1

Introduction to Arabic

Arabic is a Semitic language akin to Hebrew, Aramaic, and Amharic, and more distantly related to indigenous language families of North Africa. It possesses a rich literary heritage dating back to the pre-Islamic era, and during the rise and expansion of the Islamic empire (seventh to twelfth centuries, AD), it became the official administrative language of the empire as well as a leading language of international scholarly and scientific communication. It is today the native language of over 200 million people in twenty different countries as well as the liturgical language for over a billion Muslims throughout the world.

1 Afro-Asiatic and the Semitic language family

The Semitic language family is a member of a broader group of languages, termed Afro-Asiatic (also referred to as Hamito-Semitic). This group includes four subfamilies in addition to Semitic, all of which are indigenous languages of North Africa: (1) Tamazight (Berber) in the Northwest (Morocco, Mauretania, Algeria, Tunisia and Libya); (2) the Chad languages (including Hausa) in the Northwest Central area; (3) ancient Egyptian and Coptic; and (4) the Cushitic languages of Northeast Africa (Somalia, the Horn of Africa).[1] The Semitic part of the family was originally based farthest East, in the Levant, the Fertile Crescent, and the Arabian peninsula.

Arabic, Hebrew, Aramaic (including Syriac), and Amharic are living language members of the Semitic group, but extinct languages such as Akkadian (Assyrian and Babylonian), Canaanite, and Phoenician are also Semitic. The Semitic language family has a long and distinguished literary history and several of its daughter languages have left written records of compelling interest and importance for the history of civilization.[2]

[1] See Zaborski 1992 for a brief description of the Afro-Asiatic language family and its general characteristics.

[2] For a general description of Arabic and the Semitic group, see Bateson 1967 (2003), 50–58 and Versteegh 1997, 9-22. For a more detailed discussion of the Semitic family and an extensive bibliography, see Hetzron 1987 and especially 1992, where he provides a list of fifty-one Semitic languages. For book-length introductions to comparative Semitic linguistic structure, see Wright 1966, Gray 1934, and especially Moscati 1969.

2 An overview of Arabic language history

The earliest stages of the Arabic language (Proto-Arabic or Old Arabic) are documented from about the seventh century BC until approximately the third century AD, but because of the paucity of written records, little is known about the nature of the language of those times. The only written evidence is in the form of epigraphic material (brief rock inscriptions and graffiti) found in northwest and central Arabia.[3]

The next period, the third through fifth centuries, is usually referred to as Early Arabic, a transitional period during which the language evolved into a closer semblance of Classical Arabic. There are again few literary artifacts from this age, but it is known that there was extensive commercial and cultural interaction with Christian and Jewish cultures during this time, an era of both Roman and Byzantine rule in the Levant and the Fertile Crescent.[4]

3 Classical Arabic

The start of the literary or Classical Arabic era is usually calculated from the sixth century, which saw a vigorous flourishing of the Arabic literary (or poetic) language, especially in public recitation and oral composition of poetry, a refined and highly developed formal oral art practiced by all Arab tribal groups and held in the highest esteem. During the sixth century, the Arabic ode, or *qaSîda*, evolved to its highest and most eloquent form. It was characterized by sophisticated metrics and a "highly conventionalized scheme . . . upwards of sixty couplets all following an identical rhyme."[5]

The form of language used in these odes is often referred to as the standard poetic language or the poetic koinè, and there are conflicting theories as to its nature – whether it was an elevated, distinctive, supra-tribal language shared by the leadership of the Arabic-speaking communities, or whether it was the actual vernacular of a region or tribe which was adopted by poets as a shared vehicle for artistic expression. In particular, debate has centered around the existence and use of desinential (i.e., word-final) case and mood inflection, a central feature of classical poetry but one which fell increasingly out of use in spoken Arabic, and which no longer exists in the urban vernaculars of today. Since little is

[3] A condensed but authoritative overview of the history and development of Arabic is provided in the article "Arabiyya" in the *Encyclopedia of Islam* (1960, I:561–603). See also Kaye 1987 and Fischer 1992. On the pre-Islamic period in particular, see Beeston 1981 and Versteegh 1997, 23–52. A good general reference in Arabic is Hijazi 1978.

[4] For a comprehensive, multi-volume study of the Arab world and its relations with Rome and Byzantium in late classical antiquity see Shahîd 1981, 1984, 1989, and 1995.

[5] Arberry 1957, 15. For further discussion of pre-Islamic Arabic poetry, see Nicholson 1987. See also Zwettler 1978 for a survey and analysis of the Arabic oral poetry tradition.

known about the nature of the everyday spoken Arabic of pre-Islamic times or the different levels of linguistic formality that might have been used on different occasions, certainty has not been reached on this point, although theories abound.[6]

In the seventh century AD the Prophet Muhammad was gifted over a period of years (622–632 AD) with the revelation of verses which constituted a holy book, the *Qurʾân*, in Arabic, which became the key text of the new monotheistic religion, Islam. The text was rendered into an official version during the reign of the Caliph ᶜUthmân (644–656 AD). From that time on, Arabic was not only a language of great poetic power and sophistication, but also permanently sacralized; as the chosen language for the *Qurʾân*, it became the object of centuries of religious study and exegesis, theological analysis, grammatical analysis and speculation.[7] Throughout the European medieval period, from the seventh through the twelfth centuries, the Arabic-speaking world and the Islamic empire expanded and flourished, centered first in Mecca and Madina, then Damascus, and then Baghdad.[8] Arabic became an international language of civilization, culture, scientific writing and research, diplomacy, and administration. From the Iberian peninsula in the West to Central and South Asia in the East stretched the world of Islam, and the influence of Arabic. The vast empire eventually weakened under the growing influence and power of emerging independent Muslim dynasties, with inroads made by the Crusades, Mongol invasions from the East, and with the expulsion of Muslims from the Iberian peninsula in the West. Arabic remained the dominant language in North Africa, the Levant, the Fertile Crescent, and the Arabian Peninsula, but lost ground to indigenous languages such as Persian in the East, and Spanish in the West.[9]

The language era from the thirteenth century to the eighteenth is generally known as "Middle Arabic," although there is some ambiguity to this term.[10] During this time, the Classical Arabic of early Islam remained the literary language, but the spoken Arabic of everyday life shifted into regional variations, each geographical

[6] On the nature of the standard poetic language and the pre-Islamic koinè, see Zwettler 1978, especially Chapter 3; Rabin 1955; Fück 1955; Corriente 1976; and Versteegh 1984, especially Chapter 1.

[7] For a brief introduction to the origins of Islam and the Qurʾânic revelations, see Nicholson 1930, especially Chapter 4.

[8] The main dynasties of the Caliphate are: the Orthodox Caliphs (632–661 AD); the Umayyads, based in Damascus (661–750 AD); and the Abbasids, based in Baghdad (750–1258 AD).

[9] Arabic has remained the dominant language in countries where the substratum language was originally Semitic or Afro-Asiatic, but not where the substratum languages were Indo-European, such as Persia or the Iberian peninsula. Aside from nationalistic and political considerations, linguistic compatibility between Arabic and its sister languages may have enabled certain populations to adapt more easily and throughly to Arabic. See Bateson 1967 (2003), 72–73 on this topic.

[10] Versteegh (1997, 114–29) has a cogent discussion of the issues related to "Middle Arabic." See also Blau 1961.

area evolving a characteristic vernacular.[11] The spoken variants of Arabic were not generally written down and therefore not preserved or anchored in any way to formalize them, to give them literary status or grammatical legitimacy. They continued to evolve along their own lively and supple paths, calibrating to the changes of everyday life over the centuries, but never reaching the status of separate languages.[12]

4 The modern period

The modern period of Arabic dates approximately from the end of the eighteenth century, with the spread of literacy, the concept of universal education, the inception of journalism, and exposure to Western writing practices and styles such as editorials, short stories, plays, and novels. Many linguists make a distinction between Classical Arabic (CA), the name of the literary language of the previous eras, and the modern form of literary Arabic, commonly known (in English) as Modern Standard Arabic (MSA). Differences between CA and MSA are primarily in style and vocabulary, since they represent the written traditions of very different historical and cultural eras, from the early medieval period to the modern. In terms of linguistic structure, CA and MSA are largely but not completely similar. Within MSA, syntax and style range from complex and erudite forms of discourse in learned usage to more streamlined expression in the journalistic, broadcasting, and advertising worlds. The high degree of similarity between CA and MSA gives strong continuity to the literary and Islamic liturgical tradition.

In Arabic, both CA and MSA are referred to as *al-lugha al-fuSHâ* اللغة الفصحى, or simply, *al-fuSHâ* الفصحى, which means "the most eloquent (language)." Badawi (1985) draws a helpful distinction between *fuSHâ al-ʿaSr* فصحى العصر (of the modern era) (MSA) and *fuSHâ al-turâth* فصحى التراث (of heritage) (CA). This is by no means a clear or universally accepted delineation, and opinion in the Arab world is apparently divided as to the scope and definition of the term *fuSHâ* فصحى.[13]

[11] There is speculation that the written/spoken Arabic dichotomy began much earlier, during the ninth century. See Blau 1961, Versteegh 1984, Fück 1955. For an evaluation of the main theories of Arabic dialect evolution and an extensive bibliography on the topic, see Miller 1986 and Bateson 1967 (2003), 94–114.

[12] This contrasts distinctively with the situation in the Scandinavian countries, for example, where a similar situation prevailed in that a mother language, known as Common Scandinavian, prevailed from about AD 550–1050, and then evolved into six official, literary languages (Danish, Dano-Norwegian, New-Norwegian, Swedish, Faroese, and Icelandic), plus many dialects. Despite the fact that the offshoots are all considered independent languages, "within this core [mainland Scandinavia] speakers normally expect to be understood [by each other] when speaking their native languages" (Haugen 1976, 23–24).

[13] See Parkinson's informative 1991 article for an extensive discussion of *fuSHâ*. In his study of Egyptian native Arabic speakers' ability with *fuSHâ*, he came to the conclusion that "The important point here is that people do not agree on a term, and that further they do not agree on what specific part of the communicative continuum, i.e., what specific varieties, any particular term should refer to" (33).

5 Arabic today

The Arab world today is characterized by a high degree of linguistic and cultural continuity. Arabic is the official language of all the members of the Arab League, from North Africa to the Arabian Gulf.[14] Although geography (including great distances and land barriers such as deserts and mountains) accounts for much of the diversity of regional vernaculars, a shared history, cultural background and (to a great extent) religion act to unify Arab society and give it a profound sense of cohesion and identity.

MSA is the language of written Arabic media, e.g., newspapers, books, journals, street signs, advertisements – all forms of the printed word. It is also the language of public speaking and news broadcasts on radio and television. This means that in the Arab world one needs to be able to comprehend both the written and the spoken forms of MSA. However, in order to speak informally with people about ordinary everyday topics, since there is no universally agreed-upon standard speech norm, Arabs are fluent in at least one vernacular form of Arabic (their mother tongue), and they understand a wide range of others. This coexistence of two language varieties, the everyday spoken vernacular and a higher literary form is referred to in linguistic terms as "diglossia."

5.1 Diglossia

The divergence among the several vernacular forms of Arabic, and between the vernaculars as a whole and the standard written form, make the linguistic situation of the Arab world a complex one.[15] Instead of having one universally agreed-upon standard speech norm, each major region of the Arab world (such as the Levant, the Arabian Gulf, the western Arabian peninsula, western North Africa, Egypt, and the Sudan) has as its own speech norm, a spoken vernacular coexistent with the written standard – MSA. Vernacular speech is much more flexible and mutable than the written language; it easily coins words, adapts and adopts foreign expressions, incorporates the latest cultural concepts and trends, and propagates slang, thus producing and reflecting a rich, creative, and constantly changing range of innovation. Vernacular or colloquial languages have evolved their own forms of linguistic artistry and tradition in terms of popular songs, folk songs, punning and jokes, folktales and spontaneous performance art.

[14] Morocco, Algeria, Tunisia, Libya, Egypt, Sudan, Mauretania, Jordan, Lebanon, Palestine, Syria, Iraq, Kuwait, Bahrein, Qatar, UAE, Oman, Saudi Arabia, and Yemen.

[15] For more on diglossia, see Ferguson 1959a and 1996, and Walters, 1996. See also *Southwest Journal of Linguistics* 1991, which is a special issue devoted to diglossia. Haeri 2003 is a book-length study of the relationships among Classical Arabic, MSA, and colloquial Arabic in Egypt.

Their changeability, however, also means that Arabic vernaculars may vary substantially from one another in proportion to their geographical distance. That is, neighboring vernacular dialects such as Jordanian and Syrian are easily mutually intelligible to native Arabic speakers; however, distant regional dialects, such as Moroccan and Kuwaiti, have evolved cumulative differences which result in the need for conscious effort on the part of the speakers to accommodate each other and adjust their everyday language to a more mainstream level. Educated native Arabic speakers have enough mutual awareness of dialect characteristics that they can identify and adjust rapidly and naturally to the communicative needs of any situation.[16] This spontaneous yet complex adjustment made by Arabic speakers depends on their knowledge of the vast reservoir of the mutually understood written language, which enables them to intercommunicate. Therefore, Arabic speakers share a wealth of resources in their common grasp of the literary language, MSA, and they can use this as a basis even for everyday communication.

In the re-calibration of Arabic speech to be less regionally colloquial and more formal, however, some researchers have identified another variation on spoken Arabic, an intermediate level that is termed "cultivated," "literate," "formal," or "educated" spoken Arabic.[17] Thus, the Arabic language situation is characterized not simply as a sharp separation between written forms and spoken forms, but as a spectrum or continuum of gradations from "high" (very literary or formal) to "low" (very colloquial), with several levels of variation in between.[18] As Elgibali states (1993, 76), "we do not . . . have intuition or scholarly consensus concerning the number, discreteness and/or stability of the middle level(s)."

These levels are characterized by (at least) two different sociolinguistic dimensions: first, the social function; that is, the situations in which speakers find themselves – whether those situations are, for example, religious, formal, academic, casual or intimate. Secondly, these levels are conditioned by the educational and regional backgrounds of the speakers. In this intricate interplay of speech norms, situations, and backgrounds, educated native Arabic speakers easily find their way, making spontaneous, subtle linguistic adjustments to suit the dimensions of the occasion and the interlocutors.

[16] For a detailed discussion of variation in Arabic see Elgibali 1993.

[17] This is known as "cultivated" speech in Arabic: *ʿâmmiyyat al-muthaqqafîn* عامّيّة المثقّفين, or *lughat al-muthaqqafîn* لغة المثقّفين. A number of Arabic linguists have researched and discussed this phenomenon, but there is no consensus as to the nature, extent, definition, and use of this part of the Arabic language continuum. The focus of the dispute centers around the ill-defined and unstable nature of this particular form of spoken Arabic and whether or not it can be distinguished as an identifiable linguistic level of Arabic. For more discussion of this point, see Badawi 1985, Elgibali 1993, El-Hassan 1978, Hary 1996, Mitchell 1986, Parkinson 1993, and Ryding 1990 and 1991.

[18] See, for example, the five levels distinguished in Badawi 1985 and the "multiglossia" of Hary 1996.

5.2 Modern Standard Arabic: MSA

MSA is the written norm for all Arab countries as well as the major medium of communication for public speaking and broadcasting.[19] It serves not only as the vehicle for current forms of literature, but also as a resource language for communication between literate Arabs from geographically distant parts of the Arab world. A sound knowledge of MSA is a mark of prestige, education, and social standing; the learning of MSA by children helps eliminate dialect differences and initiates Arab children into their literary heritage and historical tradition. It aids in articulating the connections between Arab countries and creating a shared present as well as a shared past. Education in the Arab countries universally reinforces the teaching and maintenance of MSA as the single, coherent standard written language.

A number of excellent Western pedagogical texts have been developed over the past fifty years in which MSA is discussed, described, and explained to learners of Arabic as a foreign language.[20] However, up to this point, there has been no comprehensive reference grammar designed for use by western students of MSA.

5.3 Arabic academies

Grammatical and lexical conservatism are hallmarks of MSA. Arabic language academies exist in several Arab capitals (Cairo, Damascus, Baghdad, Amman) to determine and regulate the procedures for incorporation of new terminology, and to conserve the overall integrity of MSA.[21] Although foreign words are often borrowed into Arabic, especially for ever-expanding technical items and fields, the academies try to control the amount of borrowing and to introduce and encourage Arabic-derived equivalents, such as the Arabic word *hâtif* هاتف (pl. *hawâtif* هواتف) for 'telephone' (based on the Arabic lexical root *h-t-f*), to counteract the widespread use of the Arabized European term: *tiliifûn* تليفون.

According to Versteegh (1997, 178) "From the start, the goal of the Academy was twofold: to guard the integrity of the Arabic language and preserve it from dialectal and foreign influence, on the one hand, and to adapt the Arabic language to the needs of modern times, on the other." Another researcher states

> Arab academies have played a large role in the standardization of modern written and formal Arabic, to an extent that today throughout the Arab world there is more or less one modern standard variety. This is the variety used in newspapers, newsreel

[19] For a discussion and definition of this particular term, see McLaughlin 1972.
[20] See, for example, Abboud and McCarus 1983; Abboud, Attieh, McCarus, and Rammuny 1997; Brustad, Al-Batal, and Al-Tonsi 1995 and 1996; Cowan 1964; Middle East Centre for Arab Studies (MECAS) 1959 and 1965; Rammuny 1994; Ziadeh and Winder 1957.
[21] For more detail on Arabic language academies see Holes 1995, 251–55 and Stetkevytch 1970, 23–25 and 31–33.

> broadcasting, educational books, official and legal notices, academic materials, and instructional texts of all kinds. The three academies that have had the greatest influence are those based in Cairo, Damascus and Baghdad. Among the common objectives of these academies is the development of a common MSA for all Arabic-speaking peoples. (Abdulaziz 1986, 17).

5.4 Definitions of MSA

A fully agreed-upon definition of MSA does not yet exist, but there is a general consensus that modern Arabic writing in all its forms constitutes the basis of the identity of the language. Modern writing, however, covers an extensive range of discourse styles and genres ranging from complex and conservative to innovative and experimental. Finding a standard that is delimited and describable within this great range is a difficult task; however, there is an identifiable segment of the modern Arabic written language used for media purposes, and it has been the focus of linguists' attention for a number of years because of its stability, its pervasiveness, and its ability to serve as a model of contemporary written usage. Dissemination of a written (and broadcast) prestige standard by the news media is a widespread phenomenon, especially in multilingual, diglossic, and multi-dialectal societies.

One of the most complete descriptions of MSA is found in Vincent Monteil's *L'arabe moderne* in which he refers to "le néo-arabe" as "l'arabe classique, ou régulier, ou écrit, ou littéral, ou littéraire, sous sa forme moderne" (1960, 25). That is, he understands "modern Arabic" to be the modern version of the old classical language. He also states that "on pourrait aussi le traiter d'arabe 'de presse', étant donné le rôle déterminant qu'a joué, et que joue encore, dans sa diffusion . . . *lughat al-jarâ'id*" (1960, 27). Defining MSA through its function as the language of the Arabic news media is a useful way to delimit it since it is not officially codified as a phenomenon separate from Classical Arabic and because Arabic speakers and Arabic linguists have differing opinions on what constitutes what is referred to as *al-lugha al-fuSHâ*. As Monteil also remarks, "s'il est exact de reconnaître . . . que l'arabe moderne 'se trouve être une langue assez artificielle, une langue plus ou moins fabriquée' plutôt qu'un 'usage codifié,' il faut déclarer . . . que 'c'est une langue vivante' et qui 'correspond à un besoin vital'" (1960, 28). It is these characteristics of newspaper language, its vitality and practicality, that make it a prime example of modern written Arabic usage.

Elsaid Badawi's phrase, *fuSHâ al-ʿaSr* فصحى العصر, is his Arabic term for MSA (1985, 17), which he locates on a continuum (at "level two") between Classical Arabic ("level one") and Educated Spoken Arabic ("level three"). As he points out, the levels "are not segregated entities," (1985, 17) but shade into each other gradually. He identifies level two (MSA) as "mostly written" rather than spoken, and levels

two and three as essentially "in complementary distribution" with each other (1985, 19), that is, they function in separate spheres, with some overlap.

Leslie McLoughlin, in his 1972 article "Towards a definition of Modern Standard Arabic," attempts to identify distinctive features of MSA from one piece of "quality journalism" (57) and provides the following definition which he borrows from M. F. Saᶜîd: "that variety of Arabic that is found in contemporary books, newspapers, and magazines, and that is used orally in formal speeches, public lectures, learned debates, religious ceremonials, and in news broadcasts over radio and television" (58). Whereas Saᶜîd states that MSA grammar is explicitly defined in grammar books (which would bring it close to CA), McLoughlin finds several instances in which MSA differs from CA, some of which are lexical and some of which are syntactic (72–73).

In her *Arabic Language Handbook* (1967; 2003, 84), Mary Catherine Bateson identified three kinds of change that differentiate MSA from CA: (1) a "series of 'acceptable' simplifications" in syntactic structures, (2) a "vast shift in the lexicon due to the need for technical terminology," and (3) a "number of stylistic changes due to translations from European languages and extensive bilingualism."

In the research done for this book, a wide variety of primarily expository texts, including Arabic newspaper and magazine articles, as well as other forms of MSA, were consulted and put into a database over a period of ten years. The morphological and syntactic features of the language used in these writings were then analyzed and categorized. This resulted in the finding that few structural inconsistencies exist between MSA and CA; the major differences are stylistic and lexical rather than grammatical. Particular features of MSA journalistic style include more flexible word order, coinage of neologisms, and loan translations from western languages, especially the use of the *ʾiDaafa* إضافة or annexation structure to provide equivalents for compound words or complex concepts. It is just this ability to reflect and embody change while maintaining the major grammatical conventions and standards that make journalistic Arabic in particular, a lively and widely understood form of the written language and, within the style spectrum of Arabic as a whole, a functional written standard for all Arab countries.

2

Phonology and script

This chapter covers the essentials of script and orthography as well as MSA phonological structure, rules of sound distribution and patterning, pronunciation conventions, syllable structure, and word stress. Four features of Arabic script are distinctive: first, it is written from right to left; second, letters within words are connected in cursive style rather than printed individually; third, short vowels are normally invisible; and finally, there is no distinction between uppercase and lowercase letters. These features can combine to make Arabic script seem impenetrable to a foreigner at first. However, there are also some features of Arabic script that facilitate learning it. First of all, it is reasonably phonetic; that is, there is a good fit between the way words are spelled and the way they are pronounced. And secondly, word structure and spelling are very systematic.

1 The alphabet

There are twenty-eight Arabic consonant sounds, twenty-six of which are consistently consonants, but two of which – *waaw* and *yaaʾ* – are semivowels that serve two functions, sometimes as consonants and other times as vowels, depending on context.[1] For the most part, the Arabic alphabet corresponds to the distinctive sounds (phonemes) of Arabic, and each sound or letter has a name.[2] Arabic letter shapes vary because Arabic is written in cursive style, that is, the letters within a word are systematically joined together, as in English handwriting. There is no option in Arabic for "printing" or writing each letter of a word in independent form. There is no capitalization in Arabic script and therefore no distinction between capital and small letters. Letters are instead distinguished by their position in a word, i.e., whether they are word-initial, medial, or final. This is true

[1] "Certain consonants have some of the phonetic properties of vowels . . . they are usually referred to as approximants (or frictionless continuants), though [/w/ and /y/] are commonly called semivowels, as they have exactly the same articulation as vowel glides. Although phonetically vowel-like, these sounds are usually classified along with consonants on functional grounds" Crystal 1997, 159. See also section 4.2.2. this chapter.

[2] For further reading about the Arabic alphabet and its close conformity with the phonemes of the language, see Gordon, 1970, 193–97.