This article is about the general language or language group. For the literary standard, see Modern Standard Arabic. For the various vernaculars, see varieties of Arabic. For the small family also encompassing the North Arabian languages, see Arabic languages.


The literary language, called Modern Standard Arabic or Literary Arabic, is the only official form of Arabic. It is used in most written documents as well as in formal spoken occasions, such as lectures and news broadcasts.

Arabic is a Central Semitic language, closely related to Aramaic, Hebrew, Ugaritic and Phoenician. Standard Arabic is distinct from and more conservative than all of the spoken varieties, and the two exist in a state known as diglossia, used side-by-side for different societal functions.

Some of the spoken varieties are mutually unintelligible,[5] both written and orally, and the varieties as a whole constitute a sociolinguistic language. This means that on purely linguistic grounds they would likely be considered to constitute more than one language, but are commonly grouped together as a single language for political or religious reasons (see below). If considered multiple languages, it is unclear how many languages there would be, as the spoken varieties form a dialect chain with no clear boundaries. If Arabic is considered a single language, it is perhaps spoken by as many as 422 million speakers (native and non-native) in the Arab world,[6] making it one of the six most spoken languages in the world. If considered separate languages, the most spoken variety would most likely be Egyptian Arabic[7] with 89 million native speakers[8]—still greater than any other Afroasiatic language. Arabic also is a liturgical language of 1.6 billion Muslims.[9][10] It is one of six official languages of the United Nations.[11]

The modern written language (Modern Standard Arabic) is derived from the language of the Quran (known as Classical Arabic or Quranic Arabic). It is widely taught in schools and universities, and is used to varying degrees in workplaces, government, and the media. The two formal varieties are grouped together as Literary Arabic, which is the official language of 26 states and the liturgical language of Islam. Modern Standard Arabic largely follows the grammatical standards of Quranic 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 post-Quranic era, especially in modern times.

Arabic is the only surviving member of the Ancient North Arabian dialect group attested in pre-Islamic Arabic inscriptions dating back to the 4th century.[12] 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 forms.

Arabic has influenced many languages around the globe throughout its history. Some of the most influenced languages are Persian, Urdu, Kurdish, Somali, Swahili, Bosnian, Kazakh, Bengali, Hindi, Malay, Indonesian, Tigrinya, Pashto, Punjabi, Tagalog, Sindhi and Hausa. During the Middle Ages, Literary 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. Many words of Arabic origin are also found in ancient languages like Latin and Greek. Arabic influence, mainly in vocabulary, is seen in Romance languages, particularly Spanish, Catalan, Galician, Portuguese, and Sicilian, owing to both the proximity of Christian European and Muslim Arab civilizations and 800 years of Arabic culture and language in the Iberian Peninsula, referred to in Arabic as al-Andalus.

Arabic has also borrowed words from many languages, including Hebrew, Aramaic, Greek, Persian and Syriac in early centuries, Kurdish in medieval times and contemporary European languages in modern times, mostly English and French.

1 History

1.1 Pre-Islamic Arabic and Arabic language

Many Semitic languages were spoken in Arabia during antiquity. The first of these languages to share the
The earliest inscriptions of Arabic in the (northern) Arabic script date CE 512-568. The Aramaic alphabet was introduced to the Arab people through traders from the Mediterranean Empire, and the Arab people began using the script during the Christian period in the Middle East. From the Aramaic, the script continued to develop through two separate kingdoms in the region: The Nabataean and the Palmyran. The Arabic script that is widely used today developed from the Nabataean Kingdom’s writing script. While the Nabataean alphabet and writing system met a great deal of the needs, it did not provide letters or symbols for /t/, /h/, /g/, /z/ and /d/, which were not represented by Aramaic script. The Aramaic writing system also only provided fifteen letter shapes for 28 consonants. In order to differentiate between consonants that had the same shape, a system of placing dots around the letters developed. It took over 100 years in order to codify these dots around letters (Bateson, 55). Evidence of the development of codified dots is recorded on numerous different codes and tombstones. In the 8th century, the dots were finally codified enough that all texts used dots with the exception of purely decorative writings that were not meant to be read. In addition to the issue of codifying the dots above letters, there was also the issue of how to represent vowel sounds in Arabic script, a language made up of an all-consonants script (an abjad). During the 7th century a dotting system also developed to mark vowels. Red dots were used to mark vowels while black lines were used to mark consonants. Eventually smaller versions of the letters representing short vowels were placed above consonants in order to indicate that a vowel was present.

1.2 Pre-Islamic poetry and early Islamic literature

The oral poetic tradition had been alive and well for centuries in the Arabian Peninsula before it was eventually recorded. Arab poets blossomed in the 6th century AD but their work was not recorded or written down until the 8th or 9th century AD. There were linguistic oddities in regards to spelling found throughout the poems. The poems had been recorded, but there were different spelling and pronunciation techniques used by different authors when trying to record a poem that had previously only been recited. These differences reflected how different dialects had a large impact on written Arabic and how texts were recorded. To solve this problem, grammarians and scholars asked Bedouins to recite poems in order to hear how they pronounced the poem as their voices were believed to be pure. Consensus was then used to determine the correct pronunciation of a word so that the word could also be spelt correctly. Eventually, scholars and grammarians developed a system for standardizing Classical Arabic so that texts and words would be written in a way that the majority of the population could understand. Muslims believe that the Quran was revealed to Muhammad in 632 CE. Both a formal and informal version of Arabic existed during the Pre-Islamic Period. The informal dialect was used on coin and tomb inscriptions while the formal variety of Arabic was used on letters and contracts. However, contracts were sometimes written in a mixture of Formal and Informal Arabic, reflecting the large influence dialects had on the written language in Pre-Islamic Arabia.

1.3 The Islamic conquests and Arabic language

The Islamic conquests introduced Arabic to new non-Arab regions, such as North Africa, Spain and Persia. As a result, Osmani, Kurdish, Persian, Urdu, Malay, Berber, Swahili and Hausa all adopted some Arabic alphabet into their writing systems and Arabic adopted 12 letters from others (for example, Persian P CH Zh and Ga). These sounds entered Arabic through loanwords. In the 8th century Islamic scholars in Lower Iraq feared the influence that the recently conquered non-Arabs would have on the language and scholars became more conservative and a more standardized Classical Arabic writing system developed. However, non-Arabs had a huge influence on the religious writings at the time since many of the intellectual elite were non-Arab.

1.4 Modern era

During the colonial era, the European powers occupying Arab nations actively encouraged the public spread and use of colloquial Arabic dialects and suppressed the
use and teaching of classical Arabic. This caused great diversity in dialects throughout the near east, northern Africa, and even eastern European dialects of Arabic like Czech Arabic and Slavo-Anderski Arabic.\textsuperscript{[20]} After wiping out a third of the Algerian population between 1830 and 1872 for example, the French then closed all Qur’anic schools and banned public usage of Arabic; Arabic was actually declared a foreign language in 1938 and while about half the population was literate in Arabic at the beginning of French colonization, 90% of the native population was illiterate in both Arabic and French by its end in the 1960s.\textsuperscript{[27]}

2 Classical, Modern Standard and spoken Arabic

See also: List of Arabic dictionaries

Arabic usually designates one of three main variants: Classical Arabic, Modern Standard Arabic and colloquial or dialectal Arabic. Classical Arabic is the language found in the Quran, used from the period of Pre-Islamic Arabia to that of the Abbasid Caliphate. Theoretically, Classical Arabic is considered normative, 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-‘Arab). In practice, however, modern authors almost never write in pure Classical Arabic, instead using a literary language with its own grammatical norms and vocabulary, commonly known as Modern Standard Arabic (MSA).

MSA is the variety used in most current, printed Arabic publications, spoken by some of the Arabic media across North Africa, the Horn of Africa and the Middle East, and understood by most educated Arabic speakers. “Literary Arabic” and “Standard Arabic” ( Phú sūfūhá) 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 dialect (e.g., the energetic mood) are almost never used in Modern Standard Arabic.

- No modern spoken variety of Arabic has case distinctions. 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. In addition, MSA has borrowed or coined a large number of terms for concepts that did not exist in Quranic times, and MSA continues to evolve.\textsuperscript{[28]} Some words have been borrowed from other languages—notice that transliteration mainly indicates spelling and not real pronunciation (e.g., لَفْم ‘film’ or ديموقراطية ‘democracy’).

However, the current preference is to avoid direct borrowings, preferring to either use loan translations (e.g., فرع ‘branch’, also used for the branch of a company or organization; جناح ‘wing’, is also used for the wing of an airplane, building, air force, etc.), or to coin new words using forms within existing roots (ازِتِمَات ‘apoptosis’, using the root موت ‘death’ put into the Xh form, or جامعة ‘university’, based on جمع ‘to gather, unite’; جمهورية ‘republic’, based on جمْهُور ‘multitude’). An earlier tendency was to repurpose older words; that has fallen into disuse (e.g., هاتف ‘telephone’ e ‘invisible caller (in Sufism)’; جريدة ‘newspaper’ < ‘palm-leaf stalk’).

Colloquial or dialectal Arabic refers to the many national or regional varieties which constitute the everyday spoken language. Colloquial Arabic has many regional variants; geographically distant varieties usually differ enough to be mutually unintelligible, and some linguists consider them distinct languages.\textsuperscript{[29]} The varieties are typically unwritten. They are often used in informal spoken media, such as soap operas and talk shows,\textsuperscript{[30]} 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 (predominately Roman 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. Historically, Algerian Arabic was taught in French Algeria under the name ċarjā. 
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).

3 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. 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, similar to the issue with the varieties of Chinese, Hindi and Urdu, Serbian and Croatian, Scots and English, etc. Similar to how speakers of Hindi and Urdu will 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.

4 Influence of Arabic on other languages

Main article: Influence of Arabic on other languages
See also: List of Arabic loanwords in English

The influence of Arabic has been most important in Islamic countries, because it is the language of the Islamic sacred book, Quran. Arabic is also an important source of vocabulary for languages such as Baluchi, Bengali, Berber, Bosnian, Catalan, Chechen, Dagestani, English, French, German, Gujarati, Hausa, Hindustani, Italian, Indonesian, Kazakh, Kurdish, Kutchi, Malay, Malayalam, Pashto, Persian, Portuguese, Punjabi, Rohingya, Saraiki, Sicilian, Sindhi, Somali, Spanish, Swahili, Tagalog, Turkish, Urdu, Uzbek and Wolof, as well as other languages in countries where these
languages are spoken.

In addition, English has many Arabic loanwords, some directly but most through the medium of other Mediterranean languages. Examples of such words include admiral, adobe, alchemy, alcohol, algebra, algorithm, alkaline, almanac, amber, arsenal, assassin, candy, catat, cipher, coffee, cotton, goulash, jar, kismet, lemon, loofah, magazine, mattress, sherbet, sofa, sumac, tariff and many other words. 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 taszālīṭ ‘prayer’ < salat) (صلاة salāḥ), academic terms (like Uyghur menitiq ‘logic’), economic items (like English coffee) to placeholders (like Spanish fulano ‘so-and-so’) and everyday terms (like Hindustani lekin ‘but’, or Spanish taza meaning ‘cup’and hasta meaning ‘until’), 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 entered through Persian though Persian is an Indo-Iranian language. Older Arabic loanwords in Hausa were borrowed from Kanuri.

Some words in English and other European languages are derived from Arabic, often through other European languages, especially Spanish and Italian. Among them are commonly used words like “coffee” (قهوة qahwah), “cotton” (قطن qutn) and “magazine” (مكتبة makhāzin). English words more recognizably of Arabic origin include “algebra”, “alcohol”, “alchemy”, “alkali”, “zenith,” and “nadir”.

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

As, throughout the Islamic world, Arabic occupied a position similar to that of Latin in Europe, many of the Arabic concepts in the field 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 right until the 18th and 19th century, when swaths of Arab-inhabited lands were under Ottoman rule.

5 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, Ethiopian; and to a lesser degree Hebrew (mainly religious concepts). In addition, many cultural, religious and political terms have entered Arabic from Iranian languages, notably Middle Persian, Parthian, and (Classical) Persian, and Hellenistic Greek (κηνιον kēniōν) as in the well-known form “Al-Jazeera,” means “island” and has its origin in the Syriac ķavt.” (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:

- madinah/medina (مدينة, city or city square), a word of Aramaic or Hebrew origin;
- jazīrah (جزيرة), as in the well-known form “Al-Jazeera,” means “island” and has its origin in the Syriac ķavt.
- lājvard (لازورد) is taken from Persian 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 Spanish.

6 Arabic alphabet and nationalism

There have been many instances of national movements to convert Arabic script into Latin script or to Romanize the language.

6.1 Lebanon

The Beirut newspaper La Syrie pushed for the change from Arabic script to Latin script 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.
6.2 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. [44][45] 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. [44][45] 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 script, 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. [44][45] Ahmad Lutfi As Sayid and Muhammad Azmi, two Egyptian intellectuals, agreed with Musa and supported the push for Romanization. [44][46] 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. [44][46] However, this effort failed as the Egyptian people felt a strong cultural tie to the Arabic alphabet. [44][46] In particular, the older Egyptian generations believed that the Arabic alphabet had strong connections to Arab values and history, which is easy to believe due to the long history of the Arabic alphabet (Shrivtiel, 189).

7 Arabic and Islam

Classical Arabic is the language of the Qur'an. Arabic is closely associated with the religion of Islam because the Qur'an is written in the language, but it is nevertheless also spoken by Arab Christians, Mizrahí Jews and Iraqi Mandaeans. Most of the world’s Muslims do not speak 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.

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. [47][48] Judaism has a similar account with the Tower of Babel.

8 External history

See also: Pre-Islamic Arabic inscriptions

Among the earliest surviving texts in Ancient North Arabian, a group of languages closely related to but not a direct predecessor of Arabic, are the Hasean inscriptions of in eastern Arabia, from about the 4th century BC, written not in the modern Arabic alphabet, nor in its Nabataean ancestor, but in variants of the epigraphic South Arabian musnad. More numerous are the 6th-century BC Liyanite texts from southeastern Arabia and the Thamudic texts found throughout Arabia and the Sinai, and not actually connected with Thamud.

Classical Arabic co-existed with the Old North Arabian languages. In the 5th century BC, Herodotus (Histories I,131; III,8) quotes the epithet of a goddess in its pre-classical Arabic form as 'AIlát (Ἀλιλάτ, i. e.,'al-‘ilat), which means “the goddess”. [49] Apart from this isolated theonym, Arabic is first attested in an inscription in Qaryat al-Faw (formerly Qaryat Dhat Kahil, near Sulayyil, Saudi Arabia) in the 1st century BC. [50][51] Later come the Safaitic and Hismaic inscriptions beginning in the 1st century AD (in the South Arabian script) and the many Arabic personal names attested in Nabataean inscriptions (which inscriptions are, however, Aramaic).

The oldest inscription in Classical Arabic known in 1985 goes back to 328 AD and is known as the Namārah inscription, written in the Nabataean alphabet and named after the place where it was found in southern Syria in April 1901. [52] By the fourth century AD, the Arab kingdoms of the Lakhmids in southern Iraq and the
Ghassanids in southern Syria appeared. The Kindite Kingdom emerged in Central Arabia. Their courts were responsible for some notable examples of pre-Islamic Arabic poetry and for some of the few surviving pre-Islamic Arabic inscriptions in the Arabic script.[53]

9 Internal history

The Semitic languages changed a great deal between Proto-Semitic and the establishment of the Central Semitic languages, particularly in terms of grammar. Innovations of the Central Semitic languages—all maintained in Classical Arabic—include

- The conversion of the suffix-conjugated stative formation into a past tense.
- The conversion of the prefix-conjugated preterite-tense formation 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.

9.1 Phonological history

Of the 29 Proto-Semitic consonants, only one has been lost: */ʒ/, which merged with /ʃ/.[54] But the consonant */ɡ/ is still found in many colloquial Arabic dialects. 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[55] - became palatalized to /g/ or /j/ by the time of the Quran and /dʒ/, /ɡ/, /ʒ/ or /ɡ/ in MSA (see Arabic phonology#Local variations for more detail).[56] An original voiceless alveolar lateral fricative */ɦ/ became /f/.[57] Its emphatic counterpart */h/ was considered by Arabs to be the most unusual sound in Arabic (Hence the Classical Arabic’s appellation lughat al-dād or “language of the ḍād”); for most modern dialects, it has become an emphatic stop /dˤ/ with loss of the laterality[57] or with complete loss of any pharyngealization or velarization, /d/. (The classical dād pronunciation of pharyngealization /h/ still occurs in the Mehri language and the similar sound without velarization 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 at the eighth century and glottalized in Proto-Semitic.[57]

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 Bedouin tribes in the eastern part of the Arabian Peninsula, who spoke the most conservative variants of Arabic. Even at the time of Mohammed 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/.


together suggest the existence of the koine:

• Loss of the dual (grammatical 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., *halaltu 'I untied' → *halēt(u).
• Conversion of separate words *li 'to me', *laka 'to you', etc. into indirect-object clitic suffixes.
• Certain changes in the cardinal number system, e.g., *khamsat *ayyām 'five days' → *khām(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' < *ja'a bi- 'come with'; *shāf 'see'; *ēsh 'what' (or similar) < *ayyu *shay 'which thing'; *illi (relative pronoun).
• Merger of *la'y and */θ/.

10.3 Dialect groups

• Egyptian Arabic is spoken by around 53 million in Egypt (55 million worldwide).[59] 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, Palestinian Territories, Israel, Cyprus and Turkey.

Magrebi Arabic, also called "Darija" spoken by about 70 million people in Morocco, Algeria, Tunisia, Libya and Malta. It is very hard to understand for Arabic speakers from the Mashriq or Mesopotamia, the easiest being Libyan Arabic and the hardest Moroccan Arabic and Maltese language (which is close to Tunisian Arabic). The others such as Algerian Arabic can be considered “in between”.

Maltese, spoken on the island of Malta, is the only dialect to have established itself as a fully separate language, with independent literary norms. Sicilian Arabic, spoken on the island of Sicily until the 14th century, developed into Maltese in Malta. In the course of its history the language has adopted numerous loanwords, phonetic and phonological features, and even some grammatical patterns, from Italian, Sicilian and English. It is also the only Semitic language written in the Latin script. Furthermore, Maltese or Sicilian Arabic are closely related to Tunisian Arabic due to the cultural and historical ties between Tunisia and Malta. The languages are partially mutually intelligible.

Mesopotamian Arabic, spoken by about 7 million people in Iraq (where it is called “Aamiyah”), eastern Syria and southwestern Iran (Khuzestan).

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 neighbors to the north; rather, the Sudanese have a dialect similar to the Hijazi dialect.

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 Qatar citizens speak Najdi Arabic (Bedawi).

Yemeni Arabic spoken in Yemen, Somalia, Djibouti 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 Hijaz, western Saudi Arabia

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

Bahraini Arabic (600,000 speakers), spoken by Bahraini Shi’ah in Bahrain and Qatar, 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.

Central Asian Arabic, spoken in Uzbekistan, Tajikistan and Afghanistan, is highly endangered

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

Shirvani Arabic, spoken in Azerbaijan and Dagestan until the 1930s, now extinct.

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

11 Phonology

Main article: Arabic phonology

The “colloquial” spoken varieties 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 broadcasts, formal lectures, parliamentary discussions and to some extent between speakers of different colloquial varieties. Even when the literary language is spoken, however, it is normally only spoken in its pure form when reading a prepared text out loud. 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.

11.1 Literary Arabic

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. Note that 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 Proto-Semitic /g/ has many different pronunciations in the modern spoken varieties, e.g., [d3 ~ 3 ~ j ~ q0 ~ q]s. Speakers whose native variety has either [d3] or [3] will use the same pronunciation when speaking MSA. Even speakers from Cairo, whose native Egyptian Arabic has [q], normally use [q] when speaking MSA. The [j] of Persian Gulf speakers is the only variant pronunciation which isn’t found in MSA; [d3] is used instead.

Another example: Many colloquial varieties are known for a type of vowel harmony in which the presence of an “emphatic consonant” triggers different pronunciations of nearby vowels (especially the low vowels /a:/, which are backed to [a(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.)

11.1.1 Vowels

Modern Standard Arabic has six pure vowels, 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 [æ(ː)] in most situations, but a back [a(ː)] in the neighborhood of emphatic consonants. Some accents and dialects, such as those of the Hijaz, have central [a(ː)] in all situations. The vowels /a/ and /i/ 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 /a/ and /i/ tends towards [iː] and [iː] 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 /h/; /dˤ/; /j/; and /tˤ/, if not followed immediately by /i(ː)/. Frequently, the uvular fricatives /x y/ also trigger emphatic allophones; occasionally also the pharyngeal consonants /h/ (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., /wɔt/ [waqt] ‘time’; /wɔtæn/ [wat’on] ‘homeland’; /wæstˈɑl-mædɪnə/ [wasṭ’al-madīnə] ‘downtown’ (sometimes [wasṭ’al-madī:ni] 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 [æj]: hence صيف [sajf] ‘summer’ but صيف [sajf] ‘sword’; صيف [sajf] ‘summer’. However, in accents with no emphatic allophones of /a/, e.g., in the Hijaz), the pronunciation [aj] occurs in all situations.

11.1.2 Consonants

The phoneme [d–d3–j–q–q’s] is represented by the Arabic letter jīm (ﺟ) and has many standard pronunciations. [d3] is characteristic of north Algeria, Iraq, also in most of the Arabian peninsula but with an allophonic [3] in some positions; [3] occurs in most of the Levant and most 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 [g] or
[11.1 Literary Arabic]

Arabic occur in only two places: at the end of the sentence (due to pausal pronunciation) and in words such as *حَارَ 'hot', مَدَدَ 'stuff, substance', and *تاَهِضَ '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 -نَا '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 *إِسْتَهَارُ 'he bought'. *اجتماع '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., *إِجْتِمَاعُ 'meeting', /ʔa.l.i.jtimaː/.
- If the word occurs after another word ending in a vowel, the initial vowel of the word is elided, e.g., *بيتَ الْمِدَرِيْرِ 'house of the director' /ba.julmudiːr/.
- If the word occurs at the beginning of an utterance, or words such as *إِسْتَهَارُ 'he bought'. or *إِجْتِمَاعُ 'meeting'. or *الْبَيْتِ هَوَاء 'the house is ...', /ʔal.bajtu.wa ... /.

11.1.4 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 *يَنْكَتَبُ 'he subscribed' (whether or not the final short vowel is pronounced),

[11.1.3 Syllable structure]

Arabic has two kinds of syllables: open syllables (CV) and (CVV)—and closed syllables (CVC), (CVVC) and (CVCC). The syllable types with three morae (units of time), i.e. CVC and CVV, are termed heavy syllables, while those with four 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 *حَارَ 'hot', مَدَدَ 'stuff, substance', and *تاَهِضَ '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 -نَا 'us, our', due to the deletion of final short vowels.)
exist:

nouns and verbs. The following levels of pronunciation

language, despite forming part of the formal paradigm of

mood endings

where the following changes occur:

nounced as written, except at the end of an utterance,

mal level actually used in speech. All endings are pro-

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., irmīl 'throw!'‘.)

• The entire indefinite noun endings -in and -un (with

nunciation) are left off. The ending -an is left off of

ouns preceded by a tā marbūṭah s (i.e. the -t in the

nding-at- that typically marks feminine nouns), but

ounced as -ā in other nouns (hence its writing in

his fashion in the Arabic script).

• The tā marbūṭah itself (typically of feminine nouns)

onvoked as h. (At least, this is the case in

tremely formal pronunciation, e.g., some Quranic

citations. In practice, this h is usually omitted.)

Formal short pronunciation

This is a formal level of pronunciation sometimes seen. It is somewhat like prono-
nouncing 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. How-

ever, the following short vowels are pronounced:

• feminine plural -na

• shortened vowels in the jussive/imperative of

defective verbs, e.g., irmi‘l ‘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

ot’, sawfa (future-tense marker)

• The nunation endings -an, -in, -un are not pro-
nounced. However, they are pronounced in adver-

tial accusative formations, e.g., taqrīban ʻādatan

‘almost, approximately’, ʻādatan ʻāda ‘usually’.

• The tā marbūṭah ending ʻs is unpronounced, except

in construct state nouns, where it sounds as f (and in

adverbial accusative constructions, e.g., ʻādatan ʻāda ‘usually’, where the entire -tan is pronounced).

• The masculine singular nisbah ending -iyy is actually

ounced -i and is unstressed (but plural and fem-

ine singular forms, i.e. when followed by a suffix, 

still sound as -iyy-).

• Full endings (including case endings) occur when a

lic object or possessive suffix is added (e.g., -nā

‘us/our’.)


(but mak-ta-ba-tun ‘library’ in short pronunciation), ka-

ta-bā (Modern Standard Arabic) ‘they wrote’ = ka-ta-bu

(dialect), ka-ta-bā-h(u) (Modern Standard Arabic) ‘they

wrote it’ = ka-ta-bā (dialect), ka-ta-ba-tā (Modern Stan-

dard Arabic) ‘they (dual, fem) wrote’, ka-ta-ba-tu (Mod-

ern Standard Arabic) ‘I wrote’ (short form or diale.

) 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

uation where they are not pronounced, as in the above

xample of mak-ta-ba-tun ‘library’ in full pronunciation,

ut 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’, gā-hi-rah ‘Cairo’. This also

affects the way that Modern Standard Arabic is pro-

ounced in Egypt. In the Arabic of Sanaa, stress is of-

ten 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.)

11.1.5 Levels of pronunciation

The final short vowels (e.g., the case endings -a -i -u and

and mood endings -a -a) are often not pronounced in this lan-

guage, despite forming part of the formal paradigm of

ouns and verbs. The following levels of pronunciation exist:

Full pronunciation

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., irmīl ‘throw!’‘.)
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 -tì are pronounced -t -t -tì. But masculine ānta 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 -a-) 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., -rm- is pronounced as a three-consonant cluster, but -rm- needs to be broken up).
  - Or, a short vowel is never added, but consonants like r /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. (But note that 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 -ān vs. oblique -ān) 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 abā 'father of', akhā '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).

11.2 Colloquial varieties

Further information: Varieties of Arabic

11.2.1 Vowels

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 [ɑː] 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 /w/; while in Egyptian Arabic, it usually spreads throughout the entire word, including prefixes and suffixes. In Moroccan Arabic, /i u/ also have emphatic allophones [ʊ~ɔ e~ɛ].

Unstressed short vowels, especially /i u/, are deleted in many contexts. Many sporadic examples of short vowel change have occurred (especially /aU/→/u/ and interchange /i/→/u/). Most Levantine dialects merge short /i u/ into /u/ 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 /a/, which is deleted in many
contexts. (The labialization plus /a/ is sometimes interpreted as an underlying phoneme /ːa/.) This essentially causes the wholesale loss of the short-long vowel distinction, with the original long vowels /aː iː uː/, 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 all circumstances, including adjacent to emphatic consonants). In Moroccan Arabic, these have subsequently merged into original /iː uː/.

11.2.2 Consonants

In some dialects, there may be more or fewer phonemes than those listed in the chart above. For example, non-Arabic [v] is used in the Maghrebi dialects as well in the written language mostly for foreign names. Semitic [p] became [f] extremely early on in Arabic before it was written down; a few modern Arabic dialects, such as Iraqi (influenced by Persian and Kurdish) distinguish between [p] and [b]. The Iraqi Arabic also uses sounds [q], [tʃ] and uses Persian adding letters, e.g.: ُقَوْيَةُ gawjah – a plum; ُقَمْ جُحُم chimah – a truffle and so on.

Early in the expansion of Arabic, the separate emphatic phonemes [lia] and [lia] coalesced into a single phoneme [lia]. Many dialects (such as Egyptian, Levantine, and much of the Maghreb) subsequently lost interdental fricatives, converting [0 δ δ̩] into [t d d̩]. 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 [0 δ δ̩] in borrowed words as [s z z̩ d̩].

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

- /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 [q] 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 [t], as do Shi’i Bahrainis. In some Gulf dialects, it is palatalized to [d̩] or [j]. It is pronounced as a voiced uvular constractive [r] 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.

- /tˤ/ 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] in most words in much of the Persian Gulf.

- /k/ usually retains its original pronunciation, but is palatalized to [q] in many words in Israel and the Palestinian Territories, Iraq, and much of the Arabian Peninsula. Often a distinction is made between the suffixes -ak/ (‘you’, masc.) and -ik/ (‘you’, fem.), which become /-ak/ and /-iʃ/, respectively. In Sana’a, Omani, and Bahraini -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 [a] 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 /tˤ/ vs. /ʔ/ in Moroccan Arabic, because the former is pronounced as an affricate [tʃ] but the latter is not.)

12 Grammar

Main article: Arabic grammar

12.1 Literary Arabic

Main article: Modern Standard 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:
### 12.1 Literary Arabic

#### 12.1.1 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 /-at/, which is reduced to /-ah/ or /-a/ 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 /al-/. Indefinite singular nouns (other than those that end in long ā) add a final /-al/ to the case-marking vowels, giving /-unl/, /-anl/ or /-inl/ (which is also referred to as nunation or tanwin).

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 /-al/ or /-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 (/-ni/) and for nouns or prepositions (/-al/ after consonants, /-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.

#### 12.1.2 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 indicatice occur only in the non-past, and the future tense is signaled by prefixing sa- or sawfa onto the non-past. The past and non-past differ in the form of the stem (e.g., past kataba vs. non-past -ktuba-), 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, kataba 'to write'. Note that in Modern Standard Arabic, many final short vowels are dropped (indicated in parentheses below), and the energetic mood (in either long or short form, which have the same meaning) is almost never used.

### 12.1.3 Derivation

Unlike most languages, Arabic has virtually no means of deriving words by adding prefixes or suffixes to words. Instead, they are formed according to a finite (but fairly large) number of templates applied to roots.

For verbs, a given root can construct up to fifteen different verbs, 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 forms encode concepts such as the causative, intensive and reflexive. These forms can be viewed as analogous to verb conjugations in languages such as Spanish in terms of the additional complexity of verb formation that they induce. (Note, however, that their usage in constructing vocabulary is somewhat different, since the same root can be conjugated in multiple forms, with different shades of meaning.)

Examples of the different verbs formed from the root k-t-b 'write' (using h-m-r 'red' for Form IX, which is limited to colors and physical defects):

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. maktaba 'desk, office' < k-t-b 'write', maṭbakh 'kitchen' < t-b-kh 'cook').

The only three genuine suffixes are as follows:

- The feminine suffix -ah; variously derives terms from women from related terms for men, or more generally terms along the same lines as the corresponding masculine, e.g. makaṭaba 'library' (also a writing-related place, but different from maktab, as above).

- The nisbah suffix -yy-. 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 -iyah. 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 jumḥūriyyah 'republic' (lit. "public-ness", < jumhūr 'multitude, general public'), and the Gaddafi-specific variation jamāhīrīyyah 'people's republic' (lit. "masses-ness", < jamāhīr 'the masses', pl. of jumhūr, as above).

### 12.2 Colloquial varieties

Main article: Varieties of Arabic

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 “nunciation” and the internal passive.

The following is an example of a regular verb paradigm in Egyptian Arabic.
13 Writing system

Main articles: Arabic alphabet and Arabic Braille

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 script, notably naskh, which is used in print and by computers, and ruq‘ah, which is commonly used in handwriting.[67]

13.1 Calligraphy

Main article: Islamic 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.

13.2 Romanization

Main article: Romanization of Arabic

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 ݰ is used to represent both a consonant, as in “you” or “yet”, and a vowel, as in “me” or “eat”.) 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 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 (ض). 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 (ض), may be represented by D. Its emphatic counterpart, (ض), may be written as D.
13.3 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 position is 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, “five and seventy and nine-hundred and a thousand.”

14 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. See also: Arabic Language International Council

15 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. [69]

16 Arabic speakers and other languages

Historically, Arab linguists considered the Arabic language to be superior to all other languages, and took almost no interest in learning any language other than Arabic. With the sole example of Medieval linguist Abu Hayyan al-Gharnati - who, while a scholar of the Arabic language, was not ethnically Arab - scholars of the Arabic language made no efforts at studying comparative linguistics, considering all other languages inferior. [70]

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.” [71] Arab-American professor Franck Salamah went as far as to declare Arabic a dead language conveying dead ideas, blaming its stagnation for Arab intellectual stagnation and lamenting that great writers in Arabic are judged by their command of the language and not the merit of the ideas they express with it. [72]

17 See also

18 References

Notes

[1] “Världens 100 största språk 2010” (The World’s 100 Largest Languages in 2010), in Nationalencyklopedin
[45] Shrivtiel, p. 188

[46] Shrivtiel, p. 189


[49] Woodard, Roger D. Ancient Languages of Syria-Palestine and Arabia. p 208


[54] Lipinski (1997:124)

[55] Al-Jallad, 42


[57] Watson (2002:2)


[60] Borg and Azzopardi-Alexander Maltese (1997:xiii) “The immediate source for the Arabic vernacular spoken in Malta was Muslim Sicily, but its ultimate origin appears to have been Tunisia. In fact Maltese displays some areal traits typical of Maghrebian Arabic, although during the past eight hundred years of independent evolution it has drifted apart from Tunisian Arabic”.


[65] e.g., Thelwall (2003:52)


[67] Hanna & Greis (1972:2)


[71] Suleiman, p. 93


Bibliography

- Badawi, Mohamed; Caroli, Christian A. (2011), As-Sabil: Grundlagen der arabischen Grammatik (in German), Konstanz
- Irwin, Robert (2006), For Lust of Knowing, London: Allen Lane
• Lipinski, Edward (1997), *Semitic Languages*, Leuven: Peeters
• Procházka, S. (2006), "'Arabic'"*, Encyclopedia of Language and Linguistics* (2nd ed.)
• Vaglieri, Laura Veccia, *Grammatica teorico-pratica della lingua araba*, Rome: I.P.O.

19 External links

• Arabic: a Category III language Languages which are difficult for native English speakers.
• Dr. Nizar Habash’s, Columbia University, Introduction to Arabic Natural Language Processing
• Google Ta3reeb – Google Transliteration
• Transliteration Arabic language pronunciation applet
• USA Foreign Service Institute Arabic basic course
• How to speak Arabic
• Alexis Neme (2011), A lexicon of Arabic verbs constructed on the basis of Semitic taxonomy and using finite-state transducers
• Alexis Neme and Eric Laporte (2013), Pattern-and-root inflectional morphology: the Arabic broken plural
• Alexis Neme and Eric Laporte (2015), Do computer scientists deeply understand Arabic morphology? - and how could we ever understand Arabic morphology? available also in Arabic, Indonesian, French
• Jastrow, Morris (1905), "Arabic Language and Literature". *New International Encyclopedia*. 


[21] 19 External links

• Arabic: a Category III language Languages which are difficult for native English speakers.
• Dr. Nizar Habash’s, Columbia University, Introduction to Arabic Natural Language Processing
• Google Ta3reeb – Google Transliteration
20.3 Content license

- File:Wikibooks-logo-en-noslogan.svg  
  Source: https://upload.wikimedia.org/wikipedia/commons/d/df/Wikibooks-logo-en-noslogan.svg  
  License: CC BY-SA 3.0  
  Contributors: Own work  
  Original artist: User:Bastique, User:Ramac et al.

- File:Wikipedia-logo-v2.svg  
  License: CC BY-SA 3.0  
  Original artist: version 1 by Nohat (concept by Paullusmagnus); Wikimedia.

- File:Wikiversity-logo.svg  
  Source: https://upload.wikimedia.org/wikipedia/commons/9/91/Wikiversity-logo.svg  
  License: CC BY-SA 3.0  
  Contributors: Snorky (optimized and cleaned up by verdy_p)  
  Original artist: Snorky (optimized and cleaned up by verdy_p)

- File:Wikivoyage-Logo-v3-icon.svg  
  Source: https://upload.wikimedia.org/wikipedia/commons/d/dd/Wikivoyage-Logo-v3-icon.svg  
  License: CC BY-SA 3.0  
  Contributors: Own work  
  Original artist: AleXXw

- File:Wiktionary-logo-en.svg  
  License: Public domain  
  Original artist: Vectorized by Fvasconcellos (talk · contribs), 
  based on original logo tossed together by Brion Vibber

20.3 Content license

- Creative Commons Attribution-Share Alike 3.0