# Chapter 1
# The Structure of Arabic Language and Orthography

**Elinor Saiegh-Haddad and Roni Henkin-Roitfarb**

**Abstract** This chapter was designed to promote our understanding of the triangulation, in Arabic, of language, orthography and reading. We focus on topics in the structure of the Arabic language and orthography that pertain to literacy research and practice. It is agreed that the development of basic reading skills is influenced by linguistic (mainly phonological and morpho-syntactic) and orthographic variation among languages. Therefore, the chapter devotes particular attention to these aspects of the linguistic structure of Arabic and to the way this structure is represented in the Arabic orthography. Further, in light of the importance of oral language processing skills in the acquisition of reading, the chapter also discusses Arabic diglossia: it describes the linguistic distance between Colloquial or Spoken Arabic and Standard or Literary Arabic, the primacy of Standard Arabic linguistic structures in the written form of the language, and the effect of this on several linguistic processes in literacy acquisition.

**Keywords** Arabic · Diacritics · Diglossia · Language · Morphology · Orthography · Phonology · Reading · Spelling · Syntax

## 1.1 Introduction

Arabic is the native language of approximately 300 million people worldwide and is an official language in 27 states. Also, as the language of the *Quran* it is the religious and liturgical language of all Muslims everywhere. Significantly, some local spoken variety of this language is spontaneously acquired by all native speakers as their mother tongue. This variety is known as *Spoken* (or *Colloquial*) *Arabic*, a collective term that refers to the whole range of Arabic vernaculars in numerous local dialects. These are generally classified into two regional clusters: *Eastern*

E. Saiegh-Haddad (✉)
Bar-Ilan University, Israel
e-mail: saieghe@mail.biu.ac.il

R. Henkin-Roitfarb
Ben-Gurion University of the Negev, Beersheba, Israel
e-mail: henkin@bgu.ac.il

E. Saiegh-Haddad, R. M. Joshi (eds.), *Handbook of Arabic Literacy,* Literacy Studies 9,
DOI 10.1007/978-94-017-8545-7_1, 

and *Western* dialects. Eastern Arabic is spoken throughout the Fertile Crescent, in the Arabic-speaking regions of Asia, in Egypt, in the Sudan, and in partially Arabized parts of East Africa. Western Arabic is spoken in the region referred to as the *Maghreb*, including Morocco, Algeria, Tunisia, Mauritania and Libya. The regional distinction between Eastern and Western Arabic coincides with contrasting linguistic differences of phonological, morphological, phonotactic, and lexical nature, pertaining most saliently to the inflection of the imperfect verb, syllable structure, and many items of lexicon.

In contrast with the dialects, the literary varieties of Arabic, namely *Classical Arabic*, *Literary Arabic* and their modern descendant, known as (*Modern*) *Standard Arabic* (MSA), have no native speakers.[1] These literary varieties constitute the primary language of literacy,[2] namely the language children are taught to read and write at school and the only variety considered, until recently, proper for writing Arabic. As such, it is the only variety with a standardized written form. Although Spoken Arabic may be phonetically represented using the Arabic alphabet (notwithstanding some spoken sounds that have no corresponding letters) there is no consensus regarding the appropriate orthographic representation of Spoken Arabic, or even as to whether it is legitimate (culturally and ideologically) to put this non-prestigious form of the language into writing.[3]

## 1.2 The Structure of Arabic

### *1.2.1 Phonology: Consonants, Vowels, Diphthongs*

The rich consonantal inventory of Modern Standard Arabic comprises 28 phonemes (two of which are actually semi-vocalic, see below). Four coronals, */s t d ð/*, represented by the letters س ت د ذ respectively, whose primary articulation involves the tongue blade and the dental-alveolar location, have phonemic counterparts characterized phonetically by a velarized co-articulation known in traditional Arabic grammar as إطْباق *ʔiṭba:q* 'covering, lidding'. Articulation of these sounds

[1] These terms have historically referred to different language varieties—*Classical Arabic* referred to the language of pre-Islamic poets; *Literary Arabic* referred to the prose language of medieval Islam, while (*Modern) Standard Arabic* refers to the modern use of this language, a descendant of the former two older forms (Bateson 2003, p. 75). The distinction, however, is not strictly adhered to.

[2] Writing in some of the colloquial prestige dialects has been noted since the fifteenth century, but most prominently since the nineteenth century in the Cairene dialect for several genres of literary prose, poetry, and drama. This 'culture of the colloquial' has been challenged and evoked some opposition and debate in Egypt (Davies 2006).

[3] Historically, Colloquial Arabic is argued by scholars to have descended from "some form of inter-tribal speech in use during the period of the [Islamic] conquests containing a greater or lesser admixture of ClA [Classical Arabic], and owe their variations to the indigenous influences" (Bateson 2003, p. 94). The popular belief that Colloquial Arabic is a direct deterioration of Classical Arabic, believed to have been the spoken language of the pre-Islamic era until spoiled by foreign substrata in the newly conquered territories, has been refuted in the light of evidence that Classical Arabic was never generally spoken (ibid).

involves raising the tongue body toward the back of the soft palate (Davis 2009, p. 636),[4] so that it "seems to fill the cavity above like a lid" (Bakalla 2007, p. 459). Additional co-articulations characterize these four phonemes, including constriction of the top of the pharynx (Al-Ani 2008, p. 599; Bakalla 2007, p. 460; Broselow 2008, p. 611; Holes 2004, p. 57). They are subsequently labeled 'pharyngealized', 'velarized', or 'emphatic', and are conventionally transcribed with a diacritic underdot */ṣ ṭ ḍ ðˤ/*. In the Arabic alphabet these phonemes are represented by the letters ص ط ض ظ respectively.[5]

These velarized emphatics share with other back consonants (velar غ */ɣ/* and خ */x/*; and uvular ق */q/*) the feature of تَفْخيم *tafxi:m* 'thickening, magnifying, emphasizing' (Bakalla 2009, p. 421) caused by the tongue raising (in the primary articulation of the latter but as a secondary co-articulation in the former). In modern dialects, all these مُسْتَعْلِيَة *mustaʕliya* 'raised' consonants (velarized and velar), also ر */r/* in many cases (Holes 2004, p. 58), tend to trigger a phonological assimilation process known as 'velarization spread' or 'emphasis spread'. This process results in the lowering and backing of neighboring vowels and in the velarization of surrounding consonants within the word, and sometimes even across a word boundary, until blocked by a high or front environment. Velarization spread may proceed forward, as in *ṣa:d* [*ṣa:ḍ*] 'to hunt', where the emphatic $C_1$ */ṣ/* partially assimilates the non-emphatic $C_2$ */d/* with respect to velarization, turning it into a *[ḍ]* allophone. Alternatively, velarization spread may proceed backward, as in *wasaṭ* [*wạṣạṭ*] 'middle', where the emphatic */ṭ/* velarizes the preceding non-emphatic */s/*, turning it into allophonic *[ṣ]*. The vowels in both cases become velarized as a result of this process. The two directions of spread have been claimed to stand in asymmetrical relation: regressive spread, like regressive assimilation in general, is more frequent and 'stronger'—it is more categorical (i.e. non-gradient) and less subject to blocking by consonants and high vowels (Davis 2009, p. 637).[6]

'Marginal' (Al-Ani 2008, p. 600) or 'secondary' emphatics, primarily */l m b/* in the vicinity of back vowels, may also trigger backing effects in many dialects. Notably, phonemic value has been claimed for secondary emphatics, such as */ṛ ṃ ḷ/* in Negev Arabic, e.g., *na:ṛ* 'fire', *ʔaṃṃ* 'mother', *xa:ḷ* 'maternal uncle', respectively. But minimal pairs cannot be established since the secondary emphatics are limited to a low vocalic environment (Davis 2009, p. 637) and are thus conditioned allophones (phonetic variants of phonemes) in contrast with the true or primary emphatic phonemes which are by definition non-conditioned. Moreover, for example, in the Negev Arabic pair *xaḷḷ-i:*(*h*)[7] 'my vinegar' vs. *xall-i:h* 'leave him' (Shawarbah 2012, p. 55), velarization in the former affects the entire lexeme [*χạḷḷ*], and a pair cannot be minimal if it differs

[4] According to other descriptions, the back of the tongue is raised towards the velum, i.e. the extreme back of the palate (Bakalla 2007, p. 459; Shawarbah 2012, p. 54).

[5] In many modern dialects, including Negev Arabic, *ḍ* and *ðˤ* have merged and are pronounced as an interdental emphatic, like the historical/*ðˤ*/.

[6] But Al-Ani (2008, p. 600) claims the opposite: "The progressive spreading is the most common, whereas regressive spreading is very rare".

[7] The 1st person sg. possessive and accusative suffixes in Negev Arabic, stressed *-i:* 'my' and *-ni:* 'me' respectively, may end in an *h*-like off-glide, so that *ʔibni:h* 'my son' is indistinguishable from the imperative *ʔibni:h* 'build it' (Blanc 1970, p. 131; Henkin 2010, p. 14).

in more than one segment. The same is true for the oft-cited 'minimal pair' *walla:h* 'he appointed him' vs. *waḷḷa:h* 'by God' (see for example, Al-Ani 2008, p. 600). Since the latter word is emphatic throughout [*wạḷḷạ:ḥ]*, the pair is far from minimal. Notably, the velarized consonant develarizes in a front environment, as in *l-illa:h* 'to God', which shows it to be a conditioned allophone. In any case, it is agreed among Arabists that the phonological scope of emphasis and rules of velarization spread are highly dialect-specific: "dialects may differ in the domain of emphasis spread, the direction of emphasis spread, the set of consonants that trigger emphasis spread, and the set of segments that block emphasis spread" (Broselow 2008, p. 610 citing Watson 2002, pp. 273–275). Moreover, the phonological scope of emphasis emanating from both 'primary emphatics', i.e. the four conventionally recognized emphatics of Classical Arabic, and 'secondary emphatics', such as /*l̩ m̩ b̩*/, is a suprasegmental phenomenon pertaining to both phonetics and phonotactics. Notably, it tends to influence the phonetic realization of consonants and vowels in MSA which, in the absence of an accepted MSA norm, will reflect the speaker's native dialect (Holes 2004, p. 58). Most importantly for our study, this spreading phenomenon results in a large set of velarized allophones. Some of these allophonic variants coincide with Arabic phonemes that have orthographic representation in the Arabic alphabet, including (ظ ط ض ص). This, as we will explain later, becomes an important issue in spelling Arabic and a source of orthographic opacity.

Two of the 28 conventional 'consonants', namely the glides /*w*/ and /*y*/, are in fact better considered semi-vowels (or semi-consonants): like consonants and unlike vowels, the glides may open a syllable (Holes 2004, p. 57); but in other respects, including the articulatory, acoustic and even orthographic (see Sect. 1.3: Orthography), they act like a prolongation of the corresponding vowels */u/* and */i/* respectively: the letter و represents both the semi-consonantal glide /*w*/ and the long vowel /*u:*/; correspondingly, the letter ي represents simultaneously the semi-consonantal glide /*y*/ *a*nd the long vowel */i:/*.

Notwithstanding the large consonantal inventory of Standard Arabic, its vocalic inventory is small, consisting of just 6 vowel phonemes. The three short vowels are low /*a*/, high front */i/*, and high back /*u*/, corresponding to their respective long equivalents: */a:/, /i:/,* and /*u:*/ (Broselow 2008, p. 609), as in *walad* 'boy', *bint* 'girl', *ʔumm* 'mother'; *na:s* 'people', *di:n* 'religion', *du:r* 'houses', respectively. In fact, some linguists (cf. Holes 2004, p. 57) recognize even fewer vocalic phonemes—just three (short) vowels, and an element of length applicable to both vowels and consonants: a geminated or lengthened consonant such as *ll* by this approach is prosodically equivalent to a long vowel, such as */a:/*. But it must be remembered that the distributional properties of lengthened vowels and geminated consonants are very different: a geminated *ll* may 'split' to two distinct, non-adjacent ones *l*V*l*. Thus, the root *DLL* gives both *dall* 'to guide' (with a geminated *ll*) and *dali:l* 'proof' (where the two root consonants $C_2$-*l* and $C_3$-*l* are separated by a vowel */i:)*. In contrast, a long vowel such as */a:/* cannot 'split' to two non-adjacent short ones, in a sequence such as *a*C*a*.

Ancient Arabic dialects, specifically eastern ones, appear to have had a fourth long vowel, the result of إمالة *ʔima:la* 'inclination, deflection', namely raising and fronting from an original /*a:*/ towards /*e:*/ or even /*i:*/ (Levin 2007; Versteegh 2001, p. 42; Wright 1975 I, p. 10). Medial (word internal) ʔ*ima:la* of several types has been

recognized in modern dialects. Minimal pairs in some sub-dialects of Negev Arabic include *jdæ:d* 'new' (plural) / *jda:d* 'forefathers', *bæ:liy* 'worn out' (participle) / *ba:li* 'my mind' (Henkin 2010, p. 53). Two secondary phonemes in many dialects are /*e:*/ and /*o:*/, resulting from diphthong contraction (see below): *mawt⇒mo:t* 'death'; *sayf⇒se:f* 'sword'.

The term 'diphthong', known in Arabic as صَوْت مُرَكَّب *ṣawt murakkab* 'compound sound', is applied in Semitic linguistics to a combination of a vowel and a glide, rather than to a sequence of two adjacent vowels forming the peak of a syllable, as in other languages. In traditional Arabic grammar just two falling diphthongs are recognized: *aw* and *ay* (al-Ani 2008, p. 599; Iványi 2006, p. 640). Widespread contraction or monophthongization of these in the dialects, especially in front phonetic environments, has given rise to two additional long vowels of Spoken Arabic, *e:* and *o:*. Both are at least partially phonemic, as witnessed by minimal pairs such as *de:r* 'monastery' vs. *di:r* 'put' (imperative); *do:r* 'turn, role' vs. *du:r* 'houses'. However, not all native speakers perceive the difference between /*e:*/ and /*i:*/, or between /*o:*/ and /*u:*/, even in dialects where some phonemic status has been established (cp. Blanc 1970, p. 118 for Negev Arabic).

### *1.2.2 Phonotactics: Root Structure, Syllable Structure, Stress*

All 28 Arabic consonants may function as root radicals. However, there are some constraints on the distribution of some consonants, mainly on the co-occurrence of root consonants that are identical, homorganic or otherwise similar. For example, $C_2$ and $C_3$ may be identical, as in *RDD*, whence *radd* 'to return'; but $C_1$ and $C_2$ cannot be identical. A comprehensive table, devised by Greenberg (Frisch 2008, p. 625), presents the co-occurrence of all consonant groups with each other on a gradient of similarity and co-occurrence, and a principle of similarity and preference in inverse correlation. Moreover, Frisch (2008, p. 628) proposes a functional base for the principle of dissimilation, namely that similarity poses a cognitive load and is therefore undesirable: "forms without repetition are easier to produce, perceive, and hold in short-term memory". Some basic principles are as follows (Broselow 2008, p. 610):

> Generally, roots are unlikely to contain adjacent labial consonants (/*b f m*/). Adjacent coronals are avoided if they also share similar manners of articulation; thus, roots with adjacent coronal sonorants, coronal stops, or coronal fricatives are rare, and even combinations of a coronal stop and a coronal fricative are unlikely. In the posterior regions, combinations of velar and uvular consonants are avoided, as are combinations of guttural consonants.[8]

All syllables in Modern Standard Arabic begin with a single consonant (C) or glide, serving as the syllable onset and necessarily followed by a vowel (V), as the syllable nucleus or peak. The minimal syllable is thus CV, as in the preposition *li* 'to'. This is known as an open syllable, because it ends in a vowel, which is characterized by relative openness of the vocal tract. It is monomoraic, i.e. it contains one

[8] Holes (2004, p. 99) precludes homorganic non-identical root radicals in general. Exceptions include the sonorants, which can co-occur with any other consonant in any position.

mora,[9] and is thus light. Each additional mora, be it vowel length or an additional consonant, adds heaviness. A bimoraic syllable, consisting of CV: or CVC, is thus 'heavy' (Broselow 2008, p. 612; Jesry 2009, p. 388; Kager 2009, p. 344).[10] It may be open (CV: as *ma:* 'what') or closed (CVC, as *man* 'who'). Syllables with 3–4 moras, considered 'extra heavy', or 'super heavy' in this system, are limited to pausal status. One sub-class of this category is a syllable containing both a long vowel and a closing consonant (CV:C), e.g., *ba:b* 'door'—this structure may occur word-internally in special cases, such as *ʕa:m.ma* 'public' (fm.) (Holes 2004, p. 61); another is a syllable that is 'doubly' closed with two consonants: CVCC, e.g., *kalb* 'dog' or even CV:CC, e.g., *ma:rr* 'passer by'—this last type, however, is limited to geminate consonants (Broselow 2008, p. 610 ff.; Jesry 2009, p. 388).

Importantly, Arabic syllable boundaries vary with morphological processes such as declension that the words might undergo. Since syllabification in junctural (connected) prose operates across the boundaries of words in sequence, we find Standard Arabic pausal (basic) forms resyllabified in non-pausal connected or context status, e.g., pausal *jadd* 'grandfather' vs. context *jaddun* (*jad.dun*); pausal *maktab* (*mak.tab)* 'office' vs. *maktabu š-šurṭa* (*mak.ta.buš.šur.ṭa)* 'the police office' in a construct phrase. The Standard Arabic sequence *min* 'from' and *l-bayt* 'the house' potentially forms a 3-consonant cluster (*nlb*). Since Arabic does not permit 3-consonant clusters in principle, an anaptyctic (helping vowel) is inserted to break the cluster, forming *min-al-bayt* (*mi.nal.bayt)* 'from the house'.

It is noteworthy that Arabic vernaculars may vary in their syllable structure and their phonotactic constraints. For instance, Palestinian Arabic allows many 2-consonant clusters in syllable-initial positions (e.g., *tra:b* 'soil' or *kla:b* 'dogs') or across morpheme-boundaries in some grammatical forms (e.g., definite nouns *l-be:t* 'the house'). Yet, syllable final clusters are not as prevalent. The sonority principle of final anaptyxis is $C_1VC_2C_3 \Rightarrow C_1VC_2VC_3$ if Sonority $C_2 <$ Sonority $C_3$ (Zemánek 2006a, p. 86). In other words, a rise in sonority within a final $C_2C_3$ cluster will call for anaptyxis, so *qabl* 'before' (sonority rises from $C_2$*b* to $C_3$*l*) $\Rightarrow$*qabil*. Notably, the sonority hierarchy for final clusters is directly contrary to the sonority hierarchy for initial clusters, where anaptyxis is called for in the case of falling sonority. Thus, perfectly acceptable word-initial clusters of a $C_1$ stop or fricative and a $C_2$ sonorant of higher sonority, such as *dr, bl, tn, fl, sm* in ***dr****u:s* 'lessons', ***bl****a:d* 'country', ***tn****e:n* 'two', ***fl****a:n* 'so-and-so', ***sm****i:n* 'fat', will need anaptyxis in word final position, as in *ba.****dir*** 'full moon', *qa.****bil*** 'before', *ma.****tin*** 'corpus', *ṭi.****fil*** 'child', *ʔi.****sim*** 'name', respectively. Word-final clustering is more generally acceptable in the case of dropping sonority: *ʔaka****lt*** 'I/you ate', *ka****lb*** 'dog', *ħa****md*** 'praise', though again, dialects vary with respect to clustering in such cases.

Arabic stress is non-phonemic (Holes 2004, p. 62) or non-distinctive (Kager 2009, p. 344), and is predictable (though dialect-dependent), given the weight

[9] A mora is a prosodic weight unit for classifying syllable structure. It counts all units excluding the onset consonant.

[10] Holes (2004, p. 62 ff.) considers bimoraic syllables 'light' too; 'heavy' syllables in this system contain 3–4 moras. Al-Ani (2008, p. 601) similarly considers CVC a light syllable. A little further on in the article, however, Al-Ani (2008, p. 602) posits an in-between category of 'medium' or bimoraic syllables, such as *kam* 'how many' and *ma:* 'what'.

and number of syllables in the word.[11] In Standard Arabic, a word (in pausal status only) can contain just one extra-heavy syllable (of four elements or more)—that syllable is necessarily final, and receives stress, e.g., *ki.**ta:b*** 'book', *ka.**tabt*** 'I/you wrote'. In the absence of extra-heavy syllables, stress falls on the rightmost non-final heavy syllable (Kager 2009, p. 349): *mu.dar.ri.**su:.**na* 'teachers'; *yas.ta.**ʈi:.**ʕu* 'he is able'; *kas.**sar**.tu.hu* 'I broke it', ***mak**.tab* or ***mak.**ta.bun* 'office'. Otherwise, stress falls on the first syllable, e.g., ***ba.**ra.ka* 'blessing', ***ka**.ta.bu:* 'they wrote'.[12] Stress variation in Modern Standard Arabic is due, at least in part, to the fact that, as in the issue of syllable structure, here too speakers are influenced by their native dialects, which vary considerably in their stress rules. The Standard Arabic stress scheme just outlined is very similar to that of Eastern Arabic dialects (Kager 2009, p. 350).

### *1.2.3 Morphology: Root, Pattern*[13]

Arabic, like other Semitic languages, is characterized by a predominantly non-linear or non-concatenative morphological structure (Larcher 2006; McCarthy 1981), the hallmark of which is a جَذْر *jaðr* 'root' and a derivational or inflectional pattern ميزان صَرْفيّ *mi:za:n ṣarfiyy*.

In Semitic languages, morphological derivation and inflection typically involve two bound morphemes: a triliteral (and sometimes quadriliteral) root (e.g., $_{C1}$*K*-$_{C2}$*T*-$_{C3}$*B*) and a word pattern or template (Broselow 2008, p. 610; Holes 2004, p. 99), such as $C_1a{:}C_2iC_3$ e.g., *ka:tib* 'writer' (active participle) or $maC_1C_2u{:}C_3$, e.g., *maktu:b* 'written' (passive participle). The root is an unpronounceable bound morpheme, "a skeleton of consonants" (Bentin and Frost 1995, p. 273) that provides the core meaning, or the semantic family. The pattern is a non-pronounceable bound morpheme too—a fixed prosodic template with slots for the root consonants. The insertion of the root consonants within the word pattern produces a unique lexical item with a unique meaning and a well-defined grammatical category directly discernible by the specific word pattern. It is noteworthy that while patterns are

[11] Holes ibid presents rare cases where phonemic status may be attributed to stress. This is due to neutralization of word final gemination, which results in minimal pairs such as dialectal ***sA**kat 'he was silent' vs. sa**kAt**+t⇒sa**kAt**.* 'I was/you were silent'. But he notes that such cases are "marginal and artificial".

[12] More elaborate stress rules (Holes 2004, p. 62 ff.) account for cases like *yas.**ta**.mi.ʕu* 'he listens', *muš.ki.**la**.tu.ka* 'your problem' and, particularly, when all the non-final syllables are light, e.g., *ma.li.ka.tu.hu* 'his queen'. In this case there is no general agreement as to whether the stress fell on the first syllable in Classical Arabic ***ma**.li.ka.tu.hu* (Kager 2009, p. 349), or was limited to the last three syllables (Broselow 2008, p. 613), namely *ma.li.**ka**.tu.hu*, the Arab grammarians having totally ignored the issue of stress in their writings.

[13] In the following two sections we discuss mainly Modern Standard Arabic. In demonstrating the forms, however, we choose variants that are as close as possible to those of Spoken Arabic. We thus prefer pausal forms that omit final short vowels in the same way as dialectal variants, e.g., *katab* (and not *kataba)* 'to write', Impf. *yaktub* (rather than *yaktubu),* unless the omitted vowels are the issue discussed, or when historical morpho-phonological processes are being shown, e.g., *ramaya⇒rama*: 'to throw'.

primarily vocalic templates (vowel patterns), some patterns involve gemination of root consonants or vowel length, and others are augmented with certain consonants, such as */ʔ s t n/*. In the case of verbs, these augmented patterns are called أَفْعالمَزيدَة *ʔafʕa:l mazi:da* 'augmented verbs', namely all Arabic verb patterns except for pattern I, referred to as فِعْل مُجَرَّد *fiʕl mujarrad* 'bare verb', because it consists only of the root consonants and vocalic pattern. Importantly, the additional consonants of the augmented verbs, as well as the long vowels of word patterns, are an indispensable part of the orthographic representation of words, even in unvoweled Arabic script (see Sect. 1.3: Orthography).

The root-pattern morphological structure is common to almost all Arabic content words and some function words, such as *qabl* 'before'; their semantic identity is largely determined by the consonantal root. Interestingly, even loan words, such as *talfizyo:n* 'television' and *talifo:n* 'telephone', are treated by speakers as having an internal root-pattern structure; via a derivational process known as 'root extraction', new quadriliteral roots *TLFZ* and *TLFN* are derived and combine with the quadriliteral pattern $C_1aC_2C_3aC_4$ to form the verbs *talfaz* 'to televise' and *talfan* 'to phone'. Root consonants usually preserve their phonemic identity when combining with word patterns to form Arabic lexemes. Yet, because of velarization spread (the phonological assimilation process described earlier) some root consonants may become emphatic. This phonetic change is not represented, however, in the orthographic structure of Arabic words and this may lead to orthographic opacity (see Sect. 1.3: Orthography).

All consonants, including glides, can function as root-radicals. A root containing a glide, however, is considered مُعْتَلّ *muʕtall* 'weak',[14] being prone to morpho-phonological changes. These contrast with the 'strong' or 'sound' roots called صَحيح *ṣaħi:ħ* 'correct' whose radicals remain phonologically stable (Akesson 2009, p. 121; Holes 2004, p. 110 ff.; Versteegh 2001, p. 85 ff.; Versteegh 2007b, p. 309). In a $C_1$-glide root, known as مِثال *miθa:l* 'assimilated', e.g., *WJD* 'find', the glide may be elided in the Impf. **yawjidu* ⇒*yajidu* 'he finds'; a $C_2$-glide root, known as أَجْوَف *ʔajwaf* 'hollow', e.g., *QWL*, undergoes several changes, e.g., **qawaltu* ⇒*qultu* 'I said', Impf. **ʔaqwulu* ⇒*ʔaquwlu* ⇒*ʔaqu:lu* 'I say'; **qawalat* ⇒*qa:lat* 'she said'; a $C_3$-glide root, known as ناقِص *na:qiṣ* 'defective', such as *RMY,* is also prone to morpho-phonological changes, e.g., **ramaya* ⇒*rama:* 'to throw', Impf. **yarmiyu* ⇒*yarmiy* ⇒*yarmi:* 'he throws' (Akesson 2009, pp. 121–122; Chekayri 2007, p. 164 ff.).[15]

Most traditional Arabic dictionaries are alphabetically ordered by consonantal roots and they specify in each entry the specific meaning that results from the

[14] Some scholars include hamzated verbs, i.e. verbs containing *hamza* (see Sect. 1.3: Orthography), in the category of weak verbs (e.g., Voigt 2009, p. 700 ff.).

[15] The grammarians set up phonotactic rules according to a scale of relative lightness and strength of the phonemes that corresponds to sonority (Holes 2004, p. 113): vowels are lightest and strongest, consonants heaviest and weakest; within the vowels, the hierarchy is *a*>*i*>*u*. In contact, the lighter-stronger phoneme overrules and only sequences of rising lightness are permitted. So the triphthong *iyu* in **yarmiyu* above will contract to *iy* ⇒*i:*, as also in **qa:ḍiyu* ⇒*qa:ḍi:* 'judge' (Versteegh 2001, p. 86 ff.; Voigt 2009, p. 699). The homogeneous triphthongs **awa,***aya* are simplified by elision of the glide, as we saw in **qawala* ⇒*qa:la* and **ramaya* ⇒*rama*: above.

combination of the root with the pattern. Regular renditions of a word meaning from its root and pattern, known in Arabic grammatical terminology as قِياسيّ *qiya:siyy* 'analogous' or 'regular', need not be listed as these may be computationally constructed. In contrast, dictionaries attempt to list all meanings, known in traditional terminology as سَماعيّ *sama:ʕiyy* 'heard', i.e. based on hearing (Versteegh 2001, p. 85) or learned by ear. In the latter case, a word's meaning might not be a straightforward combinatorial function of the root meaning and the function of the word-pattern. This is because roots may be affiliated with more than one semantic family; some of these families may be remarkably distinct. Also, roots may undergo semantic broadening and adopt new areas of meaning while other areas might become obsolete. Finally, word patterns are not perfectly regular nor are they systematic.

It is possible to categorize word patterns in Arabic into two classes: *verbal patterns* and *nominal patterns*. Verbal patterns combine with roots to derive verbs, whereas nominal patterns combine with roots to derive nouns. There are 15 distinct triliteral verbal patterns (measures or forms, Hebrew *binyanim*) in Arabic, 10 of which are still productive (Holes 2004, p. 100 ff.; Larcher 2009, p. 640 ff.), though not necessarily in all dialects: I *faʕal*, II *faʕʕal*, III *fa:ʕal*, IV *ʔafʕal*, V *tafaʕʕal*, VI *tafa:ʕal*, VII *ʔinfaʕal*, VIII *ʔiftaʕal*, IX *ʔifʕall*, X *ʔistafʕal;* the remainder are rare and non-productive.[16] Quadriliterals have two distinct patterns *faʕlal and tafaʕlal*, $C_1aC_2C_3aC_4$ and $taC_1aC_2C_3aC_4$, respectively. Each verbal pattern in Arabic is associated with a set of morpho-syntactic inflectional patterns used in the conjugation of the verb for tense, person, number, gender, and mood.

Nominal patterns form a very large set. For example, Wright's grammar of Classical Arabic lists 44 nominal patterns derived from the first verbal pattern only. Holes (2004, p. 106) notes eleven among them as the most common in modern use. He also lists 13 additional patterns used in deriving nouns from augmented verbs. Boudelaa and Marslen-Wilson (2010, p. 483) report the occurrence of 2,324 different word patterns in current use in MSA; 'broken plural' patterns alone (see 1.2.4: Morpho-syntax) exceed 36 (Versteegh 2001, p. 84).

If patterns were perfectly systematic and predictable, "the lexicographer would only need to list the roots, and the speaker could combine them at will with the desired pattern to express, e.g., 'the place where such and such takes place', 'a professional practitioner of such and such', 'one who pretends to be such and such', etc." (Bateson 2003, pp. 1–2), but in fact, there is no such uniformity. Even though patterns are conceived to have clearly defined functions, they are not perfectly systematic. So, from the verb *jalas* Impf. *yajlis* 'to sit' we find *majlis* 'place or time of a meeting' in the $maC_1C_2iC_3$ pattern for place and time of an action distinct from *majlas,* a verbal substantive of the type known as مَصْدَر ميميّ *maṣdar mi:miyy* '*M*-verbal noun'. But in other cases, the verbal noun is identical to the noun designating place or time, or to the passive participle in the case of the derived verbal patterns (Wright 1975 I, pp. 124–129), e.g., *mujtamaʕ* 'gathering place' and also

[16] The numbering of these forms is a western innovation. Arabic terminology knows them just by name (Versteegh 2001, p. 87).

'gathered people, society'. This contributes to morphological opacity—difficulty in recovering the meaning of a word from its root-pattern morphological structure.

Another factor contributing to morphological opacity in Arabic is the fact that "many patterns are the result of a series of derivational steps, some of which are semantically systematic, while others seem arbitrary" (Bateson 2003, p. 2). So *qawmiyya* 'nationalism' is derived in stages from *qawm* 'race, people, nation'+attributive suffix *-iyy*=>*qawmiyy* 'national'+feminine suffix *-a* for an abstract noun (ibid, p. 20).[17]

### *1.2.4 Morpho-syntax: Parts of Speech, Inflection, Declension, Clitics*

Arabic words have been traditionally classified into three classes إِسْم *ʔism* 'noun' (including substantive and adjective), فِعْل *fiʕl* 'verb', and حَرْف *ħarf* 'particle' (including adverbs as well as prepositions and conjunctions). Both nouns and verbs inflect for gender (مُذَكَّر *muðakkar* 'masculine', مُؤَنَّث *muʔannaθ* 'feminine') and for number (مُفْرَد *mufrad* 'singular', مُثَنّى *muθanna:* 'dual', and جَمْع *jamʕ* 'plural'), although the morphemes marking these categories differ.

There are two pluralization mechanisms for nominal forms: سالِم *sa:lim* 'sound' or 'sane' concatenated plural on the one hand and so-called مُكَسَّر *mukassar* 'broken' or تَكْسير *taksi:r* 'breaking' non-concatenated plural on the other hand (Wright 1975 I, p. 191 ff.). The sound plural masculine suffixes, in general use for participles in the augmented verbal patterns (II -X), some animate nouns and adjectives (Versteegh 2001, p. 83) are *u:n(a)* **or** *i:n(a)* depending on case (see below), and the feminine suffix, also common in loans, is *-a:t;* so, for example, *muʕallim-u:na* 'teachers', in the oblique cases (accusative and genitive) *muʕallim-i:na*; fm. *muʕallim-a:t*. The broken plural patterns are numerous and diverse, e.g., *ʔaqla:m* 'pens' from *qalam*; *kila:b* 'dogs' from *kalb*; *kutub* 'books' from *kita:b*; *mulu:k* 'kings' from *malik; maka:tib* 'offices' from *maktab*. Dual nouns are suffixed with *-a:ni* or *-ayni* depending on case. In the head noun of a construct phrase and before possessive suffixes, the final syllable of the sound plural (and also the dual forms) is omitted, thus *muʔallim-u:-hum* 'their teachers'; *walada: l-ja:r* 'the neighbor's two sons'.

Verbs inflect for person (as well as number and gender)—مُتَكَلِّم *mutakallim* 'speaker', مُخاطَب *muxa:ṭab* 'addressee', and غائِب *ɣa:ʔib* 'absentee' (Wright 1975 I, p. 52). They may be structurally classified into two conjugations: the suffix conjugation combines perfective aspect with past tense, e.g., *katab-tu* 'I wrote, I have written' (the completed action is set in the past); the prefix conjugation combines

[17] The attributive suffix named نِسْبَة *nisba* 'relationship, attribution', is transcribed in the linguistic literature and dictionaries as *-i:, -iy, or -iyy*. We prefer the latter, reflecting most faithfully the morpho-phonological gemination occurring in MSA and seen in vocalized Arabic orthography. Gemination of this morpheme is absent from many dialects and this affects stress patterns in the spoken varieties.

imperfective aspect with non-past (present and future); secondary differentiations are encoded in particles, modal endings, and auxiliary verbs, e.g., *sa-ʔ-aktub* 'I will write' (the incomplete action of writing is explicitly set in the future by the particle *sa-*); *ṣirtu ʔ-aktub* 'I have begun to write, I began writing' (the incomplete action of writing is non-past, ongoing; its initiation is denoted by the auxiliary verb *ṣa:r* 'to become, begin', itself set in the perfective past).

Common to both nouns and verbs in Standard Arabic are عَلامات الإعْراب *ʕala:ma:t al-ʔiʕra:b* 'ʔiʕra:b-endings'. These vocalic word endings denote the syntactic categories of case and mood respectively. Nouns in non-pausal position take one of three case-endings: the nominative *-u*(*n*) which, being a high vowel, is called مَرْفوع *marfu:ʕ* 'raised', the accusative-adverbial *-a*(*n*) called مَنْصوب *manṣu:b* 'erected', and the genitive *-i(n)* which is مَجْرور *majru:r* 'pulled along' by a preceding preposition or construct-head of the إضافَة *ʔiḍa:fa* 'construct'. The imperfective verb resembles the noun in taking the former two endings—to denote the indicative and subjunctive moods respectively—and is thus called مُضارع *muḍa:riʕ* 'similar (to the active participle)'; the third mood, the jussive, is denoted by a zero-ending, whence the term مَجْزوم *majzu:m* 'apocopated' (Wright 1975 I, p. 60). The imperative أمْر *ʔamr* 'command' is considered a distinct mood in Arabic grammatical tradition; the classical 'energetic' form, known as تَأْكيد *taʔki:d* 'corroboration', is likewise listed as a mood in some modern reference works (e.g., Wright 1975 I, p. 51) or at least a modal category (Larcher 2009, p. 640).

The noun is determined by the article *(a)l-* 'the', by possessive suffixes, e.g., *ʔumm-i:* 'my mother', or by a following noun in construct (genitive) status, e.g., *ʔumm-u l-walad-i* 'the boy's mother'. Indetermination in Standard Arabic is marked by تَنْوين *tanwi:n* 'nunation', e.g., *ja:r-un* 'a neighbor'. Nouns are primarily triptotes, declining for all three cases; but there is a group of diptotes admitting just partial declension and hence known as غَيْر مُنْصَرِف *ɣayr munṣarif* 'non-declined (for *tanwi:n*)' or غَيْر قابِل للتَصْريف *ɣayr qa:bil l-it-taṣri:f* 'not allowing declination'. In the indefinite state they admit just *-u* or *-a* (not *tanwi:n)*, but behave regularly in definite status. This partial lack of declension is attributed by the grammarians to a deviation from default unmarked Arabic substantive basic forms (msc., sg., indefinite) in at least two of nine criteria of deviations, such as تَأْنيث *taʔni:θ* 'being feminine', وَصْفِيّة *waṩfiyya* 'being an adjective', عُجْمة *ʕujma* 'being a foreign word', تَرْكيب *tarki:b* 'being a compound', عَلَمِيّة *ʕalamiyya* 'being a proper noun', وَزْن الفِعْل *wazn al-fiʕl* 'a verbal pattern' (Versteegh 2001, p. 82; Wright 1975 I, p. 234 ff., especially p. 245). For example, the personal name *Yazi:d* 'loses' its capacity for triptosis by the two criteria of 'verbality' + 'proper noun'; *ʔakbar* 'bigger'– adjective + verbal form; *ħamra:ʔ* 'red'– adjective + feminine.

The adjective, named صِفَة *ṣifa* 'attribute', is a sub-class of the noun, characterized by admitting elative (comparative, superlative) forms, e.g., *kabi:r* 'big' vs. *ʔakbar* 'bigger, biggest'. Every adjective may be employed as a substantive and stand alone, e.g., *kari:m* 'a noble or generous man' (Bateson 2003, p. 44; Beeston 1970, p. 34, 67; Fischer 2006a, p. 18).

Arabic does not have a separate lexical category of adverbs. Adverbial functions are fulfilled by noun phrases and prepositional phrases, such as *ʔams* 'yesterday';

*bi-l-ʔamsi* ‘on the eve’ (Beeston 1970, p. 89), and most pervasively the accusative-adverbial case ending *-a(n),* as in *jidd-an* ‘very’, *layl-an* ‘at night’, *al-yawm-a* ‘today’.[18]

The morphological structure of Arabic also comprises a predominant system of clitics. These are morphemes that are grammatically independent, but phonologically dependent on another word or phrase. They are pronounced (and in Arabic also written) like affixes but function at the phrase level much like the English contracted forms -’ll in ‘he’ll’, or -’ve in ‘I’ve’. In Arabic, clitics may attach to the word as unstressed prefixes (proclitic) or suffixes (enclitic) and can co-occur within the same word, resulting in one-word phrases and clauses, as in بِبَيْتِهِ *bi-bayt-i-hi* ‘in his house’ or وَسَيَأْخُذُهُ *wa-sa-yaʔxuð-hu* ‘and he will take him’. Pronominal clitics are suffixed to verbs (as direct objects), to nouns (as possessives), and to prepositions; clitics that are prefixed to the content lexeme include several prepositions, conjunctions, and other particles, such as the article *(a)l-*, the asservative (emphasizing) *la-*, and future marker *sa-*.

### *1.2.5 Syntax*

Typologically, inflected languages do not need strict word order because syntactic functions are encoded morphologically (e.g., in case endings) and are thus independent of word order. Yet, “although Arabic is an inflected language, it does have a relatively rigid word order which allows for stylistic deviations” (Bateson 2003, p. 45). Moreover, word order is highly significant in the syntactic conception of the Arab grammarians. They traditionally classified Arabic clauses/sentences into two types (Fischer 2006b, p. 398; Versteegh 2001, pp. 79–81): one is the verbal clause (جُمْلة فِعْلِيَّة *jumla fiʕliyya*) which opens with a verb and proceeds in a default sequence of Verb-Subject-Object-Adverbial(s), e.g., *kataba r-rija:lu l-maktu:ba l-yawma,* literally ‘wrote the men the letter today’; the other, classified in the Arabic grammatical tradition as a nominal clause (جُمْلة إسْمِيَّة *jumla ʔismiyya)*, may naturally have no verb at all and constitute a Subject-Complement-Adverbial(s) sequence, e.g., *ʔar-rija:lu huna: l-yawma* ‘the men (are) here today’; more interestingly, however, a nominal sentence may also begin with a noun followed by a verb in a Subject-Verb-Object-Adverbial(s) sequence, e.g., *ʔar-rija:lu katabu: l-maktu:ba l-yawma* ‘the men wrote the letter today’. The apparent paradox, in western eyes, of a nominal sentence containing a verb, is very rational for the Arab grammarians. The

[18] In the Greek and Latin grammatical tradition the term ‘declension’ is exclusive to nouns. As mentioned earlier, however, Arab grammarians see the imperfect verb as مُضارع *muḍa:riʕ* ‘similar’ to the participle, and have focused their attention on the parallelism between verbal and nominal endings. They subsume both under the term إعْراب *ʔiʕra:b*, treated under syntax (نَحْو *naħw)*, rather than morphology (تَصْرِيف-صَرْف *ṣarf ~taṣri:f)*, which deals with inflections of person, number, etc. (Versteegh 2001, p. 74). In this tradition “the endings/*-u, -a*, -0/ of the imperfect verb are case endings” (Versteegh 2001, p. 85). We shall accordingly use the term ‘declension’ for verbal modal endings too, as is common in the writings of modern Arabists (e.g., Larcher 2009, p. 639; Versteegh 2001, pp. 76–79).

verb in a verbal clause profiles the action which initiates it. As such, it is not fully governed by the following subject and therefore is not in full agreement with it: in *kataba* (msc.sg.) *r-rija:lu* (msc.pl.) there is agreement in gender but not in number; the action is declared, as it were, semi-independently of the following subject, which is downgraded to almost an afterthought. In the nominal clause, however, the clause-initial subject is actually a topic in a left-dislocation syntagm. So *ʔar-rija:lu katabu:* is actually 'the men—they wrote', where 'the men' is a dislocated topic and the rest, a verbal sentence, is the comment. Verbal agreement is full in this structure (*katabu:* pl.), but is perceived to be to the covert subject pronoun 'they' rather than to the dislocated topic 'the men'. The syntactic behavior reflects a major semantic opposition, as formulated by Wright (1975 II, p. 251–252):

> The difference between verbal and nominal sentences, to which the native grammarians attach no small importance, is properly this, that the former relates an act or event, the latter gives a description of a person or thing.

## 1.3 Orthography

Arabic is written from right to left in a cursive script. All 28 letters of the alphabet represent consonants, except for aleph which, however, may act as a 'bearer', metaphorically كُرْسيّ *kursiyy* 'chair' of an additional sign. This is the *hamza*, representing the 28th consonant, a glottal stop (Holes 2004, p. 89).

The Arabic script is believed to have originated in the earlier Nabatean script (Bateson 2003, p. 54 ff.). The Nabatean script, itself descended from the Aramaic alphabet, was used first to write the Nabatean dialect of Aramaic, and subsequently for writing Arabic. As Arabic had more consonants than Aramaic, the script was modified to represent the extra Arabic consonants. The ligatures, which were adopted from the early Canaanite alphabets to form cursive script, also resulted in the loss of some phonological distinctions. Therefore, some originally distinct Aramaic letters became indistinguishable in shape, so that in the early writings 15 distinct letter-shapes had to represent 28 sounds.

In order to disambiguate pairs or triplets of letters that were identical in their basic shape (رَسْم *rasm*) and represented multiple sounds, e.g., modern س/ش, غ/ع, د/ذ, ر/ز, ص/ض, ط/ظ, a system of consonant pointing was developed, named إعْجام (*ʔiʕja:m*) 'foreignizing', which consisted in the use of distinguishing dots. Each ambiguous grapheme was allocated a distinct number of dots for each of its sounds, one (ن), two (ت), or three (ث); placement of the dots, above(خ ن) or below (ج ب) the letter was also distinctive. It was not until the eighth century AD that this pointing system was standardized and stabilized as an inherent component of the Arabic alphabet, with the dots eventually considered part of the letter.

The writing system reflects some dialectal differences between the western *ħija:ziy* dialect of early seventh century Mecca, which dictated the Quranic orthography, and the prestigious eastern dialects of Najd, on which subsequent

standardized pronunciation was based a century later (Beeston 1970, p. 26 ff.). Discrepancies between the western and eastern dialects were resolved by diverse means in the script, which could not be altered for its religious sanctity. This had significant repercussions for the resulting orthography. A particularly prominent example is the glottal stop, which had by that time disappeared from the Meccan dialect to be replaced by a glide or long vowel depending on its phonetic environment. This situation is reflected in the consonantal script. So, for example, the word *suʔa:l* 'question' was pronounced *suwa:l* in the Meccan dialect, and written سوال. In the consequent standardization process, the *hamza*, still very much alive in the eastern dialects on which the grammarians of Lower Iraq based their codification decisions, was restored over the consonantal body, and is now written سُؤال with the letter و *W* now acting as the bearer of the *hamza* (Goldenberg 2013, p. 39). Another example of this discrepancy in orthographic convention is the so called أَلِف مَقْصورَة *ʔalif maqṣu:ra* 'shortened aleph'. It often represents a historical Meccan final diphthong /*ay*/, written in the consonantal script with the letter ي *Y* (Beeston 1970, p. 27; Holes 2004, p. 91). In the eastern dialects, however, this diphthong contracted to a long /*a:*/, pronounced [*a*] today, as in the verb *baka:* 'to cry' or the preposition *ʔila:* 'to'. These are written بَكى and إلى respectively, namely with the final ى Y grapheme, but without its diacritic dots.[19]

The adapted Nabatean alphabet did not represent vowels. The Arabic alphabet is thus considered a consonantal alphabet, or an *abjad* (Daniels 1992). An *abjad* is a type of writing system where each symbol always or usually stands for a consonant, leaving the reader to supply the appropriate vowels. This system was nicely suited to the Arabic root and word pattern morphological structure, where the most basic semantic meaning is carried by the consonantal root and where vowel information may be recovered from the vocalic word pattern. Each of the 28 letters of the Arabic alphabet (except aleph) represents a consonant. Three of these letters, ا و ي are called حُروفالعِلّة *ħuru:f al-ʕilla* 'letters of defectiveness'. They act as *matres lectionis* 'mothers of reading' and are used to represent the three Standard Arabic long vowels: high front /*i:*/, high back /*u:*/, and low /*a:*/, respectively. These three letters are also called حُروف اللّين والمَدّ *ħuru:f al-li:n wal-madd* 'letters of softness and elongation' because according to traditional views they indicate elongation of the preceding short vowel sound represented orthographically via a vowel mark (Versteegh 2007b, p. 309). This traditional characterization of the role of ا و ي appears to fit nicely with recent characterizations of the Arabic writing system as a mora-based system (Ratcliffe 2001). According to this view, Arabic letters represent CV moras within syllables. Any additional segment besides the mora, be it vowel length as in a CV: syllable, or another consonant (including a glide) as in a

[19] *ʔalif maqṣu:ra* is glossed by Wright (1975 I, p. 11) as the aleph "that can be abbreviated", in contrast with *ʔalif mamdu:da* 'lengthened aleph', which never shortens. In non-final context the consonantal/*y*/ may re-appear, e.g بَكَيْتَ *bakayta* 'you cried' and إِلَيْكَ *ʔilayka* 'to you', respectively. Another variant of the shortened aleph is actually spelled with an aleph in cases such as the verb غَزا *ɣaza:* 'to raid' from the root *ɣZW*.

CVC syllable, requires an additional letter, as in ما *ma:* 'what' and من *man* 'who', respectively.

The modern Arabic script is thus characterized by two sets of diacritics: the first is graphemic and consists of the dots of *ʔiʕja:m* which, as we saw above, are compulsory and are used for phonetic distinction of letter consonants. The second is phonemic and does not include any dots but rather, other superscripted marks representing the short vowels of Arabic and other features of vocalization. It is known as تَشْكيل *taški:l* 'forming'. The short vowel marks of *taški:l* are called حَرَكات *ħaraka:t* 'motions',[20] and include:

1. *fatħa* فَتْحَة 'opening (of the lips)' for a short /*a*/—a small diagonal accent mark placed above a letter;
2. *kasra* كَسْرَة 'breaking, drawing apart (of the lips)', for a short /*i*/—a similar diagonal mark below a letter;
3. *ḍamma* ضَمّة 'pressing together (of the lips)' for a short /*u*/—a small و *W* placed above a letter;
4. *taški:l* also includes *suku:n* سْكون 'silence', which is a circle-shaped diacritic placed above a letter, indicating that the consonant below is vowelless and closes a syllable; this latter information is important for orthographic segmentation and phonological decoding of the Arabic orthography, especially for beginners, given the predominance of the CV syllable in the phonological structure of Arabic words (Saiegh-Haddad 2007). Besides the four marks described above, *taški:l* also includes شَدّة *šadda*, a small ش Š without its dots (Goldenberg 2013, p. 39) placed above the letter indicating consonant doubling (or lengthening).

The *taški:l* diacritics also include the following less frequent signs:

1. مَدّة *madda* 'elongation', a tilde-like diacritic over an aleph آ, accordingly أَلِف مَمْدودَة *ʔalif mamdu:da* 'lengthened aleph'. The most common context is when a syllable-initial *hamza* (always written above or below an aleph) is to be followed by an aleph (with or without a *hamza*, i.e. vocalic or consonantal)—the two consecutive alephs are replaced by one elongated aleph, e.g., آكِلونَ *ʔa:kilu:na* 'eating' (pl. participle) instead of أاكِلونَ (Wright 1975 I, p. 25).[21]
2. هَمْزَة وَصْل *hamzat waṣl* 'connecting *hamza*' or وَصْلَة waṣla 'connector' (Wright 1975 I, p. 19 ff.) which indicates that a *hamza*, predominantly that of the determiner (*ʔa)l-,* is not pronounced in juncture, e.g., *w- + (ʔa)l-walad⇒wal-walad* 'and the boy' although its bearer, the aleph, is written, as in وَٱلَوَلَد.
3. أَلِف خَنْجَرِيَّة *ʔalif xanjariyya* 'dagger aleph' or 'superscript aleph', a short vertical stroke on top of a consonant indicating a long */a:/* where aleph is normally not written. This diacritic, familiar from some high-frequency words like هٰذا *ha:ða:* 'this', is seldom indicated.

[20] The term *ħaraka:t* refers properly to "the phonemes that are known in the Western tradition as 'short vowels'…" (Versteegh 2007a, p. 232), but often includes the graphemes too.

[21] The *madda* is less frequently written over an aleph designating a long*/a:/* before a *hamza*, e.g., *ja:ʔ* 'he came' is usually written جاء, properly جآء (Wright 1975 I, p. 24).

A distinct sub-category of *taški:l* is عَلامات الإعْراب *ʕala:ma:t al-ʔiʕra:b* '*ʔiʕra:b*-endings'. These have the morpho-syntactic function of indicating mood and case (see Sect. 1.2.4: Morpho-syntax). The modal endings of verbs and the case endings of definite nouns consist of the three Arabic short vowels, and are represented in the Arabic orthography using the same phonemic symbols of *fatħa, kasra, and ḍamma*. The case endings of indefinite nouns in non-pausal status are called تَنْوين *tanwi:n* 'nunation'. Phonologically and orthographically distinct from other diacritical marks, they consist of the three vowel signs doubled to indicate that the vowel sound is followed by the consonant /*n*/: وَلَدٌ *walad**un*** 'a boy (nominative)'; وَلَدًا *walad**an*** 'a boy (accusative)'; وَلَدٍ *walad**in*** 'a boy (genitive)'.

The fact that the Arabic writing system is corroborated by an optional system of *taški:l* to mark vocalization results in two scripts: مَشْكول *maški:l*, a fully vocalized (vowelized or voweled) and an unvocalized script. The bulk of Arabic script is unvocalized. Indeed, *taški:l* is commonly used only in religious texts, in children's literature, and sporadically in ordinary texts when an ambiguity of pronunciation might arise, as its main purpose is to provide a phonetic aid, by showing the correct pronunciation.

It is noteworthy that Arabic also employs a partially vocalized script, where the phonemic diacritics, mainly *fatħa, kasra, ḍamma, suku:n* and *šadda*, necessary for word recognition (or lexical access) are marked word internally, but not the morpho-syntactic *ʔiʕra:b*-endings. This script is used in special purpose texts, such as those intended for native speakers when beginning to read. The main intent of partial vocalizing is to mark the phonological information required for word recognition rather than for accurate declension according to the rules of Standard Arabic.[22]

In the cursive Arabic script all but six letters may ligate (attach) forward, to a following letter. The six exceptions are known as حُروف الرَّفْس *ħuru:f ar-rafs* 'kicking letters' (ا د ذ ر ز و). All letters can ligate back to a preceding letter (unless that happens to be a kicking letter). This state of affairs results in a maximum of four allographic forms per letter, as determined by two factors: its position in the word—initial, medial, or final, and whether or not it ligates forward. The combination of position and ligation creates the four letter forms: a) a form for word-initial letters and word-medial letters preceded by a kicking letter; b) a form for word-medial letters ligating both ways; c) a form for final letters that ligate to the preceding letter, and d) a form for final letters preceded by a kicking letter. It is noteworthy that while a few of the Arabic letters actually have four distinct forms (e.g., غـغغ all representing the consonant /*ɣ*/ or ه ـه ـه all representing the consonant /*h*/), most letters have only two distinct forms with the other two differentiated just by the ligature (e.g., خـخخ representing the consonant /*x*/, or ك ـك ـك representing the consonant /*k*/).

[22] The introduction of vowel marks into the Arabic orthography was initiated by the medieval grammarian *ʔabu: l-ʔaswad ad-duʔali:*, using red dots in different arrangements and positions. This system was changed in the late eighth century by *ʔal-fara:hi:di:* into a system similar to what we see today. *ʔal-fara:hi:di:* found the task of writing Arabic tedious when using two different colors, one for letters and another, red, for vocalization. Also, the ʔi*ʕja:m* (consonant dots) had been introduced by then. This meant that without a color distinction the two systems could become confused. As a result, *ʔal-fara:hi:di:* introduced the use of superscripted letters to mark vocalization, thus distinguishing visually between the two systems, vocalization and consonant diacritics.

Like most scripts, Standard Arabic script is conservative in many respects. For example, it leaves many instances of historical phonological assimilation unmarked, most prominently the assimilation of the consonant *l-* of the determiner *(a)l-* to following 'sun letters' حُروف شَمْسِيَّة *ħuru:f šamsiyya* (ن ظ ط ض ص ش س ز ر ذ د ث ت). This group of letters representing coronal consonants takes this label because the word شَمْس *šams* 'sun' begins with such a letter, in contrast to the word قَمَر *qamar* 'moon', which represents all the other, non-assimilating consonants (Wright 1975 I, p. 15).

## 1.4 Diglossia

Arabic is a prototypical case of the concept *diglossia*, which emerged in sociolinguistic theory to describe a situation in which in a given society there is more than one language variety in complementary functional use. In his famous 1959 article, Ferguson defines diglossia as follows:

> DIGLOSSIA is a relatively stable language situation in which, in addition to the primary dialects of the language (which may include a standard or regional standards), there is a very divergent, highly codified (often grammatically more complex) superposed variety, the vehicle of a large and respected body of written literature, either of an earlier period or in another speech community, which is learned largely by formal education and is used for most written and formal spoken purposes but is not used by any section of the community for ordinary conversation. (p. 336).

According to Ferguson, a diglossic context is characterized by a stable co-existence of two linguistically-related language varieties: a *High,* primarily written, variety and a *Low* spoken variety. These are used for distinct sets of complementary functions and in different spheres of social interaction. The spoken variety, which is the original mother tongue, is almost always held in low esteem and its spheres of use involve informal, interpersonal communication. The literary variety is held in high esteem and is used for written communication and formal spoken communication. Such rigid functional complementarity, it is argued, gives way only to slight and insignificant overlap (Maamouri 1998); in a diglossic context, no section of the community uses the High variety for ordinary conversation. This is arguably "the most important factor in a diglossic situation and one that makes for relative stability" (Keller 1982, p. 90).

Though Ferguson proposes a dichotomy between the spoken and written varieties, he himself recognizes that this is just an abstraction. The much more complex linguistic situation in Arabic diglossia has subsequently been described in terms of levels, or even a continuum, with speakers shifting between as many as four (Meiseles 1980) or five (Badawi 1973) varieties, ranging between colloquial/vernacular and literary/standard forms. It is argued that there are "gradual transitions" (Blanc 1960) between the various varieties, and "theoretically an infinite number of levels" (Bassiouney 2009, p. 15). A code switching approach has also been proposed (Boussofara-Omar 2006, p. 634). We shall continue to use the well-established term 'diglossia' and its derivatives, understanding it in this modern conceptual framework as a continuum along which shifting, switching, and mixing occur constantly.

In diglossic Arabic, children start out speaking a local variety of Spoken Arabic, the one used in their immediate environment: at home and in the neighborhood. Once they enter school, they are formally and extensively exposed to Modern Standard Arabic as the language of reading and writing while Spoken Arabic remains the language of informal speech. Academic school-related speech is conducted in a semi standard variety, known as 'Educated Spoken Arabic' (Badawi 1973), except in Arabic lessons, where Standard Arabic is more dominant, or at least aspired to (Amara 1995). Outside the school milieu, there is a similarly stable co-existence of the two major varieties, each functioning for distinct spheres of social communication: Spoken Arabic is used by all native speakers—young and old, educated and uneducated—for informal and intimate verbal interaction in the home, at work, in the community. Standard Arabic, alternating with Educated Spoken Arabic, is at least expected to be used for formal oral interactions, such as delivering a speech or a lecture, and for writing. Thus, while Spoken Arabic is undoubtedly the primary spoken language, native speakers of Arabic, including young children, are actively and constantly engaged with Standard Arabic as well; they pray, do their homework, and study for their exams in Standard Arabic, and they also watch certain TV programs and dubbed series in this variety. Thus, besides proficiency in using Spoken Arabic, linguistic proficiency in Arabic involves, from an early age, concurrent proficiency in using Standard Arabic.

Moreover, the 'vertical' diglossic scale ranging from High to Low is supplemented by a 'horizontal', interdialectal scale with some prestigious dialects, mainly those of urban centers, serving as a kind of regional, or even national, *dialectal standard* (Holes 2004, p. 49 ff.). Such a prestigious, basically 'urbanite' regional standard may adopt some local 'ruralite' elements, particularly following mass immigrations to the urban center, and become a mixed 'dialectal koiné' (Miller 2006, p. 595) which, in turn, exercises koineizing and leveling effects on the entire region. Prominent regional standards include the contemporary dialects of Damascus, Beirut, Jerusalem, Casablanca and, probably the most prominent of all—the Cairene dialect, with a particularly strong koineizing effect, even outside Egypt (Versteegh 2001, p. 138 ff.). In inter-dialectal communication, speakers of local, 'marginal' dialects may tend to level their dialectal variety and accommodate to the regional dialectal standard, or to the Cairene dialect, to which they are exposed more and more today via the media, movies, and other means.[23]

Despite a rather stable diglossic context, two important developments in recent years are particularly relevant to children, casting doubt on classical definitions of diglossia and supporting the modern continuum conception. One is the introduction of satellite TV, and in particular children's TV channels, which dub children's programs in a Standard-like variety in order to make them available to children from different Spoken Arabic backgrounds. This has meant that Arabic native speaking

[23] Terminology varies here, as in other issues. Bassiouney (2009), for example, avoids the term 'standard' in the context of dialects, i.e. on the horizontal scale. She devotes a chapter (1.2.1, p. 18 ff.) to the difference between 'standard' and 'prestige' in the context of dialects, reserving the term 'standard' for Standard (i.e. modern Literary) Arabic.

children are more exposed, and at a rather early age, to Standard Arabic linguistic structures. The second is the introduction of social media and electronic texting and the widespread availability of these facilities to Arabic speaking children and youth. Electronic messages within this population in many Arabic speaking countries are written in Spoken Arabic (Abu Elhija 2012; Al-Khatib and Sabbah 2008; Haggan 2007; Mostari 2009; Palfreyman and Al-Khalil 2007).

In a diglossic community more than elsewhere, speakers' attitudes to their language and dialect are particularly important, because of the significance of the diglossic duality to everyday life, and the choice inherent in every communicative act. The Arabic language, as is well known, is held in the deepest esteem in the Arab world. This begs the question 'What is the Arabic language?' While writing this paper we were surprised to find ourselves disagreeing (happily, that did not happen too often concerning other issues) on the meaning of the term اللُّغَة العَرَبِيَّة *ʔal-luɣa l-ʕarabiyya* 'the Arabic language' for its speakers. For Elinor, based on her north Palestinian native dialect and several authorities on Arabic sociolinguistics, it is an umbrella term and an abstraction that refers to the full range of spoken varieties as well as to Standard Arabic (Maamouri 1998; Suleiman 2006, p. 173), contrasting with *ʔal-luɣa l-fuṣħa: / l-faṣi:ħa* 'the most eloquent/ eloquent language' for specific reference to the literary varieties, namely Classical, Literary and Standard Arabic. In this approach, Arabic speakers consider themselves monolingual native speakers of *ʔal-luɣa l-ʕarabiyya* 'the Arabic language', regardless of the specific vernacular they may speak. For Roni, however, based on her experience with Negev Arabic and other authorities on Arabic (Bateson 2003, p. 75; Fischer 2006b, p. 397; Holes 2004), the term *(ʔal-luɣa) l-ʕarabiyya* refers just to the pure Classical language, or a literary variety that aspires to that. Children learn it at school, but a speaker of the local Negev dialect would not say to another *ʔana baħkiy maʕak b-al-luɣa l-ʕarabiyya* 'I am speaking Arabic to you' but rather *ʔana baħkiy maʕak ʕaṛabiy.*

### *1.4.1 Differences between Classical and Modern Standard Arabic*

Modern Standard Arabic is a direct descendant of Classical-Literary Arabic and the linguistic structure that we have outlined in the previous section basically applies to both. However, as a modern means for interdialectal communication, Modern Standard Arabic has undergone, and is necessarily still undergoing, several changes. Among these, Bateson (2003) includes: (a) linguistic simplification and reduction of various Classical-Literary Arabic linguistic realizations; (b) a vast shift in the lexicon stemming from technical terminology and borrowing from other languages; (c) stylistic-syntactic variations due to translations from European languages and extensive societal bilingualism; and (d) a strong shift in the realization of Classical-Literary Arabic phonology, with changes in the phonetic realization of consonants and vowels and in the extent of velarization and allophonic variation due to the influence of spoken dialects (for a detailed discussion and examples, see Bateson 2003, pp. 84–92). Given these differences, some scholars use the simple

adjective (*ʔal-luɣa) l-faṣi:ħa* 'the eloquent (language)' to refer to Modern Standard Arabic, keeping it distinct from the superlative (*ʔal-luɣa) l-fuṣħ*a: 'the most eloquent (language)', namely Classical-Literary Arabic.

### *1.4.2 Differences between Literary and Spoken Varieties of Arabic*

There is intimate linguistic relatedness between Classical-Literary Arabic and its contemporary descendent Modern Standard Arabic, and both differ from Spoken Arabic in all linguistic domains. According to Bateson (2003) these include the processes that have occurred in the New Arabic type, to which all the contemporary dialects belong:

*Phonologically*, some consonants (as many as four or five) have been lost; final short vowels have been deleted; long unstressed vowels have been shortened and falling diphthongs have contracted to long vowels; new extra-heavy syllable types have developed including more clusters than were permitted in the old type, and various sorts of stress patterns have emerged.

*Morphologically,* the primary difference lies in the general reduction in inflectional categories. This includes the loss of final short vowels indicating case and mood, accompanied by the general use of the genitive-accusative forms of duals and sound masculine plurals. The dual, originally realized in the nominal, pronominal and verbal systems, has survived only partially in the noun system.

*Syntactically*, Colloquial Arabic has a more complex system of parts of speech than Classical Arabic, including an autonomous system of adverbs. This is in part due to the morphological changes delineated above, especially the loss of certain inflectional categories, which placed a heavier burden on word order.

*Lexically*, Colloquial Arabic is more open to loanwords than Classical Arabic. The primary source language varies from one place to another.

### *1.4.3 Representation of Standard and Spoken Arabic in Orthography*

Arabic orthography is primarily a representation of Classical-Literary-Standard Arabic. It maps Standard Arabic phonology, morphology, syntax, and lexicon. This means that linguistic features of Spoken Arabic, including sounds, words, and syntactic constructions, may not have a conventional form of representation in spelling. It is noteworthy that given the linguistic relatedness and partial overlap between Spoken and Standard Arabic, some Standard Arabic linguistic constructions are also available in some Spoken Arabic dialects, albeit with certain variation. These will naturally have a standard orthographic representation. Moreover, distinctive spoken structures may be phonetically represented. Yet, they do not have a conventional orthographic form.

Arabic orthography maps Standard Arabic consonants and long vowels in a rather regular fashion, with a one-to-one relationship between graphemes and phonemes. This results in a regular and transparent abjad (primarily consonantry) with a one-to-one mapping between the letters of written words and their phonological representation.[24] Morphological representation of this abjad is likewise transparent, with a rather regular mapping of the consonantal root morpheme letters and all other consonantal material, as well as the long vowels of word patterns (however, see Sect. 1.2.3: Morphology).

Despite a rather high degree of feedforward consistency in the relation between orthography and phonology when proceeding from the former to the latter (as in reading), Arabic features a few instances of feedback inconsistency, or opacity, especially in the process of moving from phonology to orthography (as in spelling). The first is the *hamza*, representing the glottal stop. This character ء, originally a small ع (Goldenberg 2013, p. 39), has a variety of different phonologically-conditioned orthographic forms and 'bearers' (see Sect. 1.3: Orthography), depending on preceding and following vowels and their alleged relative 'strength'. Another factor pertains to the absence of أَلِف خَنْجَرِيَّة *ʔalif xanjariyya* 'dagger aleph' (see Sect. 1.3: Orthography) from modern Arabic texts. This means that some words will be pronounced with a long vowel that is not represented in spelling. It is noteworthy, however, that 'dagger aleph' is very rare and limited to high frequency words, such as إِلٰه *ʔila:h* 'god' and هٰذا *ha:ða:* 'this'. This explains the tendency to leave it unmarked in modern Arabic texts. A third source of opacity is the optionally marked consonantal gemination (or doubling, or lengthening) which is represented using the superscript sign *šadda*. According to traditional views the *šadda* must not be omitted because consonantal doubling is phonemic, sometimes morphemic, in Arabic. Yet, most modern everyday writing omits the *šadda* together with the other *taški:l* diacritics. In the absence of the *šadda*, word recognition may be hampered, especially among beginners, yet consonant gemination may still be recovered from the morphological and morpho-orthographic representation of the word, as well as from lexical and contextual cues.

The widespread phonological assimilation process of velarization spread in Arabic is another source of orthographic opacity. In this process non-velarized consonants become velarized through vicinity to a velarized phoneme (see Sect. 1.2.1: Phonology). As such, because primary velarization is phonemic in Arabic, the phonetic realization of these secondarily velarized consonants might coincide with the phonemic representation of other letters in the Arabic alphabet. Consequently some letters become homographic, leading to difficulty in the orthographic encoding of sounds, or spelling. For instance, in connected speech, the first letter ت T in the word تَقَدَّم *taqaddam* 'advance' will tend to be realized with the emphatic sound *[ṭ]* and may therefore be spelled incorrectly with the letter ط which represents this emphatic. The source of this mistake is the uvular /*q*/ which triggers a partial regressive assimilation process of velarization spread, namely a backing effect in the vicinity of low vowels making spelling of these letters more difficult (Saiegh-Haddad 2013).

[24] Note that the phonetic realization of consonants, as allophonic variants of phonemes, is not graphemically marked. This is salient in the case of widespread phonological assimilation processes, such as velarization spread (see Sect. 1.2.1: Phonology).

Two morpho-phonological features are noteworthy here as additional factors contributing to orthographic opacity. One is تاء مَرْبُوطَة *ta:ʔ marbu:ṭa* 'bound T'; another is ألِف الفارقة *ʔalif al-fa:riqa* 'separating aleph'. *ta:ʔ marbu:ṭa* is not an independent letter of the Arabic alphabet. Rather, it is a variant of the letter T ت. The basic variant is called *ta:ʔ maftu:ħa* 'opened T', as the grapheme is open at the top. The 'bound' variant, 'closed' at the top, is in fact the letter H (word final shapes) ـه, a matre lectionis with diacritics ـة to differentiate it from the consonantal H and to mark it as a morphological entity, namely the basic feminine suffix of nouns and adjectives. When a word ending with *ta:ʔ marbu:ṭa* is suffixed with a personal pronoun, the consonant /t/ is restored in both speech and in writing as *ta:ʔ maftu:ħa* 'opened T'; when vocalized for the case ending, or opening a construct *ʔiḍa:fa*, the consonant /t/ is restored in speech only. It is argued that the feminine suffix used to be /t/ in all circumstances, then was realized as *[h]* in pausal status, and finally was muted to [*a*]; the letter representing it comes from the middle stage, H for [*h*] combined with the two dots of the letter ت T. Because *ta:ʔ marbu:ṭa* usually sounds like /*a*/ in pausal status and as /t/ in junctural speech as well as in suffixed and construct status, it may be confused with the *fatħa* in the former and with the letter ت in the latter. This may constitute a source of difficulty in early spelling, especially in the case of *ta:ʔ marbu:ṭa,* because while spelling in Arabic does not typically encode short vowel marks, omitting *ta:ʔ marbu:ṭa* is considered a spelling error.

In Standard Arabic, perfective verbs in the third person plural, such as *katabu:* 'they wrote' and imperfective verbs in the subjunctive and jussive moods, e.g., *lan yaktubu:* 'they will not write' and *lam yaktubu:* 'they did not write', respectively, end with a suffix called واو الجَماعَة *wa:w al-jama:ʕa* 'plural W'. In spelling, this suffix consists not only of the letter و W, as expected, but also of the letter aleph ا. This aleph is called ألِف الفارقة *ʔalif al-fa:riqa* or ألِفَ الفاصِلة *ʔalif al-fa:ṣila* 'separating aleph' or ألِف الوِقايَة *ʔalif al-wiqa:ya* 'aleph of protection', having served in the past to distinguish this suffix from the conjunction و W 'and' (Holes 2004, p. 92; Wright 1975 I, p. 11) at a time when words were not separated by spaces. As this aleph is silent, it may be missed in spelling or, conversely, wrongly vocalized in reading.

Vocalized Arabic is highly transparent for reading, since all of the phonological information required for accurate pronunciation is marked, and is regular. Excluded are secondarily velarized consonants and vowels. In contrast, unvocalized Arabic is rather opaque. This is because the phonological information represented through *taški:l*—mainly the system of vowel marks—is missing from this script. It is noteworthy here that the terms 'orthographic regularity' and 'orthographic opacity' refer to fundamentally different underlying phenomena in Arabic and in English. In English, orthographic opacity does not stem from the absence of the graphemes that represent phonological information, but rather from the ambiguity or lack of systematicity in the mappings between graphemes and phonemes. Such orthographic opacity necessitates reliance in reading and spelling on large grain-size units (Ziegler and Goswami 2005), primarily lexical. The Arabic unvocalized orthography, in contrast, represents the morphological structure rather regularly, with full representation of root consonants, as well as the consonants and long vowels of word-patterns. Given that the great majority of Arabic words are complex and have

an internal root-pattern morphological structure, the sub-lexical morphemic grain-size unit appears to be a functional linguistic unit in reading and spelling in Arabic (Frost 2006; Saiegh-Haddad 2013; Ravid 2012).

Given the systematic representation of morphemes in the Arabic unvocalized orthography, fully vocalized Arabic may be paradoxically more opaque than unvocalized Arabic, especially for spelling. This opacity is not related to orthographically regular vowel marks. Rather, it pertains to the case endings, in particular *tanwi:n*, which is not necessary for lexical access and which is associated with a number of orthographic-phonological complexities, such as the nasal sound /*n*/ that it represents, as well as its effect on the phonological quality of *ta:ʔ marbu:ṭa*. Similarly, other *ʔiʕra:b*-endings which take the form of short vowels may be mistaken for the mothers of reading (ي و ا) especially among children and beginners who fail to make accurate auditory discrimination between short and long vowels and who cannot use higher order linguistic skills to compensate for difficulties in phonological representation and awareness.

We have argued above that fully vocalized Arabic is highly transparent with graphemes (letters and diacritics) representing phonemes regularly. We have also argued that unvocalized Arabic is also highly consistent with morphemes fully and regularly represented. This claim is true, however, only if the mapping systems that we consider are Standard Arabic, on the one hand, and its orthographic representation, the Arabic orthography, on the other. Yet, from a psycholinguistic point of view, the Arabic orthography might not be said to be transparent for two reasons. First, at a higher linguistic level, it does not map the language structures (syntax, lexicon, etc.) that native Arabic speakers naturally use and master. Further, at a lower-order level, the symbolic system in the case of Arabic maps phonological units that may be unfamiliar to readers (Saiegh-Haddad 2003, 2004, 2005, 2007, 2011, 2012; Saiegh-Haddad et al. 2011). This implies that the mapping from spelling to sound, while it may be considered linguistically regular at some abstract level, may be regarded as psycholinguistically opaque.

In this chapter, we have attempted a general description of the Arabic language and orthography, with particular focus on phonological and morpho-syntactic properties, as well as on the mappings from language to orthography. This focus on phonology, morpho-syntax, and orthography was guided by our intent to provide the reader with those aspects of the Arabic language and orthography that may have a direct relevance to reading research and practice in Arabic. While it does not claim to be a comprehensive account of this extremely complex topic, we believe it provides the reader with the necessary ‘springboard’ for the rest of the book.

## References

Abu Elhija, D. (2012). Qaf in Nazarene and Iksali electronic writing. *e-Journal of non- Sequitur* 1(2). Israel: The University of Haifa (pp. 110–116). http://nonsequiturjournal.wix.com/english#. Accessed April 2013.

Akesson, J. (2009). ṣarf. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. IV, pp. 118–122). Leiden: E. J. Brill.

Al-Ani, S. H. (2008). Phonetics. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. III, pp. 593–603). Leiden: E. J. Brill.

Al-Khatib, M., & Sabbah, A. E. H. (2008). Language choice in mobile text messages among Jordanian university students. *SKY Journal of Linguistics,* 21*,* 37–65.

Amara, M. H. (1995). Arabic diglossia in the classroom: Assumptions and reality. In S. Izre'el & R. Drory (Eds.), *Israel Oriental Studies*, 15*, Language and culture in the Near East* (pp. 131–142). Leiden: E. J. Brill.

Badawi, E. (1973). *Levels of modern Arabic in Egypt*. Cairo: dar al-maʕarif. In Arabic.

Bakalla, M. H. (2007). iṭba:q. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. II, pp. 459–461). Leiden: E. J. Brill.

Bakalla, M. H. (2009). tafxi:m. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. IV, pp. 421–424). Leiden: E. J. Brill.

Bassiouney, R. (2009). *Arabic sociolinguistics*. Edinburgh: Edinburgh University Press.

Bateson, M. C. (2003). *Arabic language handbook*. Washington, D.C.: Georgetown University Press.

Beeston, A. F. L. (1970). *The Arabic language today*. London: Hutchinson University Library.

Bentin, S., & Frost, R. (1995). Morphological factors in visual word recognition in Hebrew. In L. Feldman (Ed.), *Morphological aspects of language processing* (pp. 217–292). Hillsdale: Erlbaum.

Blanc, H. (1960). Style variations in spoken Arabic: A sample of inter-dialectal conversation. In C. Ferguson (Ed.), *Contributions to Arabic linguistics* (pp. 81–158). Cambridge: Harvard University Press.

Blanc, H. (1970). *The Arabic dialect of the Negev Bedouins. Proceedings of the Israel Academy of Sciences and Humanities* (Vol. 4, pp. 112–150). Jerusalem: Israel Academy of Sciences and Humanities.

Boudelaa, S., & Marslen-Wilson, W. D. (2010). Aralex: A lexical database for modern standard Arabic. *Behaviour Research Methods,* 42*,* 481–487.

Boussofara-Omar, N. (2006). Diglossia. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. I, pp. 629–637). Leiden: E. J. Brill.

Broselow, E. (2008). Phonology. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. IV, pp. 607–615). Leiden: E. J. Brill.

Chekayri, A. (2007). Glide. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. II, pp. 164–169). Leiden: E. J. Brill.

Daniels, P. T. (1992). The syllabic origin of writing and the segmental origin of the alphabet. In P. Downing, S. D. Lima, & M. Noonan (Eds.), *The linguistics of literacy* (pp. 83–110). Amsterdam: John Benjamins.

Davies, H. (2006). Dialect literature. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. I, pp. 597–604). Leiden: E. J. Brill.

Davis, S. (2009). Velarization. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. IV, pp. 636–638). Leiden: E. J. Brill.

Ferguson, C. (1959). Diglossia. *Word,* 15*,* 325–340.

Fischer, W. (2006a). Adjective. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. I, pp. 16–21). Leiden: E. J. Brill.

Fischer, W. (2006b). Classical Arabic. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. I, pp. 397–405). Leiden: E. J. Brill.

Frisch, S. A. (2008). Phonotactics. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. III, pp. 624–628). Leiden: E. J. Brill.

Frost, R. (2006). Becoming literate in Hebrew: The grain-size hypothesis and Semitic orthographic systems. *Developmental Science,* 9*,* 439–440.

Goldenberg, G. (2013). *Semitic languages: Features, structures, relations, processes*. Oxford: Oxford University Press.

Haggan, M. (2007). Text messaging in Kuwait. Is the medium the message? *Multilingua,* 26*,* 427–449.

Henkin, R. (2010). *Negev Arabic: Dialectal, sociolinguistic, and stylistic variation*. Wiesbaden: Otto Harrassowitz. (Semitica Viva Series no. 48).
Holes, C. (2004). *Modern Arabic: Structures, functions, and varieties*. Washington, D.C.: Georgetown University Press.
Iványi, T. (2006). Diphthongs. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. I, pp. 640–643). Leiden: E. J. Brill.
Jesry, M. (2009). Syllable structure. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. IV, pp. 387–389). Leiden: E. J. Brill.
Kager, R. (2009). Stress. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. IV, pp. 344–353). Leiden: E. J. Brill.
Keller, R. (1982). Diglossia in German-speaking Switzerland. In W. Haas (Ed.), *Standard languages: Spoken and written* (pp. 70–93). Manchester: Manchester University Press.
Larcher, P. (2006). Derivation. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. I, pp. 573–579). Leiden: E. J. Brill.
Larcher, P. (2009). Verb. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. IV, pp. 638–645). Leiden: E. J. Brill.
Levin, A. (2007). *ʔima:la*. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. II, pp. 311–315). Leiden: E. J. Brill.
Maamouri, M. (1998). *Language education and human development: Arabic diglossia and its impact on the quality of education in the Arab region*. World Bank, Mediterranean Development Forum.
McCarthy, J. (1981). A prosodic theory of non-concatenative morphology. *Linguistic Inquiry,* 12*,* 373–418.
Meiseles, G. (1980). Educated spoken Arabic and the Arabic language continuum. *Archivum Linguisticum,* 11*,* 118–143.
Miller, C. (2006). Dialect Koiné. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. I, pp. 593–597). Leiden: E. J. Brill.
Mostari, H. A. (2009). What do mobiles speak in Algeria? Evidence from language. *Current Issues in Language Planning,* 10*,* 377–386.
Palfreyman, D., & Al-Khalil, M. (2007). A funky language for Teenzz to use: Representations of Gulf Arabic. In B. Danet & S. C. Herring (Eds.), *The multilingual internet: Language, culture, and communication online* (pp. 43–63). Oxford: Oxford University Press.
Ratcliffe, R. R. (2001). What do "phonemic" writing systems represent? *Written Language and Literacy,* 4*,* 1–14.
Ravid, D. (2012). *Spelling morphology: The psycholinguistics of Hebrew spelling*. New York: Springer.
Saiegh-Haddad, E. (2003). Linguistic distance and initial reading acquisition: The case of Arabic diglossia. *Applied Psycholinguistics,* 24*,* 431–451.
Saiegh-Haddad, E. (2004). The impact of phonemic and lexical distance on the phonological analysis of words and pseudo-words in a diglossic context. *Applied Psycholinguistics,* 25*,* 495–512.
Saiegh-Haddad, E. (2005). Correlates of reading fluency in Arabic: Diglossic and orthographic factors. *Reading and Writing: An Interdisciplinary Journal,* 18*,* 559–582.
Saiegh-Haddad, E. (2007). Linguistic constraints on children's ability to isolate phonemes in Arabic. *Applied Psycholinguistics,* 28*,* 605–625.
Saiegh-Haddad, E. (2011). Phonological processing in diglossic Arabic: The role of linguistic distance. In E. Broselow, & H. Ouli (Eds.), *Perspectives on Arabic Linguistics XXII* (pp. 269–280). John Benjamins Publishers.
Saiegh-Haddad, E. (2012). Literacy reflexes of Arabic diglossia. In M. Leikin, M. Schwartz, & Y. Tobin (Eds.), *Current issues in bilingualism: Cognitive and sociolinguistic perspectives* (pp. 43–55). Springer.
Saiegh-Haddad, E. (2013). A tale of one letter: Morphological processing in early Arabic spelling. *Writing Systems Research*, 5, 169–188.
Saiegh-Haddad, E., Levin, I., Hende, N., & Ziv, M. (2011). The linguistic affiliation constraint and phoneme recognition in diglossic Arabic. *Journal of Child Language,* 38*,* 297–315.

Shawarbah, M. (2012). *A grammar of Negev Arabic: Comparative studies, texts and glossary in the bedouin dialect of the ʻAzāzmih tribe*. Wiesbaden: Harrassowitz.

Suleiman, Y. (2006). ʻArabiyya. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. I, pp. 173–177). Leiden: E. J. Brill.

Versteegh, K. (2001). *The Arabic language*. Edinburgh: Edinburgh University Press.

Versteegh, K. (2007a). *ḥaraka*. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. II, pp. 232–236). Leiden: E. J. Brill.

Versteegh, K. (2007b). *illa*. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. II, pp. 308–311). Leiden: E. J. Brill.

Versteegh, K., et al. (Eds.). (2006–2009). *Encyclopedia of Arabic language and linguistics* (Vols. I-IV). Leiden: E. J. Brill.

Voigt, R. (2009). Weak verbs. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. IV, pp. 699–708). Leiden: E. J. Brill.

Watson, J. C. E. (2002). *The phonology and morphology of Arabic*. Oxford. Oxford University Press.

Wright, W. (1975). *A grammar of the Arabic language* (3rd ed.). Cambridge: Cambridge University Press.

Zemánek, P. (2006a). Anaptyxis. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. I, pp. 85–86). Leiden: E. J. Brill.

Zemánek, P. (2006b). Assimilation. In K. Versteegh (Ed.), *The encyclopedia of Arabic language and linguistics* (Vol. I, pp. 204–206). Leiden: E. J. Brill.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin,* 131, 3–29.