowels in borrowings from Literary Arabic.

Most spoken dialects have monophthongized original /aj aw/ to /eː oː/ in most circumstances, including adjacent to emphatic consonants, while keeping them as the original diphthongs in others e.g. مَوْعِد/m**aw**ʕid/. In most of the Moroccan, Algerian and Tunisian (except Sahel and Southeastern) Arabic dialects, they have subsequently merged into original /iː uː/.

Consonants

In most dialects, there may be more or fewer phonemes than those listed in the chart above. For example, [g]is considered a native phoneme in most Arabic dialects except in Levantine dialects like Syrian or Lebanese where ج is pronounced [ʒ] and ق is pronounced [ʔ]. [d͡ʒ] or [ʒ] (ج) is considered a native

phoneme in most dialects except in Egyptian and a number of Yemeni and Omani dialects where ج is pronounced [g]. [zˤ] or [ðˤ] and [dˤ] are distinguished in the dialects of Egypt, Sudan, the Levant and the Hejaz, but they have merged as [ðˤ] in most dialects of the Arabian Peninsula, Iraq and Tunisia and have merged as [dˤ] in Morocco and Algeria. The usage of non-native [p]پ and [v]ڤ depends on the usage of each speaker but they might be more prevalent in some dialects than others. The Iraqi and Gulf Arabic also has the sound [t͡ʃ] and writes it and [g] with the Persian letters چ and گ, as in گوجة*gawjah* "plum";چمة*chimah* "truffle".

Early in the expansion of Arabic, the separate emphatic phonemes [ɮˤ] and [ðˤ] coalesced into a single phoneme [ðˤ]. Many dialects (such as Egyptian, Levantine, and much of the Maghreb) subsequently lost interdentalfricatives, converting [θ ð ðˤ] into [t d dˤ]. Most dialects borrow "learned" words from the Standard language using the same pronunciation as for inherited words, but some dialects without interdental fricatives (particularly in Egypt and the Levant) render original [θ ð ðˤ dˤ] in borrowed words as [s z zˤ dˤ].

Another key distinguishing mark of Arabic dialects is how they render the original velar and uvular plosives /q/, /d͡ʒ/ (Proto-Semitic /g/), and /k/:

- ق/q/ retains its original pronunciation in widely scattered regions such as Yemen, Morocco, and urban areas of the Maghreb. It is pronounced as a glottal stop[ʔ] in several prestige dialects, such as those spoken in Cairo, Beirut and Damascus. But it

is rendered as a voiced velar plosive [g] in Persian Gulf, Upper Egypt, parts of the Maghreb, and less urban parts of the Levant (e.g. Jordan). In Iraqi Arabic it sometimes retains its original pronunciation and is sometimes rendered as a voiced velar plosive, depending on the word. Some traditionally Christian villages in rural areas of the Levant render the sound as [k], as do Shi'i Bahrainis. In some Gulf dialects, it is palatalized to [ʤ] or [ʒ]. It is pronounced as a voiced uvular constrictive [ʁ] in Sudanese Arabic. Many dialects with a modified pronunciation for /q/ maintain the [q] pronunciation in certain words (often with religious or educational overtones) borrowed from the Classical language.

- ج /ʤ/ is pronounced as an affricate in Iraq and much of the Arabian Peninsula but is pronounced [g] in most of North Egypt and parts of Yemen and Oman, [ʒ] in Morocco, Tunisia, and the Levant, and [j], [i̠] in most words in much of the Persian Gulf.

- ك /k/ usually retains its original pronunciation but is palatalized to /t͡ʃ/ in many words in Israel and the Palestinian Territories, Iraq, and countries in the eastern part of the Arabian Peninsula. Often a distinction is made between the suffixes /-ak/ ('you', masc.) and /-ik/ ('you', fem.), which become /-ak/ and /-it͡ʃ/, respectively. In Sana'a, Omani, and Bahrani /-ik/ is pronounced /-iʃ/.

Pharyngealization of the emphatic consonants tends to weaken in many of the spoken varieties, and to spread from emphatic

consonants to nearby sounds. In addition, the "emphatic" allophone [ɑ] automatically triggers pharyngealization of adjacent sounds in many dialects. As a result, it may difficult or impossible to determine whether a given coronal consonant is phonemically emphatic or not, especially in dialects with long-distance emphasis spreading. (A notable exception is the sounds /t/ vs. /tˤ/ in Moroccan Arabic, because the former is pronounced as an affricate[t͡s] but the latter is not.)

Grammar

Literary Arabic

As in other Semitic languages, Arabic has a complex and unusual morphology (i.e. method of constructing words from a basic root). Arabic has a nonconcatenative "root-and-pattern" morphology: A root consists of a set of bare consonants (usually three), which are fitted into a discontinuous pattern to form words. For example, the word for 'I wrote' is constructed by combining the root ***k-t-b*** 'write' with the pattern ***-a-a-tu*** 'I Xed' to form *katabtu* 'I wrote'. Other verbs meaning 'I Xed' will typically have the same pattern but with different consonants, e.g. *qara'tu* 'I read', *akaltu* 'I ate', *dhahabtu* 'I went', although other patterns are possible (e.g. *sharibtu* 'I drank', *qultu* 'I said', *takallamtu* 'I spoke', where the subpattern used to signal the past tense may change but the suffix *-tu* is always used).

From a single root ***k-t-b***, numerous words can be formed by applying different patterns:

- كَتَبْتُ *katabtu* 'I wrote'
- كَتَّبْتُ *kattabtu* 'I had (something) written'
- كَاتَبْتُ *kātabtu* 'I corresponded (with someone)'
- أَكْتَبْتُ *'aktabtu* 'I dictated'
- اِكْتَتَبْتُ *iktatabtu* 'I subscribed'
- تَكَاتَبْنَا *takātabnā* 'we corresponded with each other'
- أَكْتُبُ *'aktubu* 'I write'
- أُكَتِّبُ *'ukattibu* 'I have (something) written'
- أُكَاتِبُ *'ukātibu* 'I correspond (with someone)'
- أُكْتِبُ *'uktibu* 'I dictate'
- أَكْتَتِبُ *'aktatibu* 'I subscribe'
- نَتَكَاتَبُ *natakātabu* 'we correspond each other'
- كُتِبَ *kutiba* 'it was written'
- أُكْتِبَ *'uktiba* 'it was dictated'
- مَكْتُوبٌ *maktūbun* 'written'
- مُكْتَبٌ *muktabun* 'dictated'
- كِتَابٌ *kitābun* 'book'
- كُتُبٌ *kutubun* 'books'
- كَاتِبٌ *kātibun* 'writer'
- كُتَّابٌ *kuttābun* 'writers'
- مَكْتَبٌ *maktabun* 'desk, office'
- مَكْتَبَةٌ *maktabatun* 'library, bookshop'
- etc.

Nouns and adjectives

Nouns in Literary Arabic have three grammatical cases (nominative, accusative, and genitive [also used when the noun is governed by a preposition]); three numbers (singular, dual and plural); two genders (masculine and feminine); and three "states" (indefinite, definite, and construct). The cases of

singular nouns (other than those that end in long ā) are indicated by suffixed short vowels (/-u/ for nominative, /-a/ for accusative, /-i/ for genitive).

The feminine singular is often marked by ة /-at/, which is pronounced as /-ah/ before a pause. Plural is indicated either through endings (the sound plural) or internal modification (the broken plural). Definite nouns include all proper nouns, all nouns in "construct state" and all nouns which are prefixed by the definite article الـ /al-/. Indefinite singular nouns (other than those that end in long ā) add a final /-n/ to the case-marking vowels, giving /-un/, /-an/ or /-in/ (which is also referred to as nunation or tanwīn).

Adjectives in Literary Arabic are marked for case, number, gender and state, as for nouns. However, the plural of all non-human nouns is always combined with a singular feminine adjective, which takes the ة /-at/ suffix.

Pronouns in Literary Arabic are marked for person, number and gender. There are two varieties, independent pronouns and enclitics. Enclitic pronouns are attached to the end of a verb, noun or preposition and indicate verbal and prepositional objects or possession of nouns. The first-person singular pronoun has a different enclitic form used for verbs (ـني /-nī/) and for nouns or prepositions (ـي /-ī/ after consonants, ـيَ /-ya/ after vowels).

Nouns, verbs, pronouns and adjectives agree with each other in all respects. However, non-human plural nouns are grammatically considered to be feminine singular. Furthermore, a verb in a verb-initial sentence is marked as

singular regardless of its semantic number when the subject of the verb is explicitly mentioned as a noun. Numerals between three and ten show "chiasmic" agreement, in that grammatically masculine numerals have feminine marking and vice versa.

Verbs

Verbs in Literary Arabic are marked for person (first, second, or third), gender, and number. They are conjugated in two major paradigms (past and non-past); two voices (active and passive); and six moods (indicative, imperative, subjunctive, jussive, shorter energetic and longer energetic), the fifth and sixth moods, the energetics, exist only in Classical Arabic but not in MSA. There are also two participles (active and passive) and a verbal noun, but no infinitive.

The past and non-past paradigms are sometimes also termed perfective and imperfective, indicating the fact that they actually represent a combination of tense and aspect. The moods other than the indicative occur only in the non-past, and the future tenseis signaled by prefixing ـسَـ*sa-* or سَوْفَ*sawfa* onto the non-past. The past and non-past differ in the form of the stem (e.g., past ـكَتَبـ*katab-* vs. non-past ـكْتُبـ*-ktub-*), and also use completely different sets of affixes for indicating person, number and gender: In the past, the person, number and gender are fused into a single suffixal morpheme, while in the non-past, a combination of prefixes (primarily encoding person) and suffixes (primarily encoding gender and number) are used. The passive voice uses the same person/number/gender affixes but changes the vowels of the stem.

The following shows a paradigm of a regular Arabic verb, كَتَبَ*kataba* 'to write'. In Modern Standard, the energetic mood (in either long or short form, which have the same meaning) is almost never used.

Derivation

Like other Semitic languages, and unlike most other languages, Arabic makes much more use of nonconcatenative morphology (applying many templates applied roots) to derive words than adding prefixes or suffixes to words.

For verbs, a given root can occur in many different derived verb stems (of which there are about fifteen), each with one or more characteristic meanings and each with its own templates for the past and non-past stems, active and passive participles, and verbal noun. These are referred to by Western scholars as "Form I", "Form II", and so on through "Form XV" (although Forms XI to XV are rare). These stems encode grammatical functions such as the causative, intensive and reflexive. Stems sharing the same root consonants represent separate verbs, albeit often semantically related, and each is the basis for its own conjugational paradigm. As a result, these derived stems are part of the system of derivational morphology, not part of the inflectional system.

Form II is sometimes used to create transitive denominative verbs (verbs built from nouns); Form V is the equivalent used for intransitive denominatives.

The associated participles and verbal nouns of a verb are the primary means of forming new lexical nouns in Arabic. This is

similar to the process by which, for example, the English gerund "meeting" (similar to a verbal noun) has turned into a noun referring to a particular type of social, often work-related event where people gather together to have a "discussion" (another lexicalized verbal noun). Another fairly common means of forming nouns is through one of a limited number of patterns that can be applied directly to roots, such as the "nouns of location" in *ma-* (e.g. *maktab* 'desk, office' *<k-t-b* 'write', *maṭbakh* 'kitchen' *<ṭ-b-kh* 'cook').

The only three genuine suffixes are as follows:

- The feminine suffix *-ah*; variously derives terms for women from related terms for men, or more generally terms along the same lines as the corresponding masculine, e.g. *maktabah* 'library' (also a writing-related place, but different from *maktab*, as above).
- The nisbah suffix *-iyy-*. This suffix is extremely productive, and forms adjectives meaning "related to X". It corresponds to English adjectives in *-ic, -al, -an, -y, -ist*, etc.
- The feminine nisbah suffix *-iyyah*. This is formed by adding the feminine suffix *-ah* onto nisba adjectives to form abstract nouns. For example, from the basic root *sh-r-k* 'share' can be derived the Form VIII verb *ishtaraka* 'to cooperate, participate', and in turn its verbal noun *ishtirāk* 'cooperation, participation' can be formed. This in turn can be made into a nisbah adjective *ishtirākī* 'socialist', from which an abstract noun *ishtirākiyyah* 'socialism' can be derived. Other recent formations are *jumhūriyyah* 'republic' (lit. "public-ness", *<jumhūr* 'multitude, general public'),

117

and the Gaddafi-specific variation *jamāhīriyyah* 'people's republic' (lit. "masses-ness", <*jamāhīr* 'the masses', pl. of *jumhūr*, as above).

Colloquial varieties

- The spoken dialects have lost the case distinctions and make only limited use of the dual (it occurs only on nouns and its use is no longer required in all circumstances). They have lost the mood distinctions other than imperative, but many have since gained new moods through the use of prefixes (most often /bi-/ for indicative vs. unmarked subjunctive). They have also mostly lost the indefinite "nunation" and the internal passive.

Writing system

The Arabic alphabet derives from the Aramaic through Nabatean, to which it bears a loose resemblance like that of Coptic or Cyrillic scripts to Greek script. Traditionally, there were several differences between the Western (North African) and Middle Eastern versions of the alphabet—in particular, the *fa'* had a dot underneath and *qaf* a single dot above in the Maghreb, and the order of the letters was slightly different (at least when they were used as numerals).

However, the old Maghrebi variant has been abandoned except for calligraphic purposes in the Maghreb itself, and remains in use mainly in the Quranic schools (zaouias) of West Africa. Arabic, like all other Semitic languages (except for the Latin-

written Maltese, and the languages with the Ge'ez script), is written from right to left. There are several styles of scripts such as thuluth, muhaqqaq, tawqi, rayhan and notably naskh, which is used in print and by computers, and ruqʻ ah, which is commonly used for correspondence.

Originally Arabic was made up of only *rasm* without diacritical marks Later diacritical points (which in Arabic are referred to as *nuqaṯ*) were added (which allowed readers to distinguish between letters such as b, t, th, n and y). Finally signs known as *Tashkil* were used for short vowels known as *harakat* and other uses such as final postnasalized or long vowels.

Calligraphy

After Khalil ibn Ahmad al Farahidi finally fixed the Arabic script around 786, many styles were developed, both for the writing down of the Quran and other books, and for inscriptions on monuments as decoration.

Arabic calligraphy has not fallen out of use as calligraphy has in the Western world, and is still considered by Arabs as a major art form; calligraphers are held in great esteem. Being cursive by nature, unlike the Latin script, Arabic script is used to write down a verse of the Quran, a hadith, or simply a proverb. The composition is often abstract, but sometimes the writing is shaped into an actual form such as that of an animal. One of the current masters of the genre is Hassan Massoudy.

In modern times the intrinsically calligraphic nature of the written Arabic form is haunted by the thought that a

typographic approach to the language, necessary for digitized unification, will not always accurately maintain meanings conveyed through calligraphy.

Romanization

There are a number of different standards for the romanization of Arabic, i.e. methods of accurately and efficiently representing Arabic with the Latin script. There are various conflicting motivations involved, which leads to multiple systems. Some are interested in transliteration, i.e. representing the *spelling* of Arabic, while others focus on transcription, i.e. representing the *pronunciation* of Arabic. (They differ in that, for example, the same letter ي is used to represent both a consonant, as in "**y**ou" or "**y**et", and a vowel, as in "m**e**" or "**ea**t".) Some systems, e.g. for scholarly use, are intended to accurately and unambiguously represent the phonemes of Arabic, generally making the phonetics more explicit than the original word in the Arabic script. These systems are heavily reliant on diacritical marks such as "š" for the sound equivalently written *sh* in English. Other systems (e.g. the Bahá'í orthography) are intended to help readers who are neither Arabic speakers nor linguists with intuitive pronunciation of Arabic names and phrases. These less "scientific" systems tend to avoid diacritics and use digraphs (like *sh* and *kh*). These are usually simpler to read, but sacrifice the definiteness of the scientific systems, and may lead to ambiguities, e.g. whether to interpret *sh* as a single sound, as in *gash*, or a combination of two sounds, as in *gashouse*. The ALA-LC romanization solves this problem by

separating the two sounds with a prime symbol (′); e.g., *as'hal* 'easier'.

During the last few decades and especially since the 1990s, Western-invented text communication technologies have become prevalent in the Arab world, such as personal computers, the World Wide Web, email, bulletin board systems, IRC, instant messaging and mobile phone text messaging. Most of these technologies originally had the ability to communicate using the Latin script only, and some of them still do not have the Arabic script as an optional feature. As a result, Arabic speaking users communicated in these technologies by transliterating the Arabic text using the Latin script, sometimes known as IM Arabic.

To handle those Arabic letters that cannot be accurately represented using the Latin script, numerals and other characters were appropriated. For example, the numeral "3" may be used to represent the Arabic letter ⟨ع⟩. There is no universal name for this type of transliteration, but some have named it Arabic Chat Alphabet. Other systems of transliteration exist, such as using dots or capitalization to represent the "emphatic" counterparts of certain consonants. For instance, using capitalization, the letter ⟨د⟩, may be represented by **d**. Its emphatic counterpart, ⟨ض⟩, may be written as **D**.

Numerals

In most of present-day North Africa, the Western Arabic numerals (0, 1, 2, 3, 4, 5, 6, 7, 8, 9) are used. However, in Egypt and Arabic-speaking countries to the east of it, the

Eastern Arabic numerals (٠–١–٢–٣ – ٤–٥–٦–٧–٨–٩) are in use. When representing a number in Arabic, the lowest-valued positionis placed on the right, so the order of positions is the same as in left-to-right scripts. Sequences of digits such as telephone numbers are read from left to right, but numbers are spoken in the traditional Arabic fashion, with units and tens reversed from the modern English usage. For example, 24 is said "four and twenty" just like in the German language (*vierundzwanzig*) and Classical Hebrew, and 1975 is said "a thousand and nine-hundred and five and seventy" or, more eloquently, "a thousand and nine-hundred five seventy"

Language-standards regulators

Academy of the Arabic Language is the name of a number of language-regulation bodies formed in the Arab League. The most active are in Damascus and Cairo. They review language development, monitor new words and approve inclusion of new words into their published standard dictionaries. They also publish old and historical Arabic manuscripts.

As a foreign language

Arabic has been taught worldwide in many elementary and secondary schools, especially Muslim schools. Universities around the world have classes that teach Arabic as part of their foreign languages, Middle Eastern studies, and religious studies courses. Arabic language schools exist to assist students to learn Arabic outside the academic world. There are many Arabic language schools in the Arab world and other

Muslim countries. Because the Quran is written in Arabic and all Islamic terms are in Arabic, millions of Muslims (both Arab and non-Arab) study the language. Software and books with tapes are also important part of Arabic learning, as many of Arabic learners may live in places where there are no academic or Arabic language school classes available. Radio series of Arabic language classes are also provided from some radio stations. A number of websites on the Internet provide online classes for all levels as a means of distance education; most teach Modern Standard Arabic, but some teach regional varieties from numerous countries.

Status in the Arab world vs. other languages

With the sole example of Medieval linguist Abu Hayyan al-Gharnati – who, while a scholar of the Arabic language, was not ethnically Arab – Medieval scholars of the Arabic language made no efforts at studying comparative linguistics, considering all other languages inferior.

In modern times, the educated upper classes in the Arab world have taken a nearly opposite view. Yasir Suleiman wrote in 2011 that "studying and knowing English or French in most of the Middle East and North Africa have become a badge of sophistication and modernity and … feigning, or asserting, weakness or lack of facility in Arabic is sometimes paraded as a sign of status, class, and perversely, even education through a mélange of code-switching practises."

German Language

- **German** is a West Germanic language mainly spoken in Central Europe. It is the most widely spoken and official or co-official language in Germany, Austria, Switzerland, Liechtenstein, and the Italian province of South Tyrol. It is also a co-official language of Luxembourg and Belgium, as well as a national language in Namibia. German is most similar to other languages within the West Germanic language branch, including Afrikaans, Dutch, English, the Frisian languages, Low German, Luxembourgish, Scots, and Yiddish. It also contains close similarities in vocabulary to some languages in the North Germanic group, such as Danish, Norwegian, and Swedish. German is the second most widely spoken Germanic language after English.

One of the major languages of the world, German is a native language to almost 100 million people worldwide and is spoken by a total of over 130 million people. It is the most spoken native language within the European Union. German is also widely taught as a foreign language, especially in Europe, where it is the third most taught foreign language (after English and French), and the United States. The language has been influential in the fields of philosophy, theology, science, and technology. It is the second most commonly used scientific language and among the most widely used languages on websites. The German-speaking countries are ranked fifth in

terms of annual publication of new books, with one-tenth of all books (including e-books) in the world being published in German.

German is an inflected language, with four cases for nouns, pronouns, and adjectives (nominative, accusative, genitive, dative); three genders (masculine, feminine, neuter); and two numbers (singular, plural). It has strong and weak verbs. The majority of its vocabulary derives from the ancient Germanic branch of the Indo-European language family, while a smaller share is partly derived from Latin and Greek, along with fewer words borrowed from French and Modern English.

German is a pluricentric language; the three standardized variants are German, Austrian, and Swiss Standard High German. It is also notable for its broad spectrum of dialects, with many varieties existing in Europe and other parts of the world. Some of these non-standard varieties have become recognized and protected by regional or national governments.

Classification

Modern Standard German is a West Germanic language in the Germanic branch of the Indo-European languages. The Germanic languages are traditionally subdivided into three branches, North Germanic, East Germanic, and West Germanic. The first of these branches survives in modern Danish, Swedish, Norwegian, Faroese, and Icelandic, all of which are descended from Old Norse. The East Germanic languages are now extinct, and Gothic is the only language in this branch which survives in written texts. The West

Germanic languages, however, have undergone extensive dialectal subdivision and are now represented in modern languages such as English, German, Dutch, Yiddish, Afrikaans, and others.

Within the West Germanic language dialect continuum, the Benrath and Uerdingen lines (running through Düsseldorf-Benrath and Krefeld-Uerdingen, respectively) serve to distinguish the Germanic dialects that were affected by the High German consonant shift (south of Benrath) from those that were not (north of Uerdingen). The various regional dialects spoken south of these lines are grouped as High German dialects, while those spoken to the north comprise the Low German/Low Saxon and Low Franconian dialects. As members of the West Germanic language family, High German, Low German, and Low Franconian have been proposed to be further distinguished historically as Irminonic, Ingvaeonic, and Istvaeonic, respectively. This classification indicates their historical descent from dialects spoken by the Irminones (also known as the Elbe group), Ingvaeones (or North Sea Germanic group), and Istvaeones (or Weser-Rhine group).

Standard German is based on a combination of Thuringian-Upper Saxon and Upper Franconian dialects, which are Central German and Upper German dialects belonging to the High German dialect group. German is therefore closely related to the other languages based on High German dialects, such as Luxembourgish (based on Central Franconian dialects) and Yiddish. Also closely related to Standard German are the Upper German dialects spoken in the southern German-speaking countries, such as Swiss German (Alemannic dialects) and the various Germanic dialects spoken in the French region of

Grand Est, such as Alsatian (mainly Alemannic, but also Central- and Upper Franconian dialects) and Lorraine Franconian (Central Franconian).

After these High German dialects, standard German is less closely related to languages based on Low Franconian dialects (e.g. Dutch and Afrikaans), Low German or Low Saxon dialects (spoken in northern Germany and southern Denmark), neither of which underwent the High German consonant shift. As has been noted, the former of these dialect types is Istvaeonic and the latter Ingvaeonic, whereas the High German dialects are all Irminonic; the differences between these languages and standard German are therefore considerable. Also related to German are the Frisian languages—North Frisian (spoken in Nordfriesland), Saterland Frisian (spoken in Saterland), and West Frisian (spoken in Friesland)—as well as the Anglic languages of English and Scots. These Anglo-Frisian dialects did not take part in the High German consonant shift.

History

Old High German

The history of the German language begins with the High German consonant shift during the migration period, which separated Old High German dialects from Old Saxon. This sound shift involved a drastic change in the pronunciation of both voiced and voiceless stop consonants (*b*, *d*, *g*, and *p*, *t*, *k*, respectively). The primary effects of the shift were the following below.

- Voiceless stops became long (geminated) voiceless fricatives following a vowel;
- Voiceless stops became affricates in word-initial position, or following certain consonants;
- Voiced stops became voiceless in certain phonetic settings.

While there is written evidence of the Old High German language in several Elder Futhark inscriptions from as early as the sixth century AD (such as the Pforzen buckle), the Old High German period is generally seen as beginning with the *Abrogans* (written c. 765–775), a Latin-German glossary supplying over 3,000 Old High German words with their Latin equivalents. After the *Abrogans*, the first coherent works written in Old High German appear in the ninth century, chief among them being the *Muspilli*, the *Merseburg Charms*, and the *Hildebrandslied*, and other religious texts (the *Georgslied*, the *Ludwigslied*, the *Evangelienbuch*, and translated hymns and prayers). The *Muspilli* is a Christian poem written in a Bavarian dialect offering an account of the soul after the Last Judgment, and the *Merseburg Charms* are transcriptions of spells and charms from the pagan Germanic tradition. Of particular interest to scholars, however, has been the *Hildebrandslied*, a secular epic poem telling the tale of an estranged father and son unknowingly meeting each other in battle. Linguistically this text is highly interesting due to the mixed use of Old Saxon and Old High German dialects in its composition. The written works of this period stem mainly from the Alamanni, Bavarian, and Thuringian groups, all belonging to the Elbe Germanic group (Irminones), which had settled in what is now southern-central Germany and Austria between the second and sixth centuries during the great migration.

In general, the surviving texts of OHG show a wide range of dialectal diversity with very little written uniformity. The early written tradition of OHG survived mostly through monasteries and scriptoria as local translations of Latin originals; as a result, the surviving texts are written in highly disparate regional dialects and exhibit significant Latin influence, particularly in vocabulary. At this point monasteries, where most written works were produced, were dominated by Latin, and German saw only occasional use in official and ecclesiastical writing.

The German language through the OHG period was still predominantly a spoken language, with a wide range of dialects and a much more extensive oral tradition than a written one. Having just emerged from the High German consonant shift, OHG was also a relatively new and volatile language still undergoing a number of phonetic, phonological, morphological, and syntactic changes. The scarcity of written work, instability of the language, and widespread illiteracy of the time explain the lack of standardization up to the end of the OHG period in 1050.

Middle High German

While there is no complete agreement over the dates of the Middle High German (MHG) period, it is generally seen as lasting from 1050 to 1350. This was a period of significant expansion of the geographical territory occupied by Germanic tribes, and consequently of the number of German speakers. Whereas during the Old High German period the Germanic tribes extended only as far east as the Elbe and Saale rivers, the MHG period saw a number of these tribes expanding

beyond this eastern boundary into Slavic territory (known as the *Ostsiedlung*). With the increasing wealth and geographic spread of the Germanic groups came greater use of German in the courts of nobles as the standard language of official proceedings and literature. A clear example of this is the *mittelhochdeutsche Dichtersprache* employed in the Hohenstaufen court in Swabia as a standardized supra-dialectal written language. While these efforts were still regionally bound, German began to be used in place of Latin for certain official purposes, leading to a greater need for regularity in written conventions.

While the major changes of the MHG period were socio-cultural, German was still undergoing significant linguistic changes in syntax, phonetics, and morphology as well (e.g. diphthongization of certain vowel sounds: *hus* (OHG "house")→*haus* (MHG), and weakening of unstressed short vowels to schwa [ə]: *taga* (OHG "days")→*tage* (MHG)).

A great wealth of texts survives from the MHG period. Significantly, these texts include a number of impressive secular works, such as the *Nibelungenlied*, an epic poem telling the story of the dragon-slayer Siegfried (c. thirteenth century), and the *Iwein,* an Arthurian verse poem by Hartmann von Aue (c. 1203), lyric poems, and courtly romances such as *Parzival* and *Tristan.* Also noteworthy is the *Sachsenspiegel*, the first book of laws written in Middle *Low* German (c. 1220). The abundance and especially the secular character of the literature of the MHG period demonstrate the beginnings of a standardized written form of German, as well as the desire of poets and authors to be understood by individuals on supra-dialectal terms.

The Middle High German period is generally seen as ending when the 1346–53 Black Death decimated Europe's population.

Early New High German

Modern German begins with the Early New High German (ENHG) period, which the influential German philologist Wilhelm Scherer dates 1350–1650, terminating with the end of the Thirty Years' War. This period saw the further displacement of Latin by German as the primary language of courtly proceedings and, increasingly, of literature in the German states. While these states were still under the control of the Holy Roman Empire, and far from any form of unification, the desire for a cohesive written language that would be understandable across the many German-speaking principalities and kingdoms was stronger than ever. As a spoken language German remained highly fractured throughout this period, with a vast number of often mutually incomprehensible regional dialects being spoken throughout the German states; the invention of the printing press c. 1440 and the publication of Luther's vernacular translation of the Bible in 1534, however, had an immense effect on standardizing German as a supra-dialectal written language.

The ENHG period saw the rise of several important cross-regional forms of chancery German, one being *gemeine tiutsch,* used in the court of the Holy Roman Emperor Maximilian I, and the other being *Meißner Deutsch*, used in the Electorate of Saxony in the Duchy of Saxe-Wittenberg.

Alongside these courtly written standards, the invention of the printing press led to the development of a number of printers'

languages (*Druckersprachen*) aimed at making printed material readable and understandable across as many diverse dialects of German as possible. The greater ease of production and increased availability of written texts brought about increased standardization in the written form of German.

One of the central events in the development of ENHG was the publication of Luther's translation of the Bible into German (the New Testament was published in 1522; the Old Testament was published in parts and completed in 1534). Luther based his translation primarily on the *Meißner Deutsch* of Saxony, spending much time among the population of Saxony researching the dialect so as to make the work as natural and accessible to German speakers as possible. Copies of Luther's Bible featured a long list of glosses for each region, translating words which were unknown in the region into the regional dialect. Luther said the following concerning his translation method:

One who would talk German does not ask the Latin how he shall do it; he must ask the mother in the home, the children on the streets, the common man in the market-place and note carefully how they talk, then translate accordingly. They will then understand what is said to them because it is German. When Christ says 'ex abundantia cordis os loquitur,' I would translate, if I followed the papists, *aus dem Überflusz des Herzens redet der Mund*. But tell me is this talking German? What German understands such stuff? No, the mother in the home and the plain man would say, *Wesz das Herz voll ist, des gehet der Mund über*.

With Luther's rendering of the Bible in the vernacular, German asserted itself against the dominance of Latin as a legitimate language for courtly, literary, and now ecclesiastical subject-matter. Furthermore, his Bible was ubiquitous in the German states: nearly every household possessed a copy. Nevertheless, even with the influence of Luther's Bible as an unofficial written standard, a widely accepted standard for written German did not appear until the middle of the eighteenth century.

Austrian Empire

German was the language of commerce and government in the Habsburg Empire, which encompassed a large area of Central and Eastern Europe. Until the mid-nineteenth century, it was essentially the language of townspeople throughout most of the Empire. Its use indicated that the speaker was a merchant or someone from an urban area, regardless of nationality.

Prague (German: *Prag*) and Budapest (Buda, German: *Ofen*), to name two examples, were gradually Germanized in the years after their incorporation into the Habsburg domain; others, like Pozsony (German: *Pressburg*, now Bratislava), were originally settled during the Habsburg period and were primarily German at that time. Prague, Budapest, Bratislava, and cities like Zagreb (German: *Agram*) or Ljubljana (German: *Laibach*), contained significant German minorities.

In the eastern provinces of Banat, Bukovina, and Transylvania (German: *Banat, Buchenland, Siebenbürgen*), German was the predominant language not only in the larger towns – like *Temeschburg* (Timișoara), *Hermannstadt* (Sibiu) and *Kronstadt*

(Brașov) – but also in many smaller localities in the surrounding areas.

Standardisation

The most comprehensive guide to the vocabulary of the German language is found within the *Deutsches Wörterbuch*. This dictionary was created by the Brothers Grimm, and is composed of 16 parts which were issued between 1852 and 1860. In 1872, grammatical and orthographic rules first appeared in the *Duden Handbook*.

In 1901, the Second Orthographical Conference ended with a complete standardisation of the German language in its written form, and the *Duden Handbook* was declared its standard definition. The *Deutsche Bühnensprache* (lit. 'German stage language') had established conventions for German pronunciation in theatres three years earlier; however, this was an artificial standard that did not correspond to any traditional spoken dialect. Rather, it was based on the pronunciation of Standard German in Northern Germany, although it was subsequently regarded often as a general prescriptive norm, despite differing pronunciation traditions especially in the Upper-German-speaking regions that still characterise the dialect of the area today – especially the pronunciation of the ending *-ig* as [ɪk] instead of [ɪç]. In Northern Germany, Standard German was a foreign language to most inhabitants, whose native dialects were subsets of Low German. It was usually encountered only in writing or formal speech; in fact, most of Standard German was a written language, not identical to any spoken dialect, throughout the German-speaking area until well into the 19th century.

Official revisions of some of the rules from 1901 were not issued until the controversial German orthography reform of 1996 was made the official standard by governments of all German-speaking countries. Media and written works are now almost all produced in Standard German (often called*Hochdeutsch*, "High German") which is understood in all areas where German is spoken.

Geographical distribution

As a result of the German diaspora, as well as the popularity of German taught as a foreign language, the geographical distribution of German speakers (or "Germanophones") spans all inhabited continents. A 2020 estimate by Ethnologue places the total number of Standard German speakers at 132 million, of which over 75 million are native speakers.

However, an exact, global number of native German speakers is complicated by the existence of several varieties whose status as separate "languages" or "dialects" is disputed for political and linguistic reasons, including quantitatively strong varieties like certain forms of Alemannic and Low German. With the inclusion or exclusion of certain varieties, it is estimated that approximately 90–95 million people speak German as a first language, 10–25 million speak it as a second language, and 75–100 million as a foreign language. This would imply the existence of approximately 175–220 million German speakers worldwide.

Europe

As of 2012, about 90 million people, or 16% of the European Union's population spoke German as their mother tongue, making it the second most widely spoken language on the continent after Russian and the second biggest language in terms of overall speakers (after English).

German Sprachraum

The area in central Europe where the majority of the population speaks German as a first language and has German as a (co-)official language is called the "German *Sprachraum*". German is the sole official language of the following countries:

- Germany
- Austria
- 17 cantons of Switzerland
- Liechtenstein

German is a co-official language of the following countries:

- Belgium (as majority language only in the German-speaking Community, which represents 0.7% of the Belgian population)
- Luxembourg, along with French and Luxembourgish
- Switzerland, co-official at the federal level with French, Italian, and Romansh, and at the local level in four cantons: Bern (with French), Fribourg (with French), Grisons (with Italian and Romansh) and Valais (with French)

- Italian Autonomous Province of South Tyrol (also majority language)

Outside the Sprachraum

Although expulsions and (forced) assimilation after the two World wars greatly diminished them, minority communities of mostly bilingual German native speakers exist in areas both adjacent to and detached from the Sprachraum.

Within Europe, German is a recognized minority language in the following countries:

- Czech Republic (see also: Germans in the Czech Republic)
- Denmark (see also: North Schleswig Germans)
- Hungary (see also: Germans of Hungary)
- Poland (see also German minority in Poland; German is an auxiliary and co-official language in 31 communes)
- Romania (see also: Germans of Romania)
- Russia, (see also: Germans in Russia)
- Slovakia (see also: Carpathian Germans)

In France, the High German varieties of Alsatian and Moselle Franconian are identified as "regional languages", but the European Charter for Regional or Minority Languages of 1998 has not yet been ratified by the government.

Africa

Namibia

Namibia was a colony of the German Empire from 1884 to 1919. About 30,000 people still speak German as a native tongue today, mostly descendants of German colonial settlers. The period of German colonialism in Namibia also led to the evolution of a Standard German-based pidgin language called "Namibian Black German", which became a second language for parts of the indigenous population. Although it is nearly extinct today, some older Namibians still have some knowledge of it.

German remained a *de facto* official language of Namibia after the end of German colonial rule alongside English and Afrikaans, and had *de jure* co-official status from 1984 until its independence from South Africa in 1990. However, the Namibian government perceived Afrikaans and German as symbols of apartheid and colonialism, and decided English would be the sole official language upon independence, stating that it was a "neutral" language as there were virtually no English native speakers in Namibia at that time. German, Afrikaans, and several indigenous languages thus became "national languages" by law, identifying them as elements of the cultural heritage of the nation and ensuring that the state acknowledged and supported their presence in the country.

Today, Namibia is considered to be the only German-speaking country outside of the *Sprachraum* in Europe. German is used in a wide variety of spheres throughout the country, especially

in business, tourism, and public signage, as well as in education, churches (most notably the German-speaking Evangelical Lutheran Church in Namibia (GELK)), other cultural spheres such as music, and media (such as German language radio programs by the Namibian Broadcasting Corporation). The *Allgemeine Zeitung* is one of the three biggest newspapers in Namibia and the only German-language daily in Africa.

South Africa

An estimated 12,000 people speak German or a German variety as a first language in South Africa, mostly originating from different waves of immigration during the 19th and 20th centuries. One of the largest communities consists of the speakers of "Nataler Deutsch", a variety of Low German concentrated in and around Wartburg. The South African constitution identifies German as a "commonly used" language and the Pan South African Language Board is obligated to promote and ensure respect for it.

North America

In the United States, German is the fifth most spoken language in terms of native and second language speakers after English, Spanish, French, and Chinese (with figures for Cantonese and Mandarin combined), with over 1 million total speakers. In the states of North Dakota and South Dakota, German is the most common language spoken at home after English. As a legacy of significant German immigration to the country, German geographical names can be found throughout the Midwest

region, such as New Ulm and Bismarck (North Dakota's state capital).

A number of German varieties have developed in the country and are still spoken today, such as Pennsylvania German and Texas German.

South America

In Brazil, the largest concentrations of German speakers are in the states of Rio Grande do Sul (where Riograndenser Hunsrückisch developed), Santa Catarina, and Espírito Santo.

German dialects (namely Hunsrik and East Pomeranian) are recognized languages in the following municipalities in Brazil:

- Espírito Santo (statewide cultural language): Domingos Martins, Laranja da Terra, Pancas, Santa Maria de Jetibá, Vila Pavão
- Rio Grande do Sul (Riograndenser Hunsrückisch German is a designated cultural language in the state): Santa Maria do Herval, Canguçu
- Santa Catarina: Antônio Carlos, Pomerode (standard German recognized)

Small concentrations of German-speakers and their descendants are also found in Argentina, Chile, Paraguay, Venezuela, and Bolivia.

Oceania

In Australia, the state of South Australia experienced a pronounced wave of Prussian immigration in the 1840s (particularly from Silesia region). With the prolonged isolation from other German speakers and contact with Australian English, a unique dialect known as Barossa German developed, spoken predominantly in the Barossa Valley near Adelaide. Usage of German sharply declined with the advent of World War I, due to the prevailing anti-German sentiment in the population and related government action. It continued to be used as a first language into the 20th century, but its use is now limited to a few older speakers.

As of the 2013 census, 36,642 people in New Zealand spoke German, mostly descendants of a small wave of 19th century German immigrants, making it the third most spoken European language after English and French and overall the ninth most spoken language.

A German creole named *Unserdeutsch* was historically spoken in the former German colony of German New Guinea, modern day Papua New Guinea. It is at a high risk of extinction, with only about 100 speakers remaining, and a topic of interest among linguists seeking to revive interest in the language.

As a foreign language

Like English, French, and Spanish, German has become a standard foreign language throughout the world, especially in the Western World. German ranks second on par with French

among the best known foreign languages in the European Union (EU) after English, as well as in Russia and Turkey. In terms of student numbers across all levels of education, German ranks third in the EU (after English and French) and in the United States (after Spanish and French). In 2020, approximately 15.4 million people were enrolled in learning German across all levels of education worldwide. This number has decreased from a peak of 20.1 million in 2000. Within the EU, not counting countries where it is an official language, German as a foreign language is most popular in Eastern and Northern Europe, namely the Czech Republic, Croatia, Denmark, the Netherlands, Slovakia, Hungary, Slovenia, Sweden, Poland, and Bosnia and Herzegovina. German was once, and to some extent still is, a lingua franca in those parts of Europe.

Standard German

The basis of Standard German developed with the Luther Bible and the chancery language spoken by the Saxon court. However, there are places where the traditional regional dialects have been replaced by new vernaculars based on standard German; that is the case in large stretches of Northern Germany but also in major cities in other parts of the country. It is important to note, however, that the colloquial standard German differs greatly from the formal written language, especially in grammar and syntax, in which it has been influenced by dialectal speech.

Standard German differs regionally among German-speaking countries in vocabulary and some instances of pronunciation

and even grammar and orthography. This variation must not be confused with the variation of local dialects. Even though the regional varieties of standard German are only somewhat influenced by the local dialects, they are very distinct. German is thus considered a pluricentric language.

In most regions, the speakers use a continuum from more dialectal varieties to more standard varieties depending on the circumstances.

Varieties

In German linguistics, German dialectsare distinguished from varieties of standard German. The *varieties of standard German* refer to the different local varieties of the pluricentric standard German. They differ only slightly in lexicon and phonology. In certain regions, they have replaced the traditional German dialects, especially in Northern Germany.

- German Standard German
- Austrian Standard German
- Swiss Standard German

In the German-speaking parts of Switzerland, mixtures of dialect and standard are very seldom used, and the use of Standard German is largely restricted to the written language. About 11% of the Swiss residents speak *High German* (Standard German) at home, but this is mainly due to German immigrants. This situation has been called a *medial diglossia.* Swiss Standard Germanis used in the Swiss education system, while Austrian German is officially used in the Austrian education system.

A mixture of dialect and standard does not normally occur in Northern Germany either. The traditional varieties there are Low German, whereas Standard German is a High German "variety". Because their linguistic distance is greater, they do not mesh with Standard German the way that High German dialects (such as Bavarian, Swabian, and Hessian) can.

Dialects

The German dialects are the traditional local varieties of the language; many of them are not mutually intelligibile with standard German, and they have great differences in lexicon, phonology, and syntax. If a narrow definition of language based on mutual intelligibility is used, many German dialects are considered to be separate languages (for instance in the *Ethnologue*). However, such a point of view is unusual in German linguistics.

The German dialect continuum is traditionally divided most broadly into High German and Low German, also called Low Saxon. However, historically, High German dialects and Low Saxon/Low German dialects do not belong to the same language. Nevertheless, in today's Germany, Low Saxon/Low German is often perceived as a dialectal variation of Standard German on a functional level even by many native speakers. The same phenomenon is found in the eastern Netherlands, as the traditional dialects are not always identified with their Low Saxon/Low German origins, but with Dutch.

The variation among the German dialects is considerable, with often only neighbouring dialects being mutually intelligible.

Some dialects are not intelligible to people who know only Standard German. However, all German dialects belong to the dialect continuum of High German and Low Saxon.

Low German and Low Saxon

Middle Low German was the lingua franca of the Hanseatic League. It was the predominant language in Northern Germany until the 16th century. In 1534, the Luther Biblewas published. It aimed to be understandable to a broad audience and was based mainly on Central and Upper German varieties. The Early New High German language gained more prestige than Low German and became the language of science and literature. Around the same time, the Hanseatic League, a confederation of northern ports, lost its importance as new trade routes to Asia and the Americas were established, and the most powerful German states of that period were located in Middle and Southern Germany.

The 18th and 19th centuries were marked by mass education in Standard German in schools. Gradually, Low German came to be politically viewed as a mere dialect spoken by the uneducated. Today, Low Saxon can be divided in two groups: Low Saxon varieties with a reasonable level of Standard German influence and varieties of Standard German with a Low Saxon influence known as *Missingsch.* Sometimes, Low Saxon and Low Franconian varieties are grouped together because both are unaffected by the High German consonant shift. However, the proportion of the population who can understand and speak it has decreased continuously since World War II. The major cities in the Low German area are Hamburg, Hanover, Bremen and Dortmund.

Low Franconian

In Germany, Low Franconian dialects are spoken in the northwest of North Rhine-Westphalia, along the Lower Rhine. The Low Franconian dialects spoken in Germany are referred to as Low Rhenish. In the north of the German Low Franconian language area, North Low Franconian dialects (also referred to as Cleverlands or as dialects of South Guelderish) are spoken. The South Low Franconian and Bergish dialects, which are spoken in the south of the German Low Franconian language area, are transitional dialects between Low Franconian and Ripuarian dialects.

The Low Franconian dialects fall within a linguistic category used to classify a number of historical and contemporary West Germanic varieties most closely related to, and including, the Dutch language. Consequently, the vast majority of the Low Franconian dialects are spoken outside of the German language area, in the Netherlands and Belgium. During the Middle Ages and Early Modern Period, the Low Franconian dialects now spoken in Germany, used Middle Dutch or Early Modern Dutch as their literary language and Dachsprache. Following a 19th-century change in Prussian language policy, use of Dutch as an official and public language was forbidden; resulting in Standard German taking its place as the region's official language. As a result, these dialects are now considered German dialects from a socio-linguistic point of view. Nevertheless, topologically these dialects are structurally and phonologically far more similar to Dutch, than to German and form the both the smallest and most divergent dialect cluster within the contemporary German language area.

High German

The High German dialects consist of the Central German, High Franconian, and Upper German dialects. The High Franconian dialects are transitional dialects between Central and Upper German. The High German varieties spoken by the Ashkenazi Jews have several unique features and are considered as a separate language, Yiddish, written with the Hebrew alphabet.

Central German

The Central German dialects are spoken in Central Germany, from Aachen in the west to Görlitz in the east. They consist of Franconian dialects in the west (West Central German) and non-Franconian dialects in the east (East Central German). Modern Standard German is mostly based on Central German dialects.

The Franconian, **West Central German** dialects are the Central Franconian dialects (Ripuarian and Moselle Franconian) and the Rhine Franconian dialects (Hessian and Palatine). These dialects are considered as

- German in Germany and Belgium
- Luxembourgish in Luxembourg
- Lorraine Franconian (spoken in Moselle) and as a Rhine Franconian variant of Alsatian (spoken in Alsace bossue only) in France
- Limburgish or Kerkrade dialect in the Netherlands.

Luxembourgish as well as the Transylvanian Saxon dialect spoken in Transylvaniaare based on Moselle Franconian dialects. The major cities in the Franconian Central German area are Cologne and Frankfurt.

Further east, the non-Franconian, **East Central German** dialects are spoken (Thuringian, Upper Saxon and North Upper Saxon-South Markish, and earlier, in the then German-speaking parts of Silesia alsoSilesian, and in then German southern East Prussia also High Prussian). The major cities in the East Central German area are Berlin and Leipzig.

High Franconian

The High Franconian dialects are transitional dialects between Central and Upper German. They consist of the East and South Franconian dialects.

The East Franconian dialect branch is one of the most spoken dialect branches in Germany. These dialects are spoken in the region of Franconia and in the central parts of SaxonVogtland. Franconia consists of the Bavarian districts of Upper, Middle, and Lower Franconia, the region of South Thuringia (Thuringia), and the eastern parts of the region of Heilbronn-Franken (Tauber Franconia and Hohenlohe) in Baden-Württemberg. The major cities in the East Franconian area are Nuremberg and Würzburg.

South Franconian is mainly spoken in northern Baden-Württemberg in Germany, but also in the northeasternmost part of the region of Alsace in France. While these dialects are considered as dialects of German in Baden-Württemberg, they

are considered as dialects of Alsatian in Alsace (most Alsatian dialects are Low Alemannic, however). The major cities in the South Franconian area are Karlsruhe and Heilbronn.

Upper German

The Upper German dialects are the Alemannic dialects in the west and the Bavarian dialects in the east.

Alemannic

Alemannic dialects are spoken in Switzerland (High Alemannic in the densely populated Swiss Plateau, in the south also Highest Alemannic, and Low Alemannic in Basel), Baden-Württemberg (Swabian and Low Alemannic, in the southwest also High Alemannic), Bavarian Swabia (Swabian, in the southwesternmost part also Low Alemannic), Vorarlberg (Low, High, and Highest Alemannic), Alsace (Low Alemannic, in the southernmost part also High Alemannic), Liechtenstein (High and Highest Alemannic), and in the Tyroleandistrict of Reutte (Swabian). The Alemannic dialects are considered as Alsatian in Alsace. The major cities in the Alemannic area are Stuttgart, Freiburg, Basel, Zürich, Lucerne and Bern.

Bavarian

Bavarian dialects are spoken in Austria (Vienna, Lower and Upper Austria, Styria, Carinthia, Salzburg, Burgenland, and in most parts of Tyrol), Bavaria (Upper and Lower Bavaria as well

as Upper Palatinate), South Tyrol, southwesternmost Saxony (Southern Vogtlandian), and in the Swiss village of Samnaun. The major cities in the Bavarian area are Vienna, Munich, Salzburg, Regensburg, Graz and Bolzano.

Grammar

German is a fusional language with a moderate degree of inflection, with three grammatical genders; as such, there can be a large number of words derived from the same root.

Noun inflection

German nouns inflect by case, gender, and number:

- fourcases: nominative, accusative, genitive, and dative.
- threegenders: masculine, feminine, and neuter. Word endings sometimes reveal grammatical gender: for instance, nouns ending in *-ung* (-ing), *-schaft* (-ship), *-keit* or *heit* (-hood, -ness) are feminine, nouns ending in *-chen* or *-lein* (diminutive forms) are neuter and nouns ending in *-ismus* (-ism) are masculine. Others are more variable, sometimes depending on the region in which the language is spoken. And some endings are not restricted to one gender: for example, *-er* (-er), such as *Feier* (feminine), celebration, party, *Arbeiter* (masculine), labourer, and *Gewitter* (neuter), thunderstorm.
- two numbers: singular and plural.

This degree of inflection is considerably less than in Old High German and other old Indo-European languages such as Latin, Ancient Greek, and Sanskrit, and it is also somewhat less than, for instance, Old English, modern Icelandic, or Russian. The three genders have collapsed in the plural. With four cases and three genders plus plural, there are 16 permutations of case and gender/number of the article (not the nouns), but there are only six forms of the definite article, which together cover all 16 permutations. In nouns, inflection for case is required in the singular for strong masculine and neuter nouns only in the genitive and in the dative (only in fixed or archaic expressions), and even this is losing ground to substitutes in informal speech. Weak masculine nouns share a common case ending for genitive, dative, and accusative in the singular. Feminine nouns are not declined in the singular. The plural has an inflection for the dative. In total, seven inflectional endings (not counting plural markers) exist in German: *-s, -es, -n, -ns, -en, -ens, -e.*

In German orthography, nouns and most words with the syntactical function of nouns are capitalised to make it easier for readers to determine the function of a word within a sentence (*Am Freitag ging ich einkaufen.* – "On Friday I went shopping."; *Eines Tages kreuzte er endlich auf.* – "One day he finally showed up.") This convention is almost unique to German today (shared perhaps only by the closely related Luxembourgish language and several insular dialects of the North Frisian language), but it was historically common in other languages such as Danish (which abolished the capitalization of nouns in 1948) and English.

Like the other Germanic languages, German forms noun compounds in which the first noun modifies the category given by the second: *Hundehütte* ("dog hut"; specifically: "dog kennel"). Unlike English, whose newer compounds or combinations of longer nouns are often written "open" with separating spaces, German (like some other Germanic languages) nearly always uses the "closed" form without spaces, for example: *Baumhaus* ("tree house"). Like English, German allows arbitrarily long compounds in theory (see also English compounds). The longest German word verified to be actually in (albeit very limited) use is *Rindfleischetikettierungsüberwachungsaufgabenübertragungsgesetz*, which, literally translated, is "beef labelling supervision duties assignment law" [from *Rind* (cattle), *Fleisch* (meat), *Etikettierung(s)* (labelling), *Überwachung(s)* (supervision), *Aufgaben* (duties), *Übertragung(s)* (assignment), *Gesetz* (law)]. However, examples like this are perceived by native speakers as excessively bureaucratic, stylistically awkward, or even satirical.

Verb inflection

The inflection of standard German verbs includes:

- two main conjugation classes: weak and strong (as in English). Additionally, there is a third class, known as mixed verbs, whose conjugation combines features of both the strong and weak patterns.
- threepersons: first, second and third.
- twonumbers: singular and plural.
- threemoods: indicative, imperative and subjunctive (in addition to infinitive).

- twovoices: active and passive. The passive voice uses auxiliary verbs and is divisible into static and dynamic. Static forms show a constant state and use the verb "to be" (sein). Dynamic forms show an action and use the verb "to become" (werden).
- twotenses without auxiliary verbs (present and preterite) and four tenses constructed with auxiliary verbs (perfect, pluperfect, future and future perfect).
- the distinction between grammatical aspects is rendered by combined use of the subjunctive or preterite marking so the plain indicative voice uses neither of those two markers; the subjunctive by itself often conveys reported speech; subjunctive plus preterite marks the conditional state; and the preterite alone shows either plain indicative (in the past), or functions as a (literal) alternative for either reported speech or the conditional state of the verb, when necessary for clarity.
- the distinction between perfect and progressive aspect is and has, at every stage of development, been a productive category of the older language and in nearly all documented dialects, but strangely enough it is now rigorously excluded from written usage in its present normalised form.
- disambiguation of completed vs. uncompleted forms is widely observed and regularly generated by common prefixes (*blicken* [to look], *erblicken* [to see – unrelated form: *sehen*]).

Verb prefixes

The meaning of basic verbs can be expanded and sometimes radically changed through the use of a number of prefixes. Some prefixes have a specific meaning; the prefix ***zer-*** refers to destruction, as in ***zer****reißen* (to tear apart), ***zer****brechen* (to break apart), ***zer****schneiden* (to cut apart). Other prefixes have only the vaguest meaning in themselves; ***ver-*** is found in a number of verbs with a large variety of meanings, as in ***ver****suchen* (to try) from *suchen* (to seek), ***ver****nehmen* (to interrogate) from *nehmen* (to take), ***ver****teilen* (to distribute) from *teilen* (to share), ***ver****stehen* (to understand) from *stehen* (to stand).

Other examples include the following: *haften* (to stick), ***ver****haften* (to detain); *kaufen* (to buy), ***ver****kaufen* (to sell); *hören* (to hear), ***auf****hören* (to cease); *fahren* (to drive), ***er****fahren* (to experience).

Many German verbs have a separable prefix, often with an adverbial function. In finite verb forms, it is split off and moved to the end of the clause and is hence considered by some to be a "resultative particle". For example, *mitgehen*, meaning "to go along", would be split, giving *Gehen Sie mit?* (Literal: "Go you with?"; Idiomatic: "Are you going along?").

Indeed, several parenthetical clauses may occur between the prefix of a finite verb and its complement (ankommen = to arrive, er kam an = he arrived, er ist angekommen = he has arrived):

- *Er **kam** am Freitagabend nach einem harten Arbeitstag und dem üblichen Ärger, der ihn schon seit Jahren immer wieder an seinem Arbeitsplatz plagt, mit fraglicher Freude auf ein Mahl, das seine Frau ihm, wie er hoffte, bereits aufgetischt hatte, endlich zu Hause **an**.*

A selectively literal translation of this example to illustrate the point might look like this:

- He "came" on Friday evening, after a hard day at work and the usual annoyances that had time and again been troubling him for years now at his workplace, with questionable joy, to a meal which, as he hoped, his wife had already put on the table, finally home "to".

Word order

German word order is generally with the V2 word order restriction and also with the SOV word order restriction for main clauses. For yes-no questions, exclamations, and wishes, the finite verb always has the first position. In subordinate clauses, the verb occurs at the very end.

German requires a verbal element (main verb or auxiliary verb) to appear second in the sentence. The verb is preceded by the topic of the sentence. The element in focus appears at the end of the sentence. For a sentence without an auxiliary, these are several possibilities:

- *Der alte Mann gab mir gestern das Buch.* (The old man gave me yesterday the book; normal order)
- *Das Buch gab mir gestern der alte Mann.* (The book gave [to] me yesterday the old man)
- *Das Buch gab der alte Mann mir gestern.* (The book gave the old man [to] me yesterday)
- *Das Buch gab mir der alte Mann gestern.* (The book gave [to] me the old man yesterday)
- *Gestern gab mir der alte Mann das Buch.* (Yesterday gave [to] me the old man the book, normal order)
- *Mir gab der alte Mann das Buch gestern.* ([To] me gave the old man the book yesterday (entailing: as for someone else, it was another date))

The position of a noun in a German sentence has no bearing on its being a subject, an object or another argument. In a declarative sentence in English, if the subject does not occur before the predicate, the sentence could well be misunderstood.

However, German's flexible word order allows one to emphasise specific words:

Normal word order:

- *Der Direktor betrat gestern um 10 Uhr mit einem Schirm in der Hand sein Büro.*
- The manager entered yesterday at 10 o'clock with an umbrella in the hand his office.

Object in front:

- *Sein Büro betrat der Direktor gestern um 10 Uhr mit einem Schirm in der Hand.*
- His office entered the manager yesterday at 10 o'clock with an umbrella in the hand.
- The object *Sein Büro* (his office) is thus highlighted; it could be the topic of the next sentence.

Adverb of time in front:

- *Gestern betrat der Direktor um 10 Uhr mit einem Schirm in der Hand sein Büro. (aber heute ohne Schirm)*
- Yesterday entered the manager at 10 o'clock with an umbrella in the hand his office. (but today without umbrella)

Both time expressions in front:

- *Gestern um 10 Uhr betrat der Direktor mit einem Schirm in der Hand sein Büro.*
- Yesterday at 10 o'clock entered the manager with an umbrella in the hand his office.
- The full-time specification *Gestern um 10 Uhr* is highlighted.

Another possibility:

- *Gestern um 10 Uhr betrat der Direktor sein Büro mit einem Schirm in der Hand.*
- Yesterday at 10 o'clock the manager entered his office with an umbrella in the hand.

- Both the time specification and the fact he carried an umbrella are accentuated.

Swapped adverbs:

- *Der Direktor betrat mit einem Schirm in der Hand gestern um 10 Uhr sein Büro.*
- The manager entered with an umbrella in the hand yesterday at 10 o'clock his office.
- The phrase *mit einem Schirm in der Hand* is highlighted.

Swapped object:

- *Der Direktor betrat gestern um 10 Uhr sein Büro mit einem Schirm in der Hand.*
- The manager entered yesterday at 10 o'clock his office with an umbrella in the hand.
- The time specification and the object *sein Büro* (his office) are lightly accentuated.

The flexible word order also allows one to use language "tools" (such as poetic meter and figures of speech) more freely.

Auxiliary verbs

When an auxiliary verb is present, it appears in second position, and the main verb appears at the end. This occurs notably in the creation of the perfect tense. Many word orders are still possible:

- *Der alte Mann hat mir heute das Buch gegeben.* (The old man has me today the book given.)
- *Das Buch hat der alte Mann mir heute gegeben.* (**The book** has the old man me today given.)
- *Heute hat der alte Mann mir das Buch gegeben.* (**Today** has the old man me the book given.)

The main verb may appear in first position to put stress on the action itself. The auxiliary verb is still in second position.

- *Gegeben hat mir der alte Mann das Buch heute.* (**Given** has me the old man the book **today**.) The bare fact that the book has been given is emphasized, as well as 'today'.

Modal verbs

Sentences using modal verbs place the infinitive at the end. For example, the English sentence "Should he go home?" would be rearranged in German to say "Should he (to) home go?" (*Soll er nach Hause gehen?*). Thus, in sentences with several subordinate or relative clauses, the infinitives are clustered at the end. Compare the similar clustering of prepositions in the following (highly contrived) English sentence: "What did you bring that book that I do not like to be read to out of up for?"

Multiple infinitives

German subordinate clauses have all verbs clustered at the end. Given that auxiliaries encode future, passive, modality,

and the perfect, very long chains of verbs at the end of the sentence can occur. In these constructions, the past participle formed with *ge*-is often replaced by the infinitive.

- *Man nimmt an, dass der Deserteur wohl erschossen$_V$worden$_{psv}$sein$_{perf}$soll$_{mod}$*
- One suspects that the deserter probably shot become be should.
- ("It is suspected that the deserter probably had been shot")
- *Er wusste nicht, dass der Agent einen Nachschlüssel hatte machen lassen*
- He knew not that the agent a picklock had make let
- *Er wusste nicht, dass der Agent einen Nachschlüssel machen lassen hatte*
- He knew not that the agent a picklock make let had
- ("He did not know that the agent had had a picklock made")

The order at the end of such strings is subject to variation, but the second one in the last example is unusual.

Vocabulary

Most German vocabulary is derived from the Germanic branch of the Indo-European language family. However, there is a significant amount of loanwords from other languages, in particular Latin, Greek, Italian, French, and most recently English. In the early 19th century, Joachim Heinrich Campe estimated that one fifth of the total German vocabulary was of French or Latin origin.

Latin words were already imported into the predecessor of the German language during the Roman Empire and underwent all the characteristic phonetic changes in German. Their origin is thus no longer recognizable for most speakers (e.g. *Pforte*, *Tafel*, *Mauer*, *Käse*, *Köln* from Latin *porta*, *tabula*, *murus*, *caseus*, *Colonia*). Borrowing from Latin continued after the fall of the Roman Empire during Christianisation, mediated by the church and monasteries. Another important influx of Latin words can be observed during Renaissance humanism. In a scholarly context, the borrowings from Latin have continued until today, in the last few decades often indirectly through borrowings from English. During the 15th to 17th centuries, the influence of Italian was great, leading to many Italian loanwords in the fields of architecture, finance, and music. The influence of the French language in the 17th to 19th centuries resulted in an even greater import of French words. The English influence was already present in the 19th century, but it did not become dominant until the second half of the 20th century.

Thus, Notker Labeo was able to translate Aristotelian treatises into pure (Old High) German in the decades after the year 1000. The tradition of loan translation was revitalized in the 18th century with linguists like Joachim Heinrich Campe, who introduced close to 300 words that are still used in modern German. Even today, there are movements that try to promote the *Ersatz* (substitution) of foreign words that are deemed unnecessary with German alternatives.

As in English, there are many pairs of synonyms due to the enrichment of the Germanic vocabulary with loanwords from Latin and Latinized Greek. These words often have different

connotations from their Germanic counterparts and are usually perceived as more scholarly.

- *Historie, historisch* – "history, historical", (*Geschichte, geschichtlich*)
- *Humanität, human* – "humaneness, humane", (*Menschlichkeit, menschlich*)
- *Millennium* – "millennium", (*Jahrtausend*)
- *Perzeption* – "perception", (*Wahrnehmung*)
- *Vokabular* – "vocabulary", (*Wortschatz*)
- *Diktionär* – "dictionary, wordbook", (*Wörterbuch*)
- *probieren* – "to try", (*versuchen*)

The size of the vocabulary of German is difficult to estimate. The *Deutsches Wörterbuch* (*German Dictionary*) initiated by Jacob and Wilhelm Grimm already contained over 330,000 headwords in its first edition. The modern German scientific vocabulary is estimated at nine million words and word groups (based on the analysis of 35 million sentences of a corpus in Leipzig, which as of July 2003 included 500 million words in total).

The Duden is the *de facto* official dictionary of the German language, first published by Konrad Duden in 1880. The Duden is updated regularly, with new editions appearing every four or five years. As of August 2017, it was in its 27th edition and in 12 volumes, each covering different aspects such as loanwords, etymology, pronunciation, synonyms, and so forth.

The first of these volumes, *Die deutsche Rechtschreibung* (German Orthography), has long been the prescriptive source for the spelling of German. The *Duden* has become the bible of

the German language, being the definitive set of rules regarding grammar, spelling, and usage of German.

The *Österreichisches Wörterbuch* ("Austrian Dictionary"), abbreviated *ÖWB*, is the official dictionary of the German language in the Republic of Austria. It is edited by a group of linguists under the authority of the Austrian Federal Ministry of Education, Arts and Culture (German: *Bundesministerium für Unterricht, Kunst und Kultur*). It is the Austrian counterpart to the German *Duden* and contains a number of terms unique to Austrian German or more frequently used or differently pronounced there. A considerable amount of this "Austrian" vocabulary is also common in Southern Germany, especially Bavaria, and some of it is used in Switzerland as well. Since the 39th edition in 2001 the orthography of the *ÖWB* has been adjusted to the German spelling reform of 1996. The dictionary is also officially used in the Italian province of South Tyrol.

Orthography

German is written in the Latin alphabet. In addition to the 26 standard letters, German has three vowels with an umlaut mark, namely *ä*, *ö* and *ü*, as well as the eszett or *scharfes s* (sharp s): *ß*. In Switzerland and Liechtenstein, *ss* is used instead of *ß*. Since *ß* can never occur at the beginning of a word, it has no traditional uppercase form.

Written texts in German are easily recognisable as such by distinguishing features such as umlauts and certain orthographical features – German is the only major language that capitalizes all nouns, a relic of a widespread practice in

Northern Europe in the early modern era (including English for a while, in the 1700s) – and the frequent occurrence of long compounds. Because legibility and convenience set certain boundaries, compounds consisting of more than three or four nouns are almost exclusively found in humorous contexts. (In contrast, although English can also string nouns together, it usually separates the nouns with spaces. For example, "toilet bowl cleaner".)

Present

Before the German orthography reform of 1996, *ß* replaced *ss* after long vowels and diphthongs and before consonants, word-, or partial-word endings. In reformed spelling, *ß* replaces *ss* only after long vowels and diphthongs.

Since there is no traditional capital form of *ß*, it was replaced by *SS* when capitalization was required. For example, *Maßband* (tape measure) became *MASSBAND* in capitals. An exception was the use of *ß* in legal documents and forms when capitalizing names. To avoid confusion with similar names, lower case *ß* was maintained (thus "*KREßLEIN*" instead of "*KRESSLEIN*"). Capital *ß* (ẞ) was ultimately adopted into German orthography in 2017, ending a long orthographic debate (thus "*KREẞLEIN* and *KRESSLEIN*").

Umlaut vowels (ä, ö, ü) are commonly transcribed with ae, oe, and ue if the umlauts are not available on the keyboard or other medium used. In the same manner, *ß* can be transcribed as ss. Some operating systems use key sequences to extend the set of possible characters to include, amongst other things, umlauts; in Microsoft Windows this is done using Alt codes.

German readers understand these transcriptions (although they appear unusual), but they are avoided if the regular umlauts are available, because they are a makeshift and not proper spelling. (In Westphalia and Schleswig-Holstein, city and family names exist where the extra e has a vowel lengthening effect, e.g. *Raesfeld*[ˈraːsfɛlt], *Coesfeld*[ˈkoːsfɛlt] and *Itzehoe*[ɪtsəˈhoː], but this use of the letter e after a/o/u does not occur in the present-day spelling of words other than proper nouns.)

There is no general agreement on where letters with umlauts occur in the sorting sequence. Telephone directories treat them by replacing them with the base vowel followed by an e. Some dictionaries sort each umlauted vowel as a separate letter after the base vowel, but more commonly words with umlauts are ordered immediately after the same word without umlauts. As an example in a telephone book*Ärzte* occurs after *Adressenverlage* but before *Anlagenbauer* (because Ä is replaced by Ae). In a dictionary *Ärzte* comes after *Arzt*, but in some dictionaries *Ärzte* and all other words starting with *Ä* may occur after all words starting with *A*. In some older dictionaries or indexes, initial *Sch* and *St*are treated as separate letters and are listed as separate entries after *S*, but they are usually treated as S+C+H and S+T.

Written German also typically uses an alternative opening inverted comma (quotation mark) as in „*Guten Morgen!*".

Past

Until the early 20th century, German was printed in blacklettertypefaces (in Fraktur, and in Schwabacher), and

written in corresponding handwriting (for example Kurrent and Sütterlin). These variants of the Latin alphabet are very different from the serif or sans-serifAntiqua typefaces used today, and the handwritten forms in particular are difficult for the untrained to read. The printed forms, however, were claimed by some to be more readable when used for Germanic languages. The Nazis initially promoted Fraktur and Schwabacher because they were consideredAryan, but they abolished them in 1941, claiming that these letters were Jewish. It is believed that the Nazi régime had banned this script, as they realized that Fraktur would inhibit communication in the territories occupied during World War II.

The Fraktur script however remains present in everyday life in pub signs, beer brands and other forms of advertisement, where it is used to convey a certain rusticality and antiquity.

A proper use of the long s (*langes s*), ſ, is essential for writing German text in Fraktur typefaces. Many Antiqua typefaces also include the long s. A specific set of rules applies for the use of long s in German text, but nowadays it is rarely used in Antiqua typesetting. Any lower case "s" at the beginning of a syllable would be a long s, as opposed to a terminal s or short s (the more common variation of the letter s), which marks the end of a syllable; for example, in differentiating between the words *Wachſtube* (guard-house) and *Wachstube* (tube of polish/wax). One can easily decide which "s" to use by appropriate hyphenation, (*Wach-ſtube* vs. *Wachs-tube*). The long s only appears in lower case.

Orthography Reform

The orthography reform of 1996 led to public controversy and considerable dispute. The states (*Bundesländer*) of North Rhine-Westphalia and Bavaria refused to accept it. At one point, the dispute reached the highest court, which quickly dismissed it, claiming that the states had to decide for themselves and that only in schools could the reform be made the official rule – everybody else could continue writing as they had learned it. After 10 years, without any intervention by the federal parliament, a major revision was installed in 2006, just in time for the coming school year. In 2007, some traditional spellings were finally invalidated; however, in 2008, many of the old comma rules were again put in force.

The most noticeable change was probably in the use of the letter *ß*, called *scharfes s* (*Sharp S*) or *ess-zett* (pronounced *ess-tsett*). Traditionally, this letter was used in three situations:

- After a long vowel or vowel combination;
- Before a *t*;
- At the end of a syllable.

Examples are *Füße*, *paßt*, and *daß*. Currently, only the first rule is in effect, making the correct spellings *Füße*, *passt*, and *dass*. The word *Fuß* 'foot' has the letter *ß* because it contains a long vowel, even though that letter occurs at the end of a syllable. The logic of this change is that an 'ß' is a single letter whereas 'ss' are two letters, so the same distinction applies as (for example) between the words *den* and *denn*.

Phonology

Short /ɛ/ is realized as [ɛ] in stressed syllables (including secondary stress), but as [ə] in unstressed syllables. Note that stressed short /ɛ/can be spelled either with *e* or with *ä* (for instance, *hätte* 'would have' and *Kette* 'chain' rhyme). In general, the short vowels are open and the long vowels are close. The one exception is the open /ɛː/ sound of long *Ä*; in some varieties of standard German, /ɛː/ and /eː/ have merged into [eː], removing this anomaly. In that case, pairs like *Bären/Beeren* 'bears/berries' or *Ähre/Ehre* 'spike (of wheat)/honour' become homophonous (see: Captain Bluebear).

In many varieties of standard German, an unstressed /ɛr/is not pronounced [ər] but vocalised to [ɐ].

Whether any particular vowel letter represents the long or short phoneme is not completely predictable, although the following regularities exist:

- If a vowel (other than *i*) is at the end of a syllable or followed by a single consonant, it is usually pronounced long (e.g. *Hof*[hoː f]).
- If a vowel is followed by *h* or if an *i* is followed by an *e*, it is long.
- If the vowel is followed by a double consonant (e.g. *ff*, *ss* or *tt*), *ck*, *tz* or a consonant cluster (e.g. *st* or *nd*), it is nearly always short (e.g. *hoffen*[ˈhɔfən]). Double consonants are used only for this function of marking preceding vowels as short; the consonant itself is never pronounced lengthened or doubled, in

168

other words this is not a feeding order of gemination and then vowel shortening.

Both of these rules have exceptions (e.g. *hat*[hat] "has" is short despite the first rule; *Mond*[moːnt] "moon" is long despite the second rule). For an *i* that is neither in the combination *ie* (making it long) nor followed by a double consonant or cluster (making it short), there is no general rule. In some cases, there are regional differences. In central Germany (Hesse), the *o* in the proper name "Hoffmann" is pronounced long, whereas most other Germans would pronounce it short. The same applies to the *e* in the geographical name "Mecklenburg" for people in that region. The word *Städte* "cities" is pronounced with a short vowel [ˈʃtɛtə] by some (Jan Hofer, ARD Television) and with a long vowel [ˈʃtɛːtə] by others (Marietta Slomka, ZDF Television). Finally, a vowel followed by *ch* can be short (*Fach*[fax] "compartment", *Küche*[ˈkʏçə] "kitchen") or long (*Suche*[ˈzuːxə] "search", *Bücher*[ˈbyːçɐ] "books") almost at random. Thus, *Lache* is homographous between [laːxə]*Lache* "puddle" and [laxə]*Lache* "manner of laughing" (colloquial) or *lache!* "laugh!" (imperative).

Additionally, the digraph *ie* generally represents the phoneme /iː/, which is not a diphthong. In many varieties, an/r/ at the end of a syllable is vocalised. However, a sequence of a vowel followed by such a vocalised /r/ is not a phonemic diphthong: *Bär*[bɛɐ̯] "bear", *er*[eːɐ̯] "he", *wir*[viːɐ̯] "we", *Tor*[toːɐ̯] "gate", *kurz*[kʊɐ̯ts] "short", *Wörter*[vœɐ̯tɐ] "words".

In most varieties of standard German, syllables that begin with a vowel are preceded by a glottal stop[ʔ].

Consonant spellings

- **c** standing by itself is not a German letter. In borrowed words, it is usually pronounced [t͡s] (before ä, äu, e, i, ö, ü, y) or [k] (before a, o, u, and consonants). The combination **ck** is, as in English, used to indicate that the preceding vowel is short.
- **ch** occurs often and is pronounced either [ç] (after ä, ai, äu, e, ei, eu, i, ö, ü and consonants; in the diminutive suffix *-chen*; and at the beginning of a word), [x] (after a, au, o, u), or [k] at the beginning of a word before a, o, u and consonants. **Ch** never occurs at the beginning of an originally German word. In borrowed words with initial Ch before front vowels (*Chemie* "chemistry" etc.), [ç] is considered standard. However, Upper Germans and Franconians (in the geographical sense) replace it with [k], as German as a whole does before darker vowels and consonants such as in *Charakter*, *Christentum*. Middle Germans (except Franconians) will borrow a [□] from the French model. Both consider the other's variant, and Upper Germans also the standard [ç], to be particularly awkward and unusual.
- **dsch** is pronounced [d͡ʒ] (e.g. *Dschungel*/ˈd͡ʒʊŋəl/ "jungle") but appears in a few loanwords only.
- **f** is pronounced [f] as in "*f*ather".
- **h** is pronounced [h] as in "*h*ome" at the beginning of a syllable. After a vowel it is silent and only lengthens the vowel (e.g. *Reh*[ʁeː] = roe deer).

- **j** is pronounced [j] in Germanic words (*Jahr*[jaːɐ̯]) like "y" in "year". In recent loanwords, it follows more or less the respective languages' pronunciations.

- **l** is always pronounced [l], never *[ɫ] (the English "dark L").

- **q**only exists in combination with **u** and is pronounced [kv]. It appears in both Germanic and Latin words (*quer*[kveːɐ̯]; *Qualität*[kvaliˈtɛːt]). But as most words containing q are Latinate, the letter is considerably rarer in German than it is in English.

- **r** is usually pronounced in a guttural fashion (a voiced uvular fricative[ʁ] or uvular trill[ʀ]) in front of a vowel or consonant (*Rasen*[ˈʁaːzən]; *Burg*[bʊʁk]). In spoken German, however, it is commonly vocalised after a vowel (*er* being pronounced rather like [ɛɐ̯] – *Burg*[bʊɐ̯k]). In some varieties, the **r**is pronounced as a "tongue-tip" *r* (the alveolar trill[r]).

- **s** in German is pronounced [z] (as in "zebra") if it forms the syllable onset (e.g. *Sohn*[zoːn]), otherwise [s] (e.g. *Bus*[bʊs]). In Austria, Switzerland, and Southern Germany, [s] occurs at syllable onset as well. A**ss**[s] indicates that the preceding vowel is short. **st** and **sp** at the beginning of words of German origin are pronounced [ʃt] and [ʃp], respectively.

- **ß** (a letter unique to German called *scharfes S* or *Eszett*) is a ligature of a Long S (ſ) *and* a tailed z (ȝ) and is always pronounced[s]. Originating in Blackletter typeface, it traditionally replaced **ss** at the end of a syllable (e.g. *ich muss→ich muß*; *ich müsste→ich müßte*); within a word it contrasts with

ss[s] in indicating that the preceding vowel is long (compare *in Maßen*[ɪn ˈmaːsən] "with moderation" and *in Massen*[ɪn ˈmasən] "in loads"). The use of **ß** has recently been limited by the latest German spelling reform and is no longer used for **ss** after a short vowel (e.g. *ich muß* and *ich müßte* were always pronounced with a short U/Ü); Switzerland and Liechtenstein already abolished it in 1934.

- **sch** is pronounced [ʃ] (like "sh" in "shine").
- **tsch** is pronounced [tʃ] (like "ch" in "cherry")
- **tion** in Latin loanwords is pronounced [tsi̯oːn].
- **th** is found, rarely, in loanwords and is pronounced [t] if the loanword is from Greek, and usually as in the original if the loanword is from English (though some, mostly older, speakers tend to replace the English th-sound with [s]).
- **v** is pronounced [f] in a limited number of words of Germanic origin, such as *Vater*[ˈfaːtɐ], *Vogel* "bird", *von* "from, of", *vor* "before, in front of", *voll* "full" and the prefix *ver-*. It is also usedin loanwords, where it is normally pronounced [v]. This pronunciation is common in words like *Vase*, *Vikar*, *Viktor*, *Viper*, *Ventil*, *vulgär*, and English loanwords; however, pronunciation is [f] by some people in the deep south. The only non-German word in which "v" is always pronounced "f" is *Eva* (Eve).
- **w** is pronounced [v] as in "*v*acation" (e.g. *was*[vas]).
- **y** is pronounced as [y] when long and [ʏ] when short (as in *Hygiene*[hyɡi̯ˈeːnə] ; *Labyrinth*[labyˈʁɪnt] or *Gymnasium*/ɡʏmˈnaːzi̯ʊm/), except in *ay* and *ey* which are both pronounced [aɪ̯]. It is also often used

in loanwords and pronounced as in the original language, like *Style* or *Recycling.*

- **z** is always pronounced [t͡s] (e.g. *zog*[t͡soːk]), except in loanwords. A **tz** indicates that the preceding vowel is short.

Consonant shifts

German does not have any dental fricatives (as English **th**). The **th** sound, which the English language still has, disappeared on the continent in German with the consonant shifts between the 8th and 10th centuries. It is sometimes possible to find parallels between English and German by replacing the English **th** with **d** in German: "Thank" → in German *Dank*, "this" and "that" →*dies* and *das*, "thou" (old 2nd person singular pronoun) →*du*, "think" →*denken*, "thirsty" →*durstig* and many other examples.

Likewise, the **gh** in Germanic English words, pronounced in several different ways in modern English (as an **f** or not at all), can often be linked to German **ch**: "to laugh" →*lachen*, "through" →*durch*, "high" →*hoch,* "naught" →*nichts*, "light" →*leicht* or *Licht*, "sight" →*Sicht*, "daughter" →*Tochter*, "neighbour" →*Nachbar.*

Literature

The German language is used in German literature and can be traced back to the Middle Ages, with the most notable authors of the period being Walther von der Vogelweide and Wolfram

von Eschenbach. The *Nibelungenlied*, whose author remains unknown, is also an important work of the epoch. The fairy tales collected and published by Jacob and Wilhelm Grimm in the 19th century became famous throughout the world.

Reformer and theologian Martin Luther, who was the first to translate the Bible into German, is widely credited for having set the basis for the modern "High German" language. Among the best-known poets and authors in German are Lessing, Goethe, Schiller, Kleist, Hoffmann, Brecht, Heine, and Kafka. Fourteen German-speaking people have won the Nobel Prize in literature: Theodor Mommsen, Rudolf Christoph Eucken, Paul von Heyse, Gerhart Hauptmann, Carl Spitteler, Thomas Mann, Nelly Sachs, Hermann Hesse, Heinrich Böll, Elias Canetti, Günter Grass, Elfriede Jelinek, Herta Müller and Peter Handke, making it the second most awarded linguistic region (together with French) after English.

Organisations

Several organisations promote the use and learning of the German language.

Goethe Institut

The government-backed *Goethe-Institut*, (named after Johann Wolfgang von Goethe) aims to enhance the knowledge of German culture and language within Europe and the rest of the world. This is done by holding exhibitions and conferences with German-related themes, and providing training and guidance in the learning and use of the German language. For

example, the *Goethe-Institut* teaches the *Goethe-Zertifikat* German language qualification.

Deutsche Welle

The German state broadcaster *Deutsche Welle* provides radio and television broadcasts in German and 30 other languages across the globe. Its German language services are spoken slowly and thus tailored for learners. *Deutsche Welle* also provides an e-learning website for teaching German.

Chapter 5

English Language

English is a West Germanic language of the Indo-European language family, originally spoken by the inhabitants of early medieval England. It is named after the Angles, one of the ancient Germanic peoples that migrated to the area of Great Britain that later took their name, England. Both names derive from Anglia, a peninsula on the Baltic Sea which is not to be confused with East Anglia, the Eastern part of England which comprises the counties of Norfolk, Suffolk and Essex. English is most closely related to Frisian and Low Saxon, while its vocabulary has been significantly influenced by other Germanic languages, particularly Old Norse (a North Germanic language), as well as Latin and French.

English has developed over the course of more than 1,400 years. The earliest forms of English, a group of West Germanic (Ingvaeonic) dialects brought to Great Britain by Anglo-Saxon settlers in the 5th century, are collectively calledOld English. Middle English began in the late 11th century with the Norman conquest of England; this was a period in which English was influenced by Old French, in particular through its Old Normandialect. Early Modern English began in the late 15th century with the introduction of the printing press to London, the printing of the King James Bible and the start of the Great Vowel Shift.

Modern English has been spreading around the world since the 17th century by the worldwide influence of the British Empire and the United States. Through all types of printed and

electronic media of these countries, English has become the leading language of international discourse and the *lingua franca* in many regions and professional contexts such as science, navigation and law. Modern English grammar is the result of a gradual change from a typical Indo-Europeandependent-markingpattern, with a rich inflectionalmorphology and relatively free word order, to a mostly analytic pattern with little inflection, a fairly fixedsubject–verb–object word order and a complex syntax. Modern English relies more on auxiliary verbs and word order for the expression of complex tenses, aspect and mood, as well as passive constructions, interrogatives and some negation.

English is the most spoken language in the world and the third-most spoken native language in the world, after Standard Chinese and Spanish. It is the most widely learned second language and is either the official language or one of the official languages in almost 60 sovereign states. There are more people who have learned English as a second language than there are native speakers. As of 2005, it was estimated that there were over 2 billion speakers of English. English is the majority native language in the United States, the United Kingdom, Canada, Australia, New Zealand and Ireland, an official language and the main language of Singapore, and it is widely spoken in some areas of the Caribbean, Africa, South Asia, Southeast Asia and Oceania. It is a co-official language of the United Nations, the European Union and many other world and regional international organisations. It is the most widely spoken Germanic language, accounting for at least 70% of speakers of this Indo-European branch. English speakers are called "Anglophones". There is much variability among the many accents and dialects of English used in different

countries and regions in terms of phonetics and phonology, and sometimes also vocabulary, idioms, grammar, and spelling, but it does not typically prevent understanding by speakers of other dialects and accents, although mutual unintelligibility can occur at extreme ends of the dialect continuum.

Classification

English is an Indo-European language and belongs to the West Germanic group of the Germanic languages. Old English originated from a Germanic tribal and linguistic continuum along the Frisian North Sea coast, whose languages gradually evolved into the Anglic languages in the British Isles, and into the Frisian languages and Low German/Low Saxon on the continent. The Frisian languages, which together with the Anglic languages form the Anglo-Frisian languages, are the closest living relatives of English. Low German/Low Saxon is also closely related, and sometimes English, the Frisian languages, and Low German are grouped together as the Ingvaeonic (North Sea Germanic) languages, though this grouping remains debated. Old English evolved into Middle English, which in turn evolved into Modern English. Particular dialects of Old and Middle English also developed into a number of other Anglic languages, including Scots and the extinct Fingallian and Forth and Bargy (Yola) dialects of Ireland.

Like Icelandic and Faroese, the development of English in the British Isles isolated it from the continental Germanic languages and influences, and it has since diverged considerably. English is not mutually intelligible with any

continental Germanic language, differing in vocabulary, syntax, and phonology, although some of these, such as Dutch or Frisian, do show strong affinities with English, especially with its earlier stages.

Unlike Icelandic and Faroese, which were isolated, the development of English was influenced by a long series of invasions of the British Isles by other peoples and languages, particularly Old Norse and Norman French. These left a profound mark of their own on the language, so that English shows some similarities in vocabulary and grammar with many languages outside its linguistic clades—but it is not mutually intelligible with any of those languages either. Some scholars have argued that English can be considered a mixed language or a creole—a theory called the Middle English creole hypothesis. Although the great influence of these languages on the vocabulary and grammar of Modern English is widely acknowledged, most specialists in language contact do not consider English to be a true mixed language.

English is classified as a Germanic language because it shares innovations with other Germanic languages such as Dutch, German, and Swedish. These shared innovations show that the languages have descended from a single common ancestor called Proto-Germanic. Some shared features of Germanic languages include the division of verbs into strong and weak classes, the use of modal verbs, and the sound changes affecting Proto-Indo-European consonants, known as Grimm's and Verner's laws. English is classified as an Anglo-Frisian language because Frisian and English share other features, such as the palatalisation of consonants that were velar

consonants in Proto-Germanic (see Phonological history of Old English § Palatalization).

History

The earliest form of English is called Old English or Anglo-Saxon (c. year 550–1066). Old English developed from a set of West Germanic dialects, often grouped as Anglo-Frisian or North Sea Germanic, and originally spoken along the coasts of Frisia, Lower Saxony and southern Jutland by Germanic peoples known to the historical record as the Angles, Saxons, and Jutes. From the 5th century, the Anglo-Saxons settled Britain as the Roman economy and administration collapsed. By the 7th century, the Germanic language of the Anglo-Saxons became dominant in Britain, replacing the languages of Roman Britain (43–409): Common Brittonic, a Celtic language, and Latin, brought to Britain by the Roman occupation. *England* and *English* (originally *Ænglaland* and *Ænglisc*) are named after the Angles.

Old English was divided into four dialects: the Anglian dialects (Mercian and Northumbrian) and the Saxon dialects, Kentish and West Saxon. Through the educational reforms of King Alfred in the 9th century and the influence of the kingdom of Wessex, the West Saxon dialect became the standard written variety. The epic poem *Beowulf* is written in West Saxon, and the earliest English poem, *Cædmon's Hymn*, is written in Northumbrian. Modern English developed mainly from Mercian, but the Scots language developed from Northumbrian. A few short inscriptions from the early period of Old English were written using a runic script. By the 6th century, a Latin

alphabetwas adopted, written with half-uncialletterforms. It included the runic letters *wynn⟨ƿ⟩* and *thorn⟨þ⟩*, and the modified Latin letters *eth⟨ð⟩*, and *ash⟨æ⟩*.

Old English is essentially a distinct language from Modern English and is virtually impossible for 21st-century unstudied English-speakers to understand. Its grammar was similar to that of modern German, and its closest relative is Old Frisian. Nouns, adjectives, pronouns, and verbs had many more inflectional endings and forms, and word order was much freer than in Modern English. Modern English has case forms in pronouns (*he*, *him*, *his*) and has a few verb inflections (*speak*, *speaks*, *speaking*, *spoke*, *spoken*), but Old English had case endings in nouns as well, and verbs had more person and number endings.

The translation of Matthew 8:20 from 1000 shows examples of case endings (nominative plural, accusative plural, genitive singular) and a verb ending (present plural):

- *Foxas habbað holu and heofonan fuglas nest*
- Fox-as habb-að hol-u and heofon-an fugl-as nest-∅
- fox-NOM.PL have-PRS.PL hole-ACC.PL and heaven-GEN.SG bird-NOM.PL nest-ACC.PL
- "Foxes have holes and the birds of heaven nests"

Middle English

From the 8th to the 12th century, Old English gradually transformed through language contact into Middle English. Middle English is often arbitrarily defined as beginning with

the conquest of England by William the Conqueror in 1066, but it developed further in the period from 1200 to 1450.

First, the waves of Norse colonisation of northern parts of the British Isles in the 8th and 9th centuries put Old English into intense contact with Old Norse, a North Germanic language. Norse influence was strongest in the north-eastern varieties of Old English spoken in the Danelaw area around York, which was the centre of Norse colonisation; today these features are still particularly present in Scots and Northern English. However the centre of norsified English seems to have been in the Midlands around Lindsey, and after 920 CE when Lindsey was reincorporated into the Anglo-Saxon polity, Norse features spread from there into English varieties that had not been in direct contact with Norse speakers. An element of Norse influence that persists in all English varieties today is the group of pronouns beginning with *th-* (*they, them, their*) which replaced the Anglo-Saxon pronouns with *h-* (*hie, him, hera*).

With the Norman conquest of England in 1066, the now norsified Old English language was subject to contact with Old French, in particular with the Old Norman dialect. The Norman language in England eventually developed into Anglo-Norman. Because Norman was spoken primarily by the elites and nobles, while the lower classes continued speaking Anglo-Saxon (English), the main influence of Norman was the introduction of a wide range of loanwords related to politics, legislation and prestigious social domains. Middle English also greatly simplified the inflectional system, probably in order to reconcile Old Norse and Old English, which were inflectionally different but morphologically similar. The distinction between nominative and accusative cases was lost except in personal

pronouns, the instrumental case was dropped, and the use of the genitive case was limited to indicating possession. The inflectional system regularised many irregular inflectional forms, and gradually simplified the system of agreement, making word order less flexible. In the Wycliffe Bible of the 1380s, the verse Matthew 8:20 was written: *Foxis han dennes, and briddis of heuene han nestis* Here the plural suffix *-n* on the verb *have* is still retained, but none of the case endings on the nouns are present. By the 12th century Middle English was fully developed, integrating both Norse and French features; it continued to be spoken until the transition to early Modern English around 1500. Middle English literature includes Geoffrey Chaucer's *The Canterbury Tales*, and Malory's*Le Morte d'Arthur*. In the Middle English period, the use of regional dialects in writing proliferated, and dialect traits were even used for effect by authors such as Chaucer.

Early Modern English

The next period in the history of English was Early Modern English (1500–1700). Early Modern English was characterised by the Great Vowel Shift (1350–1700), inflectional simplification, and linguistic standardisation.

The Great Vowel Shift affected the stressed long vowels of Middle English. It was a chain shift, meaning that each shift triggered a subsequent shift in the vowel system. Mid and open vowels were raised, and close vowels were broken into diphthongs. For example, the word *bite*was originally pronounced as the word *beet* is today, and the second vowel in the word *about* was pronounced as the word *boot* is today. The Great Vowel Shift explains many irregularities in spelling since

English retains many spellings from Middle English, and it also explains why English vowel letters have very different pronunciations from the same letters in other languages.

English began to rise in prestige, relative to Norman French, during the reign of Henry V. Around 1430, the Court of Chancery in Westminster began using English in its official documents, and a new standard form of Middle English, known as Chancery Standard, developed from the dialects of London and the East Midlands. In 1476, William Caxton introduced the printing press to England and began publishing the first printed books in London, expanding the influence of this form of English. Literature from the Early Modern period includes the works of William Shakespeare and the translation of the Bible commissioned by King James I. Even after the vowel shift the language still sounded different from Modern English: for example, the consonant clusters/kn ɡn sw/ in *knight*, *gnat*, and *sword* were still pronounced. Many of the grammatical features that a modern reader of Shakespeare might find quaint or archaic represent the distinct characteristics of Early Modern English.

In the 1611 King James Version of the Bible, written in Early Modern English, Matthew 8:20 says, "The Foxes haue holes and the birds of the ayre haue nests." This exemplifies the loss of case and its effects on sentence structure (replacement with subject–verb–object word order, and the use of *of* instead of the non-possessive genitive), and the introduction of loanwords from French (*ayre*) and word replacements (*bird* originally meaning "nestling" had replaced OE *fugol*).

Spread of Modern English

By the late 18th century, the British Empire had spread English through its colonies and geopolitical dominance. Commerce, science and technology, diplomacy, art, and formal education all contributed to English becoming the first truly global language. English also facilitated worldwide international communication. England continued to form new colonies, and these later developed their own norms for speech and writing. English was adopted in parts of North America, parts of Africa, Australasia, and many other regions. When they obtained political independence, some of the newly independent nations that had multiple indigenous languages opted to continue using English as the official language to avoid the political and other difficulties inherent in promoting any one indigenous language above the others. In the 20th century the growing economic and cultural influence of the United States and its status as a superpower following the Second World War has, along with worldwide broadcasting in English by the BBC and other broadcasters, caused the language to spread across the planet much faster. In the 21st century, English is more widely spoken and written than any language has ever been.

As Modern English developed, explicit norms for standard usage were published, and spread through official media such as public education and state-sponsored publications. In 1755Samuel Johnson published his *A Dictionary of the English Language* which introduced standard spellings of words and usage norms. In 1828, Noah Webster published the *American Dictionary of the English language* to try to establish a norm for speaking and writing American English that was independent

of the British standard. Within Britain, non-standard or lower class dialect features were increasingly stigmatised, leading to the quick spread of the prestige varieties among the middle classes.

In modern English, the loss of grammatical case is almost complete (it is now only found in pronouns, such as *he* and *him*, *she* and *her*, *who* and *whom*), and SVO word order is mostly fixed. Some changes, such as the use of do-support, have become universalised. (Earlier English did not use the word "do" as a general auxiliary as Modern English does; at first it was only used in question constructions, and even then was not obligatory. Now, do-support with the verb *have* is becoming increasingly standardised.) The use of progressive forms in *-ing*, appears to be spreading to new constructions, and forms such as *had been being built* are becoming more common. Regularisation of irregular forms also slowly continues (e.g. *dreamed* instead of *dreamt*), and analytical alternatives to inflectional forms are becoming more common (e.g. *more polite* instead of *politer*). British English is also undergoing change under the influence of American English, fuelled by the strong presence of American English in the media and the prestige associated with the US as a world power.

Geographical distribution

As of 2016, 400 million people spoke English as their first language, and 1.1 billion spoke it as a secondary language. English is the largest language by number of speakers. English

is spoken by communities on every continent and on islands in all the major oceans.

The countries where English is spoken can be grouped into different categories according to how English is used in each country. The "inner circle" countries with many native speakers of English share an international standard of written English and jointly influence speech norms for English around the world. English does not belong to just one country, and it does not belong solely to descendants of English settlers. English is an official language of countries populated by few descendants of native speakers of English. It has also become by far the most important language of international communication when people who share no native language meet anywhere in the world.

Three circles of English-speaking countries

The Indian linguist Braj Kachru distinguished countries where English is spoken with a three circles model. In his model,

- the "inner circle" countries have large communities of native speakers of English,
- "outer circle" countries have small communities of native speakers of English but widespread use of English as a second language in education or broadcasting or for local official purposes, and
- "expanding circle" countries are countries where many people learn English as a foreign language.

Kachru based his model on the history of how English spread in different countries, how users acquire English, and the

range of uses English has in each country. The three circles change membership over time.

Countries with large communities of native speakers of English (the inner circle) include Britain, the United States, Australia, Canada, Ireland, and New Zealand, where the majority speaks English, and South Africa, where a significant minority speaks English. The countries with the most native English speakers are, in descending order, the United States (at least 231 million), the United Kingdom (60 million), Canada (19 million), Australia (at least 17 million), South Africa (4.8 million), Ireland (4.2 million), and New Zealand (3.7 million). In these countries, children of native speakers learn English from their parents, and local people who speak other languages and new immigrants learn English to communicate in their neighbourhoods and workplaces. The inner-circle countries provide the base from which English spreads to other countries in the world.

Estimates of the numbers of second language and foreign-language English speakers vary greatly from 470 million to more than 1 billion, depending on how proficiency is defined. Linguist David Crystal estimates that non-native speakers now outnumber native speakers by a ratio of 3 to 1. In Kachru's three-circles model, the "outer circle" countries are countries such as the Philippines, Jamaica, India, Pakistan, Malaysia and Nigeria with a much smaller proportion of native speakers of English but much use of English as a second language for education, government, or domestic business, and its routine use for school instruction and official interactions with the government.

Those countries have millions of native speakers of dialect continua ranging from an English-based creole to a more standard version of English. They have many more speakers of English who acquire English as they grow up through day-to-day use and listening to broadcasting, especially if they attend schools where English is the medium of instruction. Varieties of English learned by non-native speakers born to English-speaking parents may be influenced, especially in their grammar, by the other languages spoken by those learners. Most of those varieties of English include words little used by native speakers of English in the inner-circle countries, and they may show grammatical and phonological differences from inner-circle varieties as well. The standard English of the inner-circle countries is often taken as a norm for use of English in the outer-circle countries.

In the three-circles model, countries such as Poland, China, Brazil, Germany, Japan, Indonesia, Egypt, and other countries where English is taught as a foreign language, make up the "expanding circle". The distinctions between English as a first language, as a second language, and as a foreign language are often debatable and may change in particular countries over time. For example, in the Netherlands and some other countries of Europe, knowledge of English as a second language is nearly universal, with over 80 percent of the population able to use it, and thus English is routinely used to communicate with foreigners and often in higher education. In these countries, although English is not used for government business, its widespread use puts them at the boundary between the "outer circle" and "expanding circle". English is unusual among world languages in how many of its users are

not native speakers but speakers of English as a second or foreign language.

Many users of English in the expanding circle use it to communicate with other people from the expanding circle, so that interaction with native speakers of English plays no part in their decision to use the language. Non-native varieties of English are widely used for international communication, and speakers of one such variety often encounter features of other varieties. Very often today a conversation in English anywhere in the world may include no native speakers of English at all, even while including speakers from several different countries. This is particularly true of the shared vocabulary of the scientific and mathematical fields of life.

Pluricentric English

English is a pluricentric language, which means that no one national authority sets the standard for use of the language. Spoken English, for example English used in broadcasting, generally follows national pronunciation standards that are also established by custom rather than by regulation. International broadcasters are usually identifiable as coming from one country rather than another through their accents, but newsreader scripts are also composed largely in international standard written English. The norms of standard written English are maintained purely by the consensus of educated English-speakers around the world, without any oversight by any government or international organisation.

American listeners generally readily understand most British broadcasting, and British listeners readily understand most

American broadcasting. Most English speakers around the world can understand radio programmes, television programmes, and films from many parts of the English-speaking world. Both standard and non-standard varieties of English can include both formal or informal styles, distinguished by word choice and syntax and use both technical and non-technical registers.

The settlement history of the English-speaking inner circle countries outside Britain helped level dialect distinctions and produce koineised forms of English in South Africa, Australia, and New Zealand. The majority of immigrants to the United States without British ancestry rapidly adopted English after arrival. Now the majority of the United States population are monolingual English speakers, and English has been given official or co-official status by 30 of the 50 state governments, as well as all five territorial governments of the US, though there has never been an official language at the federal level.

English as a global language

English has ceased to be an "English language" in the sense of belonging only to people who are ethnically English. Use of English is growing country-by-country internally and for international communication. Most people learn English for practical rather than ideological reasons. Many speakers of English in Africa have become part of an "Afro-Saxon" language community that unites Africans from different countries.

As decolonisation proceeded throughout the British Empire in the 1950s and 1960s, former colonies often did not reject English but rather continued to use it as independent

countries setting their own language policies. For example, the view of the English language among many Indians has gone from associating it with colonialism to associating it with economic progress, and English continues to be an official language of India. English is also widely used in media and literature, and the number of English language books published annually in India is the third largest in the world after the US and UK. However English is rarely spoken as a first language, numbering only around a couple hundred-thousand people, and less than 5% of the population speak fluent English in India. David Crystal claimed in 2004 that, combining native and non-native speakers, India now has more people who speak or understand English than any other country in the world, but the number of English speakers in India is very uncertain, with most scholars concluding that the United States still has more speakers of English than India.

Modern English, sometimes described as the first global lingua franca, is also regarded as the first world language. English is the world's most widely used language in newspaper publishing, book publishing, international telecommunications, scientific publishing, international trade, mass entertainment, and diplomacy. English is, by international treaty, the basis for the required controlled natural languages Seaspeak and Airspeak, used as international languages of seafaring and aviation. English used to have parity with French and German in scientific research, but now it dominates that field. It achieved parity with French as a language of diplomacy at the Treaty of Versailles negotiations in 1919. By the time of the foundation of the United Nations at the end of World War II, English had become pre-eminent and is now the main worldwide language of

diplomacy and international relations. It is one of six official languages of the United Nations. Many other worldwide international organisations, including the International Olympic Committee, specify English as a working language or official language of the organisation.

Many regional international organisations such as the European Free Trade Association, Association of Southeast Asian Nations (ASEAN), and Asia-Pacific Economic Cooperation (APEC) set English as their organisation's sole working language even though most members are not countries with a majority of native English speakers. While the European Union (EU) allows member states to designate any of the national languages as an official language of the Union, in practice English is the main working language of EU organisations.

Although in most countries English is not an official language, it is currently the language most often taught as a foreign language. In the countries of the EU, English is the most widely spoken foreign language in nineteen of the twenty-five member states where it is not an official language (that is, the countries other than Ireland and Malta). In a 2012 official Eurobarometer poll (conducted when the UK was still a member of the EU), 38 percent of the EU respondents outside the countries where English is an official language said they could speak English well enough to have a conversation in that language. The next most commonly mentioned foreign language, French (which is the most widely known foreign language in the UK and Ireland), could be used in conversation by 12 percent of respondents.

A working knowledge of English has become a requirement in a number of occupations and professions such as medicine and computing. English has become so important in scientific publishing that more than 80 percent of all scientific journal articles indexed by *Chemical Abstracts* in 1998 were written in English, as were 90 percent of all articles in natural science publications by 1996 and 82 percent of articles in humanities publications by 1995.

International communities such as international business people may use English as an auxiliary language, with an emphasis on vocabulary suitable for their domain of interest. This has led some scholars to develop the study of English as an auxiliary language. The trademarked Globish uses a relatively small subset of English vocabulary (about 1500 words, designed to represent the highest use in international business English) in combination with the standard English grammar. Other examples include Simple English.

The increased use of the English language globally has had an effect on other languages, leading to some English words being assimilated into the vocabularies of other languages. This influence of English has led to concerns about language death, and to claims of linguistic imperialism, and has provoked resistance to the spread of English; however the number of speakers continues to increase because many people around the world think that English provides them with opportunities for better employment and improved lives.

Although some scholars mention a possibility of future divergence of English dialects into mutually unintelligible languages, most think a more likely outcome is that English

will continue to function as a koineised language in which the standard form unifies speakers from around the world. English is used as the language for wider communication in countries around the world. Thus English has grown in worldwide use much more than any constructed language proposed as an international auxiliary language, including Esperanto.

Phonology

The phonetics and phonology of the English language differ from one dialect to another, usually without interfering with mutual communication. Phonological variation affects the inventory of phonemes (i.e. speech sounds that distinguish meaning), and phonetic variation consists in differences in pronunciation of the phonemes. This overview mainly describes the standard pronunciations of the United Kingdom and the United States: Received Pronunciation (RP) and General American (GA). (See § Dialects, accents, and varieties, below.)

The phonetic symbols used below are from the International Phonetic Alphabet (IPA).

Phonotactics

An English syllable includes a syllable nucleus consisting of a vowel sound. Syllable onset and coda (start and end) are optional. A syllable can start with up to three consonant sounds, as in *sprint* /sprɪnt/, and end with up to four, as in *texts* /teksts/. This gives an English syllable the following structure, (CCC)V(CCCC) where C represents a consonant and

V a vowel; the word *strengths*/strɛŋkθs/ is thus an example of the most complex syllable possible in English. The consonants that may appear together in onsets or codas are restricted, as is the order in which they may appear. Onsets can only have four types of consonant clusters: a stop and approximant, as in *play*; a voiceless fricative and approximant, as in *fly* or *sly*; *s* and a voiceless stop, as in *stay*; and *s*, a voiceless stop, and an approximant, as in *string*. Clusters of nasal and stop are only allowed in codas. Clusters of obstruents always agree in voicing, and clusters of sibilants and of plosives with the same point of articulation are prohibited. Furthermore, several consonants have limited distributions: /h/ can only occur in syllable-initial position, and /ŋ/ only in syllable-final position.

Stress, rhythm and intonation

Stress plays an important role in English. Certain syllablesare stressed, while others are unstressed. Stress is a combination of duration, intensity, vowel quality, and sometimes changes in pitch. Stressed syllables are pronounced longer and louder than unstressed syllables, and vowels in unstressed syllables are frequently reduced while vowels in stressed syllables are not. Some words, primarily short function words but also some modal verbs such as *can*, have weak and strong forms depending on whether they occur in stressed or non-stressed position within a sentence.

Stress in English is phonemic, and some pairs of words are distinguished by stress. For instance, the word *contract* is stressed on the first syllable (/ˈkɒntrækt/*KON-trakt*) when used as a noun, but on the last syllable (/kənˈtrækt/*kən-*

TRAKT) for most meanings (for example, "reduce in size") when used as a verb. Here stress is connected to vowel reduction: in the noun "contract" the first syllable is stressed and has the unreduced vowel /ɒ/, but in the verb "contract" the first syllable is unstressed and its vowel is reduced to /ə/. Stress is also used to distinguish between words and phrases, so that a compound word receives a single stress unit, but the corresponding phrase has two: e.g. *a burnout* (/ˈbɜːrnaʊt/) versus *to burn out* (/ˈbɜːrnˈaʊt/), and *a hotdog* (/ˈhɒtdɒg/) versus *a hot dog* (/ˈhɒtˈdɒg/).

In terms of rhythm, English is generally described as a stress-timed language, meaning that the amount of time between stressed syllables tends to be equal. Stressed syllables are pronounced longer, but unstressed syllables (syllables between stresses) are shortened. Vowels in unstressed syllables are shortened as well, and vowel shortening causes changes in vowel quality: vowel reduction.

Varieties of English vary the most in pronunciation of vowels. The best known national varieties used as standards for education in non-English-speaking countries are British (BrE) and American (AmE). Countries such as Canada, Australia, Ireland, New Zealand and South Africa have their own standard varieties which are less often used as standards for education internationally. Some differences between the various dialects are shown in the table "Varieties of Standard English and their features".

English has undergone many historical sound changes, some of them affecting all varieties, and others affecting only a few. Most standard varieties are affected by the Great Vowel Shift,

which changed the pronunciation of long vowels, but a few dialects have slightly different results. In North America, a number of chain shifts such as the Northern Cities Vowel Shift and Canadian Shift have produced very different vowel landscapes in some regional accents.

Some dialects have fewer or more consonant phonemes and phones than the standard varieties. Some conservative varieties like Scottish English have a voiceless[ʍ] sound in *whine* that contrasts with the voiced [w] in *wine*, but most other dialects pronounce both words with voiced [w], a dialect feature called *wine–whine* merger. The unvoiced velar fricative sound /x/is found in Scottish English, which distinguishes *loch*/lɔx/ from *lock*/lɔk/. Accents like Cockney with "*h*-dropping" lack the glottal fricative /h/, and dialects with *th*-stopping and *th*-fronting like African American Vernacular and Estuary English do not have the dental fricatives /θ, ð/, but replace them with dental or alveolar stops /t, d/ or labiodental fricatives /f, v/. Other changes affecting the phonology of local varieties are processes such as *yod*-dropping, *yod*-coalescence, and reduction of consonant clusters.

General American and Received Pronunciation vary in their pronunciation of historical /r/ after a vowel at the end of a syllable (in the syllable coda). GA is a rhotic dialect, meaning that it pronounces /r/ at the end of a syllable, but RP is non-rhotic, meaning that it loses /r/ in that position. English dialects are classified as rhotic or non-rhotic depending on whether they elide /r/ like RP or keep it like GA.

There is complex dialectal variation in words with the open front and open back vowels/æ ɑːɒɔː/. These four vowels are

only distinguished in RP, Australia, New Zealand and South Africa. In GA, these vowels merge to three /æ ɑɔ/, and in Canadian English, they merge to two /æ ɑ/. In addition, the words that have each vowel vary by dialect. The table "Dialects and open vowels" shows this variation with lexical sets in which these sounds occur.

Grammar

As is typical of an Indo-European language, English follows accusativemorphosyntactic alignment. Unlike other Indo-European languages though, English has largely abandoned the inflectional case system in favor of analytic constructions. Only the personal pronouns retain morphological case more strongly than any other word class. English distinguishes at least seven major word classes: verbs, nouns, adjectives, adverbs, determiners (including articles), prepositions, and conjunctions. Some analyses add pronouns as a class separate from nouns, and subdivide conjunctions into subordinators and coordinators, and add the class of interjections. English also has a rich set of auxiliary verbs, such as *have* and *do*, expressing the categories of mood and aspect. Questions are marked by do-support, wh-movement (fronting of question words beginning with *wh*-) and word order inversion with some verbs.

Some traits typical of Germanic languages persist in English, such as the distinction between irregularly inflected strong stems inflected through ablaut (i.e. changing the vowel of the stem, as in the pairs *speak/spoke* and *foot/feet*) and weak stems inflected through affixation (such as *love/loved*,

hand/hands). Vestiges of the case and gender system are found in the pronoun system (*he/him, who/whom*) and in the inflection of the copula verb *to be*.

The seven word-classes are exemplified in this sample sentence:

Nouns and noun phrases

English nouns are only inflected for number and possession. New nouns can be formed through derivation or compounding. They are semantically divided into proper nouns (names) and common nouns. Common nouns are in turn divided into concrete and abstract nouns, and grammatically into count nouns and mass nouns.

Most count nouns are inflected for plural number through the use of the plural suffix *-s*, but a few nouns have irregular plural forms. Mass nouns can only be pluralised through the use of a count noun classifier, e.g. *one loaf of bread, two loaves of bread.*

Regular plural formation:

- Singular: *cat, dog*
- Plural: *cats, dogs*

Irregular plural formation:

- Singular: *man, woman, foot, fish, ox, knife, mouse*
- Plural: *men, women, feet, fish, oxen, knives, mice*

Possession can be expressed either by the possessive enclitic *-s* (also traditionally called a *genitive* suffix), or by the preposition *of*. Historically the -s possessive has been used for animate nouns, whereas the *of* possessive has been reserved for inanimate nouns. Today this distinction is less clear, and many speakers use *-s* also with inanimates. Orthographically the possessive -s is separated from a singular noun with an apostrophe. If the noun is plural formed with -s the apostrophe follows the -s.

Possessive constructions:

- With -s: *The woman's husband's child*
- With of: *The child of the husband of the woman*

Nouns can form noun phrases (NPs) where they are the syntactic head of the words that depend on them such as determiners, quantifiers, conjunctions or adjectives. Noun phrases can be short, such as *the man*, composed only of a determiner and a noun. They can also include modifiers such as adjectives (e.g. *red*, *tall*, *all*) and specifiers such as determiners (e.g. *the*, *that*). But they can also tie together several nouns into a single long NP, using conjunctions such as *and*, or prepositions such as *with*, e.g. *the tall man with the long red trousers and his skinny wife with the spectacles* (this NP uses conjunctions, prepositions, specifiers, and modifiers). Regardless of length, an NP functions as a syntactic unit. For example, the possessive enclitic can, in cases which do not lead to ambiguity, follow the entire noun phrase, as in *The President of India's wife*, where the enclitic follows *India* and not *President*.

The class of determiners is used to specify the noun they precede in terms of definiteness, where *the* marks a definite noun and *a* or *an* an indefinite one. A definite noun is assumed by the speaker to be already known by the interlocutor, whereas an indefinite noun is not specified as being previously known. Quantifiers, which include *one*, *many*, *some* and *all*, are used to specify the noun in terms of quantity or number. The noun must agree with the number of the determiner, e.g. *one man* (sg.) but *all men* (pl.). Determiners are the first constituents in a noun phrase.

Adjectives

Adjectives modify a noun by providing additional information about their referents. In English, adjectives come before the nouns they modify and after determiners. In Modern English, adjectives are not inflected so as to agree in form with the noun they modify, as adjectives in most other Indo-European languages do. For example, in the phrases *the slender boy*, and *many slender girls*, the adjective *slender* does not change form to agree with either the number or gender of the noun.

Some adjectives are inflected for degree of comparison, with the positive degree unmarked, the suffix *-er* marking the comparative, and *-est* marking the superlative: *a small boy*, *the boy is smaller than the girl*, *that boy is the smallest*. Some adjectives have irregular comparative and superlative forms, such as *good*, *better*, and *best*. Other adjectives have comparatives formed by periphrastic constructions, with the adverb *more* marking the comparative, and *most* marking the superlative: *happier* or *more happy*, *the happiest* or *most*

happy. There is some variation among speakers regarding which adjectives use inflected or periphrastic comparison, and some studies have shown a tendency for the periphrastic forms to become more common at the expense of the inflected form.

Pronouns, case, and person

English pronouns conserve many traits of case and gender inflection. The personal pronouns retain a difference between subjective and objective case in most persons (*I/me, he/him, she/her, we/us, they/them*) as well as a gender and animateness distinction in the third person singular (distinguishing *he/she/it*). The subjective case corresponds to the Old English nominative case, and the objective case is used in the sense both of the previous accusative case (for a patient, or direct object of a transitive verb), and of the Old English dative case (for a recipient or indirect object of a transitive verb). The subjective is used when the pronoun is the subject of a finite clause, otherwise the objective is used. While grammarians such as Henry Sweet and Otto Jespersen noted that the English cases did not correspond to the traditional Latin-based system, some contemporary grammars, for example Huddleston & Pullum (2002), retain traditional labels for the cases, calling them nominative and accusative cases respectively.

Possessive pronouns exist in dependent and independent forms; the dependent form functions as a determiner specifying a noun (as in *my chair*), while the independent form can stand alone as if it were a noun (e.g. *the chair is mine*). The English system of grammatical person no longer has a distinction

between formal and informal pronouns of address (the old 2nd person singular familiar pronoun *thou* acquired a pejorative or inferior tinge of meaning and was abandoned), and the forms for 2nd person plural and singular are identical except in the reflexive form. Some dialects have introduced innovative 2nd person plural pronouns such as *y'all* found in Southern American English and African American (Vernacular) English or *youse* found in Australian English and *ye* in Hiberno-English.

Pronouns are used to refer to entities deictically or anaphorically. A deictic pronoun points to some person or object by identifying it relative to the speech situation—for example, the pronoun *I* identifies the speaker, and the pronoun *you*, the addressee. Anaphoric pronouns such as *that* refer back to an entity already mentioned or assumed by the speaker to be known by the audience, for example in the sentence *I already told you that*. The reflexive pronouns are used when the oblique argument is identical to the subject of a phrase (e.g. "he sent it to himself" or "she braced herself for impact").

Prepositions

Prepositional phrases (PP) are phrases composed of a preposition and one or more nouns, e.g. *with the dog*, *for my friend*, *to school*, *in England*. Prepositions have a wide range of uses in English. They are used to describe movement, place, and other relations between different entities, but they also have many syntactic uses such as introducing complement clauses and oblique arguments of verbs. For example, in the phrase *I gave it to him*, the preposition *to* marks the recipient,

or Indirect Object of the verb *to give.* Traditionally words were only considered prepositions if they governed the case of the noun they preceded, for example causing the pronouns to use the objective rather than subjective form, "with her", "to me", "for us". But some contemporary grammars such as that of Huddleston & Pullum (2002:598–600) no longer consider government of case to be the defining feature of the class of prepositions, rather defining prepositions as words that can function as the heads of prepositional phrases.

Verbs and verb phrases

English verbs are inflected for tense and aspect and marked for agreement with present-tense third-person singular subject. Only the copula verb *to be* is still inflected for agreement with the plural and first and second person subjects. Auxiliary verbs such as *have* and *be*are paired with verbs in the infinitive, past, or progressive forms. They form complex tenses, aspects, and moods. Auxiliary verbs differ from other verbs in that they can be followed by the negation, and in that they can occur as the first constituent in a question sentence.

Most verbs have six inflectional forms. The primary forms are a plain present, a third-person singular present, and a preterite (past) form. The secondary forms are a plain form used for the infinitive, a gerund-participle and a past participle. The copula verb *to be* is the only verb to retain some of its original conjugation, and takes different inflectional forms depending on the subject. The first-person present-tense form is *am*, the third person singular form is *is*, and the form *are* is used in the second-person singular and all three plurals. The only verb past participle is *been* and its gerund-participle is *being*.

Further aspectual distinctions are shown by auxiliary verbs, primarily *have* and *be*, which show the contrast between a perfect and non-perfect past tense (*I have run* vs. *I was running*), and compound tenses such as preterite perfect (*I had been running*) and present perfect (*I have been running*).

For the expression of mood, English uses a number of modal auxiliaries, such as *can*, *may*, *will*, *shall* and the past tense forms *could*, *might*, *would*, *should*. There are also subjunctive and imperative moods, both based on the plain form of the verb (i.e. without the third person singular *-s*), for use in subordinate clauses (e.g. subjunctive: *It is important that he run every day*; imperative *Run!*).

An infinitive form, that uses the plain form of the verb and the preposition *to*, is used for verbal clauses that are syntactically subordinate to a finite verbal clause. Finite verbal clauses are those that are formed around a verb in the present or preterite form. In clauses with auxiliary verbs, they are the finite verbs and the main verb is treated as a subordinate clause. For example, *he has to go* where only the auxiliary verb *have* is inflected for time and the main verb *to go* is in the infinitive, or in a complement clause such as *I saw him leave*, where the main verb is *to see* which is in a preterite form, and *leave* is in the infinitive.

Phrasal verbs

English also makes frequent use of constructions traditionally called phrasal verbs, verb phrases that are made up of a verb root and a preposition or particle which follows the verb. The

phrase then functions as a single predicate. In terms of intonation the preposition is fused to the verb, but in writing it is written as a separate word. Examples of phrasal verbs are *to get up*, *to ask out*, *to back up*, *to give up*, *to get together*, *to hang out*, *to put up with*, etc. The phrasal verb frequently has a highly idiomatic meaning that is more specialised and restricted than what can be simply extrapolated from the combination of verb and preposition complement (e.g. *lay off* meaning *terminate someone's employment*). In spite of the idiomatic meaning, some grammarians, including Huddleston & Pullum (2002:274), do not consider this type of construction to form a syntactic constituent and hence refrain from using the term "phrasal verb". Instead, they consider the construction simply to be a verb with a prepositional phrase as its syntactic complement, i.e. *he woke up in the morning* and *he ran up in the mountains* are syntactically equivalent.

Adverbs

The function of adverbs is to modify the action or event described by the verb by providing additional information about the manner in which it occurs. Many adverbs are derived from adjectives by appending the suffix *-ly*. For example, in the phrase *the woman walked quickly*, the adverb *quickly* is derived in this way from the adjective *quick*. Some commonly used adjectives have irregular adverbial forms, such as *good* which has the adverbial form *well*.

Clause syntax

In English a sentence may be composed of one or more clauses, that may, in turn, be composed of one or more phrases (e.g. Noun Phrases, Verb Phrases, and Prepositional Phrases). A clause is built around a verb and includes its constituents, such as any NPs and PPs. Within a sentence, there is always at least one main clause (or matrix clause) whereas other clauses are subordinate to a main clause. Subordinate clauses may function as arguments of the verb in the main clause. For example, in the phrase *I think (that) you are lying*, the main clause is headed by the verb *think*, the subject is *I*, but the object of the phrase is the subordinate clause *(that) you are lying*. The subordinating conjunction *that* shows that the clause that follows is a subordinate clause, but it is often omitted. Relative clauses are clauses that function as a modifier or specifier to some constituent in the main clause: For example, in the sentence *I saw the letter that you received today*, the relative clause *that you received today* specifies the meaning of the word *letter*, the object of the main clause. Relative clauses can be introduced by the pronouns *who*, *whose*, *whom* and *which* as well as by *that* (which can also be omitted.) In contrast to many other Germanic languages there is no major differences between word order in main and subordinate clauses.