Syllable Structure in the Dialects of Arabic

1. INTRODUCTION

Although syllables are not directly recoverable from the speech signal, the assumption that segments are grouped into syllables has proven useful in explaining numerous phonological patterns of individual languages. The syllable is therefore viewed as an abstract constituent of the mental representation of sound structure (Al-Ani and May 1973), and the analysis of the syllable structure of any individual language must be grounded in comprehensive analysis of the overall phonological structure of that language.

The theory of syllable structure has been strongly influenced by the study of Arabic, for two reasons. First, many dialects of Arabic provide examples of regular and productive processes that make reference to syllable structure, as well as clear diagnostics for syllable division, including insertion of a vowel into clusters of consonants that could not otherwise be accommodated in the inventory of possible syllables and stress systems based on syllable weight (see references in section 3). Second, different varieties of spoken Arabic instantiate different inventories of syllable types, and these different inventories are associated with a constellation of different properties across the varieties (e.g., Fischer and Jastrow 1980, Mitchell 1993). Arabic therefore provides invaluable data for the linguist's quest to identify the range of variation across human languages.

This chapter surveys the types of evidence that have been used to argue for syllable structure in Arabic; the range of variation in syllable structures across different dialects; the competing analyses of syllable-sensitive phonological processes; and the implications of the Arabic data for theories of sound structure. Section 2 begins with an overview of the internal structure of syllables, the concept of syllable weight, and the relationship between syllables and higher level constituents. In section 3, we turn to specific processes that depend on syllable structure: word stress, vowel shortening, vowel insertion, and vowel deletion. Section 4 reviews proposals concerning the correlations among different properties of syllable structure and the typological claims based on these correlations.

In the following discussion, forms in square brackets represent transcriptions of surface forms using the International Phonetic Alphabet, while those between slashes represent underlying representations. Long vowels are represented as a sequence of two identical vowels; breaks between syllables are indicated with a period; morpheme boundaries and word boundaries are indicated by '+' and '#', respectively.

2. CRITICAL ISSUES AND TOPICS

This section reviews common assumptions related to the internal structure of syllables and the relationship between syllables and larger units of prosody, situating the discussion of Arabic within the context of hypotheses concerning the range of cross-linguistic variation in syllable structure. We identify the structural features that are shared by all or most varieties of Arabic as well as the dimensions along which dialects diverge.

2.1. Syllable-Internal Constituents: Onset, Nucleus, and Coda

Each syllable in a word represents a peak in prominence. The most prominent portion of a syllable is its *nucleus*, which is typically a vowel (V). Every syllable must have a nucleus, but syllables may in addition have one or more consonants (C) organized around that nucleus.

Consonant(s) preceding the nucleus constitute the *syllable onset* and consonants following the nucleus constitute the *syllable coda*. Onsets and codas are called *syllable margins*, and margins containing a single consonant are called *simple onsets* or *codas*, while those containing more than one consonant are *complex onsets* or *complex codas*. Syllables that end in a vowel (lack a coda) are called *open syllables*, while syllables containing a coda are called *closed syllables*.

While a syllable containing only a nucleus would appear to be the most basic structure, it is generally agreed that the universally preferred syllable consists of a single-consonant onset plus a nucleus (Blevins 1995). CV syllables are found in every language, and some languages (e.g., Hua of Papua New Guinea) have only CV syllables. However, alongside CV syllables, most languages admit at least one of the following additional options: onsetless (vowel-initial) syllables; syllables with complex onsets; and syllables containing codas of one or more consonants. A well-established implicational universal concerns the complexity of syllable margins: any language that allows complex onsets/codas also allows simpler onsets/codas.

Most and possibly all varieties of Arabic share two features: the presence of syllables with simple codas and the absence of onsetless syllables. Thus, syllables of the shape CV and CVC are found in all varieties of Arabic, while V or VC syllables are unattested. The requirement that syllables must have onsets implies that in a VCV sequence, the syllable boundary will always fall before the consonant (V.CV). Evidence for the onset requirement comes from the fact that vowel-initial words borrowed into Arabic are typically produced by Arabic speakers with an added glottal stop in onset position (e.g., [ʔotobiis] 'autobus'). Similarly, *hamzat al-wasl*, a glottal stop, appears before vowel-initial morphemes (e.g., [ibn] 'son') in contexts where there is no preceding consonant to provide an onset.

Divergence among Arabic varieties is found in the types of segments that may serve as syllable nuclei and in the possibility (and composition) of complex onsets or codas. In syllable nuclei, languages tend to favor segments of high sonority, where relative sonority may be defined according to some version of the following scale (Clements 1990):

(1) Sonority Scale:
most sonorous.....least sonorous
vowel.....glide.....liquid.....nasal.....obstruent (fricative, stop)

While it is not uncommon to find languages with relatively high sonority consonants in nucleus position (e.g., the liquid [l] in the second syllable of the English word *cattle*), languages that allows nuclei to consist of obstruents (fricatives and stops) are far more rare. Although the analysis of Moroccan Arabic syllable structure has been a topic of debate for decades (see, e.g., Harrell 1962, Dell and Elmedlaoui 2002)--with researchers disagreeing even on basic transcription--Moroccan Arabic (like some other North African dialects) is a strong candidate for a language with obstruent syllable nuclei. Most recently, Shaw et al. (2009, 2011) provide evidence that the timing of articulatory gestures in words like Moroccan [ktab] 'book' supports the analysis of this word as disyllabic [k.tab] rather than as a monosyllable with a complex onset. The property of allowing syllables with low-sonority nuclei may have its origin in contact with Berber dialects, which appear to share this feature (Dell and Elmedlaoui 2002, Versteegh 1997).

Another dimension along which Arabic varieties differ is the tolerance (or lack thereof) of complex syllable margins (Mitchell 1993, Kiparsky 2003). Complex onsets are proscribed in Modern Standard Arabic as well as in dialects of Cairo and environs but are common in many

other varieties, as illustrated by the pronunciation of 'big' as disyllabic [ki.biir] in Cairene Arabic but as monosyllabic [kbiir] in Syrian Arabic, for example. In contrast, complex codas are found in Modern Standard Arabic, in Egyptian dialects, and in Moroccan, but are missing in many other dialects, as illustrated by the pronunciation of 'dog' as [kalb] in Cairene Arabic but as [ka.lib] in Eastern Libyan Arabic. The analysis of syllable structure is complicated by the fact that in many dialects, complex margins are possible only at phrase edges, a topic we return to below.

Dialects that permit complex syllable margins may be further distinguished by their restrictions on possible sequences of consonants within the margin. Segments within a syllable tend to be arranged according to the Sonority Sequencing Principle (Selkirk 1984, Clements 1990) which dictates that sonority should increase approaching the syllable nucleus; for example, [kl], which rises in sonority, is more common in onset position than the falling sonority sequence [lk], while the converse is true for coda position. Certain varieties, such as Lebanese Arabic (Kenstowicz 1986) and possibly Tripoli Arabic (Kenstowicz and Abdul-Karim 1980, Kiparsky 2003), permit only falling sonority complex codas, as in [ħilm] 'dream'; sequences of rising sonority are broken by insertion of a vowel, as in [ħimil] 'load', where the underlined vowel is inserted (though Watson (2007:345) argues that vowel insertion in Tripoli Arabic is dependent on style rather than sonority). Cairene Arabic is unusual among the world's languages in allowing any combination of word-final consonants regardless of sonority, as for example in [ʔakl] 'food' with low sonority [k] (a stop) followed by higher sonority [l] (a liquid).

Despite the apparent freedom with which two consonants can combine to form a complex coda in Cairene, no words of this dialect may end in more than two consonants. According to Mitchell, "Final clusters of more than two consonants are rare outside Morocco" (Mitchell 1993: 69). While San'ani Arabic contains apparent triconsonantal codas arising via suffixation of the negative marker /ʃ/, as in [maa gultʃ] 'I didn't say', Watson argues that in such clusters, /t/ and the negative marker fuse to produce a single segment, an affricate (Watson 2002: 59-61).

Initial clusters of more than two consonants are also quite rare, though they do occur in some Levantine dialects in limited morphological contexts, as in the masculine singular imperative form [striiħ] 'rest!' (Mitchell 1993: 60).

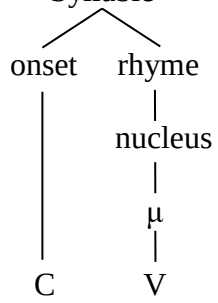
2.2. Syllable Weight and Syllable Rhyme

In many languages, phenomena such as the position of stress within a word and the minimal size of words are dependent on the composition of nucleus and coda (Hayes 1995, . Such facts have motivated the grouping of nucleus and coda into a constituent called the *syllable rhyme* (or *rime*): that portion of the syllable that is identical in rhyming words (e.g., English *meets, sweets, streets*).

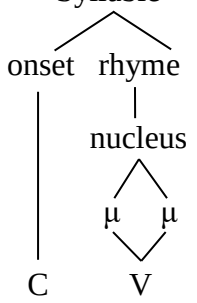
The composition of the syllable rhyme determines the *weight* of a syllable, where a unit of weight is called a *mora* (plural *moras* or *morae*, represented as μ). Each short vowel is associated with a single mora while each long vowel is associated with two moras; CV syllables are monomoraic, called *light syllables*, and CVV syllables are bimoraic, called *heavy syllables*. In many languages, including many Arabic dialects, CVV and CVC syllables pattern together in opposition to CV syllables, suggesting that each coda consonant is also associated with a mora. The structure of these different syllables is illustrated in (2):

(2) Syllable-internal structure

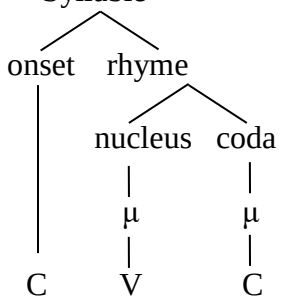
a. Light CV



b. Heavy CVV



c. Heavy CVC



The stress patterns of many varieties of Arabic are crucially dependent on syllable weight. The forms in (3) from Syrian Arabic illustrate the attraction of stress to the penultimate syllable when that syllable is heavy (3a) but not when it is light (3b):

(3) Syrian Arabic (Cowell 1964)

a. stress on penultimate syllable

ma. ˈdii. na	‘city’
da. ˈras. tu	‘I studied it (masc.)’

b. stress on antepenultimate syllable

ˈda. ra. su	‘they studied’

The regular patterns of word stress found in many Arabic dialects provide useful diagnostics for locating the boundaries between syllables (a matter of debate in English). For example, the stress contrast in (3) provides evidence that the [s] in [darastu] ‘I studied it (masc.)’ must be assigned to the coda of the preceding syllable rather than the onset of the following syllable, even though this dialect does admit [st] onsets. Similarly, the treatment of geminate (long) consonants with respect to stress indicates that geminates must be analyzed as contributing to the weight of the syllable they close, since stress falls on the penultimate syllable in words like [bi. ˈtam. mu] ‘they remain.’ We return to the role of moraic structure in conditioning word stress in section 3.

2.3. Syllables, Words, and Phrases

One view of the relationship among different levels of prosodic structure is the Strict Layering Hypothesis (Nespor and Vogel 1986) which asserts that each constituent is fully contained within and directly dominated by a constituent of the next highest level. On this hypothesis, segments are exhaustively grouped into syllables, syllables into words, and words into phrases. The implication is that the edge of a word or phrase will necessarily also be the edge of a syllable, and vice versa.

Many varieties of Arabic defy this simple picture. First, as mentioned above, many dialects restrict complex syllable margins to word or phrase edges. Second, the edges of words and syllables are not necessarily aligned. For example, in a process similar to French liaison, the

Cairene phrase [ka. ta. b# il. ga. waab] ‘he wrote the letter’ displays a mismatch between syllable structure and word structure, with the final consonant of [katab] ‘he wrote’ serving as onset to the initial syllable of the following word (Broselow 1984, Watson 2002). Thus, in some varieties of Arabic it is the phrase, rather than the individual word, that is the domain of syllabification.

3. SYLLABLE-RELATED PROCESSES

Arabic provides a particularly rich set of processes that bear on the analysis of syllable structure. In this section we consider word stress, which provides diagnostics for syllable weight; vowel insertion after CVVC and long vowel shortening; vowel insertion into consonant sequences; and vowel deletion.

3.1. Word Stress: Weight Contrasts and the Representation of Word-Final Consonants

In many Arabic dialects, word stress patterns depend on syllable weight. In Syrian Arabic, where (as in many varieties) stress falls on one of the three final syllables of a word, a word-final syllable is stressed if and only if that syllable consists of CVVC or CVCC. Lacking such a final syllable, stress falls on a heavy penultimate syllable if one is present, and otherwise on the antepenultimate:

(4) Syrian Arabic Word Stress

a. stress on final syllable		
	ki. ˈtaab	‘book’
	ma. ˈmarr	‘aisle’
b. stress on penultimate syllable		
	ki. ˈtaa. bak	‘your (masc.) book’
	ka. ˈtab. na	‘he wrote’
c. stress on antepenultimate syllable		
	ˈka. ta. bu	‘they wrote’

The question raised by this pattern is why a heavy syllable CVC is sufficient to attract stress in penultimate position but not in final position, where only CVVC or CVCC can attract stress.

The stress facts suggest the necessity for recognizing a three-way distinction in syllable weight between light, heavy, and *superheavy* (or *hyperlong*) *syllables* (CVVC, CVCC). This three-way distinction is consistent with the assumption that each coda consonant adds a mora to the syllable; since CVC syllables behave as bimoraic, it is reasonable to conclude that CVVC and CVCC should be trimoraic, although languages that require reference to a ternary oppositions of syllable weight are relatively rare (Hayes 1995, Broselow, Chen, and Huffman 1997).

An alternative analysis is one in which no word-final consonant contributes to the weight of its syllable, making final CVC light and final CVVC, CVCC heavy (bimoraic). Such an account offers a simple analysis of the Syrian stress pattern: stress falls on the rightmost heavy syllable (within the three-syllable stress window). This account has the virtue of explaining the fact that CVV, like CVVC, generally attracts stress in final position--as in Syrian [da. ra. ˈsuu] ‘they studied it (masc.)’-- without the necessity of assuming an abstract final [h] in such forms (Brame 1971, McCarthy 1979)—an assumption that, while consistent with earlier stages of Arabic (Birkeland 1952), does not account for the fact that the posited final [h] is at best only

optionally pronounced. Thus, the facts of word stress in many dialects support an analysis in which word-final consonants play a role in syllable structure that is different from the role of other coda consonants.

A variety of formal mechanisms have been marshaled to account for the special behavior of word-final consonants as well as the representations of superheavy syllables. One approach abandons the Strict Layering Hypothesis, either by assuming a recursive structure in which the syllable node dominates an additional syllable node associated with the final consonant (McCarthy 1979) or by allowing a word-final consonant (or the mora dominating it) to attach directly to the Word node (Kenstowicz 1986, Kiparsky 2003). A second approach renders word-final consonants invisible (extrametrical) at some abstract pre-surface level, incorporating the consonant into the syllable only after stress is assigned (Hayes 1995). A third alternative is to represent a word-final consonant as the onset to an empty syllable nucleus (Selkirk 1981, Angoujard 1990), an approach that is compatible with the claim made in the framework of Government Phonology that at an abstract level, CV is not only the preferred but the only possible syllable type (Lowenstamm 1996). And a fourth approach assumes that a word-final consonant shares a mora with the preceding segment rather than being associated with its own mora (Broselow et al. 1995, Broselow et al. 1997, Watson 2002, Watson 2007). We return to the issues surrounding the moraic structure of superheavy syllables in the next section.

3.2. Vowel Insertion, Vowel Shortening, and Mora Sharing in CVVC

The special status of word-final consonants in prosodic structure is reinforced by the fact that in a number of Arabic dialects, CVVC syllables are found only in word-final position. The large number of consonant-initial suffixes should give rise to non-final CVVC. However, dialects exhibit several different strategies for avoiding internal CVVC.

Vowel insertion, illustrated by data from Makkan Arabic, places a vowel (underlined below) between stem-final CVVC and a following consonant within the word (Abu-Mansour 1992), moving the stem-final C to onset position:

(5) Makkan Arabic: vowel insertion

ki. taab	‘book’
ki. taa. bi	‘my book’ (/kitaab + i/)
ki. taa. b<u>a</u>. na	‘our book’ (/kitaab + na)
mak. tab. na	‘our office’ (/maktab + na/)

An alternative strategy for avoiding CVVC is shortening of the long vowel in a closed syllable, characteristic of Cairo and the central and western dialects of the Nile Delta (Broselow 1992, Mitchell 1993, Watson 2002):

(6) Egyptian Arabic: vowel shortening

ki. taab	‘book’
ki. taa. bi	‘my book’ (/kitaab + i/)
ki. tab. na	‘our book’ (/kitaab + na/)

This process should be distinguished from the morphologically restricted vowel shortening processes found in both Makkan (Abu-Mansour 1992) and San’ani (Watson 2002). In Makkan,

for example, the long vowel of hollow verbs is shortened before the preposition /l/ 'to/for', as illustrated by contrast between the non-suffixed form [siib] 'leave!' and [siblahum] (/siib+l+hum/) 'leave for them', though the phonologically similar CVVC+C context induces epenthesis, rather than shortening, in forms like [nisiitaha] (/nisiit+ha) 'I forgot her' (Abu-Mansour 1992: 50).

One possible analysis of the motivation for both vowel insertion and vowel shortening is that both processes constitute strategies for enforcing a prohibition on trimoraic syllables. As discussed above, the facts of word stress suggest that in many dialects, word-final consonants do not add to syllable weight. The assumption that long vowels are bimoraic and that only non-final coda consonants add a mora to the syllable implies that only word-final CVVC syllables are bimoraic, while word-internal CVVC syllables must be trimoraic. The assumption that trimoraic syllables are avoided is therefore consistent with the facts of dialects like Cairene Arabic, which allows CVVC syllables only in word-final position. However, many other varieties of Arabic do include word-internal CVVC syllables; for example, in Sudanese, the dialects of Upper Egypt, and the dialects of the Levant and of the Gulf (Farwaneh 1995, Broselow 1992, Broselow et al. 1995), CVVC syllables are maintained after suffixation (e.g., [baab.na] 'our door'), though in some cases they occur alongside variants with vowel insertion (e.g., San'ani [ki.taab.na] ~ [ki.taa.ba.na] 'our book', Watson 2002:69).

The difference among dialects that allow non-final CVVC and those that lack it could be characterized simply in terms of a difference in the tolerance of trimoraic syllables. Most research on this topic, however, has agreed that the prohibition against trimoraic syllables holds even in dialects that permit non-final CVVC, which in these dialects are actually bimoraic. (The analysis of CVCC syllables is discussed in section 3.3).

One approach to representing CVVC as bimoraic is to deny the final consonant membership in the syllable containing the long vowel. This is the approach propounded by Kiparsky (2003), who characterizes CVVC as consisting of a core syllable CVV followed by a *semisyllable*: a consonant associated to the Word node, through an intervening mora but with no intervening Syllable node (in violation of Strict Layering). In Kiparsky's analysis, tolerance of nonfinal CVVC corresponds to tolerance of semisyllables, a factor which he connects with a wide range of differences across the dialects (discussed in section 4).

A second approach characterizes CVVC as contained within a single syllable, but prevents the final consonant from adding weight to the syllable by allowing it to share the second mora of the preceding vowel rather than contributing its own mora (Broselow 1992, Broselow et al. 1995, 1997; Watson 2002, 2007). In the mora-sharing analysis, tolerance of nonfinal CVVC corresponds to tolerance of mora sharing, and it is only dialects in which this is not an option that either insert a vowel after CVVC, as in Makkan Arabic, or incorporate the postvocalic C into the syllable by removing one of the vocalic moras, as in Cairene Arabic.

Support for the mora sharing analysis comes from the phonetics. In a comparison of the duration of rhyme segments of non-final CV, CVC, CVV, and CVVC produced by speakers of CVVC-tolerating dialects Jordanian, Syrian, and Lebanese, Broselow et al. (1995, 1997) found that short vowels had approximately the same duration in closed and in open syllables, while long vowels were shorter in CVVC than in CVV syllables, though still longer than underlying short vowels. Thus, they found a three-way difference in vowel length: [aa] was shorter in Levantine [kitaabhum] 'their book' than in [kitaabi] 'my book', yet still longer than the short [a] in [ʕinabhum] 'their grape.' This asymmetry is consistent with the assumption that the

prohibition on trimoraic syllables forces a long vowel to share a mora with the following coda consonant in CVVC syllables (where the vowel is bimoraic) but not in CVC syllables (where the vowel is monomoraic). Further support for the mora sharing analysis was found in the length of coda consonants, which for the CVVC-tolerating speakers were shorter in CVVC than in CVV syllables. Similar patterns were found by Khattab and Al-Tamimi (2014) for the durations of rhyme segments in syllables closed by geminate consonants in Lebanese Arabic, consistent with the hypothesis that CVVC syllables closed by a geminate are bimoraic.

Broselow et al. (1997) also compared the length of vowels in underlying CVVC+CV and CVC+CV produced by a speaker of Alexandrian Arabic, which shares the non-tolerance of word-internal CVVC with Cairene Arabic. The Alexandrian speaker's productions differed markedly from those of the Levantine speakers: the underlying long vowel in a closed syllable (/kitaab+hum/ > [kitabhum] 'their book') was realized with the same duration as an underlying short vowel ([ʕinabhum] 'their grape'). This neutralization of the contrast between underlying long and short vowels in closed syllables is consistent with the analysis of Egyptian Arabic Closed Syllable Shortening as a phonological process which removes the second vocalic mora. The Egyptian vowel shortening clearly differs from the shortening of long vowels in closed syllables in Levantine Arabic, which results in a vowel whose duration is midway between that of long vowels in open syllables and that of underlying short vowels.

Thus, one account of the distribution of CVVC is that numerous dialects share the property of avoiding trimoraic CVVC syllables, but differ in the strategies employed to implement this avoidance: insertion of a vowel to move the final C to onset position (as in Makkan); removal of one mora from a long vowel (as in dialects of Lower Egypt); and the phonetic shortening of both vowel and consonant reflecting the sharing of a mora between long vowel and coda consonant (as in Levantine). Broselow et al. (1997) provide an analysis of the cross-dialectal differences in an Optimality Theory framework as a difference in the language-specific rankings of universal constraints, where the ranking determines the preferred repair strategy. A fourth logical possibility for avoiding CVVC syllables, deletion of the coda consonant, seems not to be an established process in any variety of Arabic, presumably because of the important role played by consonants in encoding lexical information.

3.3. Vowel Insertion in Consonant Sequences

As noted in section 2.1, many dialects lack syllables with complex onsets and/or complex codas. It is notable that despite the widespread prohibitions on complex margins, concatenation of morphemes frequently gives rise to consonant sequences that cannot be exhaustively syllabified as they stand. As documented below, insertion of a vowel insertion is common in sequences of three or more consonants within words (and in some cases across words) and in sequences of two or more consonants at word or phrase edges--precisely those contexts in which the sequence of consonants cannot be incorporated into syllables without creating complex syllable margins. While the avoidance of word-internal CVCC syllables might be analyzed as a reflection of the ban on trimoraic syllables discussed in section 3.2., many dialects that contain word-internal CVVC syllables lack a similar tolerance of word-internal (and sometimes phrase-internal) CVCC (Broselow et al. 1995: 122; Watson 2007: 348-349). The non-tolerance of CVCC can be viewed either as a prohibition on complex codas that is independent of the trimoraic syllable prohibition, or as a ban on the sharing of a single mora among two segments that are close together on the sonority scale (as argued by Broselow et al. 1995). On this view, a

vowel and consonant may share a mora, but two consonants cannot.

While vowel insertion is widespread, dialects differ in the position of the inserted vowel, a difference which has attracted considerable attention (e.g., Abu-Mansour 1991; Broselow 1980, 1992; Farwaneh 1995; Itô 1989; Kiparsky 2003; Selkirk 1981; Watson 2002, 2007). In four-consonant sequences, the favored position for vowel insertion is between the second and third consonants, as illustrated by the data below from Cairene and Iraqi. Insertion of a vowel in this position is the most efficient means of ensuring that each consonant can be assigned to a neighboring vowel as a simple onset or simple coda (inserted vowels are underlined):

(7) Vowel Insertion in CCCC

a. Cairene Arabic: CC_CC
   ka. tab. ti̲l. ha 'I wrote to her' (/katab + t + l + ha/)
   wrote +1st sg.+to+ her

b. Iraqi Arabic: CC_CC
   ki. tab. ti̲l. ha 'I wrote to her' (/kitab + t + l + ha/)
   wrote +1st sg.+to+her

Dialectal differences in the position of the inserted vowel emerge in a sequence of three consonants, where a vowel appears after the second consonant in Cairene but before the second consonant in Iraqi:

(8) Vowel Insertion in CCC

a. Cairene Arabic: CC_C
   ka. tab. ti̲. lu 'I wrote to him' (/katab + t + l + u/)
   wrote +1st sg.+to+him

b. Iraqi Arabic: C_CC
   ki. ta. bi̲t. la 'I wrote to him' (/kitab + t + l + a)
   wrote +1st sg.+to+him

Vowel insertion may take place not only within words but also on the phrasal domain. For example, in Cairene Arabic a vowel is inserted when a word ending in CC is followed by a consonant-initial word within the phrase, consistent with the prohibition on final CC within phrases. Similarly, vowel insertion takes place in the cross-word context in San'ani Arabic, though only when the word-final consonant sequence is of rising sonority (Watson 2002: 65):

(9) Vowel Insertion in CC#C

a. Cairene Arabic
   bin. ti̲. # ga. mii. la 'a pretty girl'

b. San'ani Arabic
   xid. ra̲. # ba. da. wii 'a bedouin abode'

In both dialects, word-final geminates, like two-consonant sequences, also trigger vowel insertion (e.g. San'ani [ahamma̲#ay] 'the most important thing', Watson 2002: 64), which prevents the degemination that normally takes place in phrase-final position.

In addition to vowel insertion within words and phrases, many dialects exhibit vowel

insertion triggered by two-consonant sequences at constituent edges. Vowel insertion into final two-consonant sequences is quite common; in Iraqi Arabic, for example, a vowel is almost always inserted into a final sequence (as in [bin̲it] 'daughter'; cf. [binti] 'my daughter' with no inserted vowel). In other dialects, vowel insertion is variable, with a greater likelihood of insertion into final sequences that violate the Sonority Sequencing Principle (illustrated by the Lebanese pair [ħilm] 'dream' vs. [ħim̲il] 'load' cited above). Vowel insertion in these dialects seems always to place the vowel between the two final consonants rather than following them. In contrast, when vowel insertion is triggered by consonant sequences in initial position, the inserted vowel is placed before rather than inside the sequence. According to Mitchell (1993:77), vowel insertion in initial position "occurs in stylistically more formal variants of forms which elsewhere contain an initial cluster." As an example of this stylistic variation, Mitchell cites variant forms of 'book' in Jordanian Arabic: the casual pronunciation [ktaab] alongside the careful pronunciation [ʔi̲k. taab] 'book.' Insertion of a vowel before the cluster necessitates the further insertion of glottal stop to provide an onset for the newly created syllable.

Analyses of Arabic dialect vowel insertion patterns fall into two major categories: global approaches, which consider the entire string in syllable construction, and directional approaches, which build syllable structure beginning from one edge of the word or phrase.

In global approaches, the position of the inserted vowel depends on a difference in the preferred syllabic role of an initially unsyllabifiable consonant. In the onset/rhyme approach (Selkirk 1981, Broselow 1992), syllables are constructed around each vowel for a maximum syllable size of CVC. Unassigned consonants trigger insertion of the minimum number of vowels required to accommodate each consonant in a syllable. In the Cairene pattern, a single unsyllabified consonant (indicated below by parentheses) is assigned to onset position, while in the Iraqi/Jordanian pattern the consonant is assigned to coda position:

(10) Vowel Insertion, Onset/Rhyme Approach
  a. Cairene: ka. tab. (t) lu > ka. tab. ti̲. lu 'I wrote to him'
  b. Iraqi: ki. tab. (t) la > ki. tab. i̲t. la > ki. ta. bi̲t. la 'I wrote to him'

In the Iraqi pattern, the requirement that all syllables have consonantal onsets necessitates an additional step, resyllabification of a prevocalic consonant to onset position. When no preceding consonant is available (e.g., in [i̲k.taab], the consonant is assigned to rhyme position and a glottal stop inserted to supply an onset to the inserted vowel ([ʔi̲k.taab] 'book'). These vowel insertion patterns are the basis for the division of dialects into 'onset dialects' vs. 'coda dialects' (Selkirk 1981, Broselow 1992). An extension of the onset/rhyme approach is proposed by Kiparsky (2003), who argues that in dialects with the Cairene pattern (termed CV dialects), each consonant is immediately incorporated into a syllable, with vowel insertion producing the preferred CV structure. In contrast, dialects with the Iraqi pattern (VC dialects) allow an unsyllabified consonant to constitute a semisyllable (a mora directly associated with the word node). At the phrase level, semisyllables are incorporated into a syllables while maintaining their status as moraic. The surfacing of this consonant in coda position follows from the assumption that only coda segments are dominated by moras.

On the directional approach (Itô 1989, Farwaneh 1995), the organization of segments into syllables proceeds locally beginning at either right or left edge, with differences in the position of the inserted vowel reflecting differences in the direction of syllabification. Itô (1989) crucially

assumes an alternative view of mora structure whereby onset and nucleus share a mora. In this system, the initial step is to group segments into moras, with each CV sequence constituting a mora and each remaining consonant is assigned its own mora. The syllabification process then considers each pair of moras as a possible syllable, proceeding from the left edge in Cairene and from the right edge in Iraqi:

(11) Vowel Insertion, Directional Approach

a. Cairene: μ μ μ μ μ
ka ta b t lu > ka. tab. ti. lu (left-to-right)

b. Iraqi: μ μ μ μ μ
ki ta b t la > ki. ta. bi̲t. la (right-to-left)

In Cairene, the second and third moras form a syllable [tab], but the next pair ([t], [lu]) cannot constitute a well-formed Cairene syllable, triggering vowel insertion. The placement of the inserted vowel after [t] is argued to represent the default position for a single consonantal mora. On the other hand, the placement of the vowel before the [t] in Iraqi follows from the right-to-left directionality of syllable construction: following syllabification of the rightmost mora [la], the syllabification scan considers the next pair of moras [l], [t], grouping them into a single syllable by insertion of a vowel between the two consonants. Thus, the position of the inserted vowel is a function of whether the scan encounters one consonantal mora (as in Cairene) or two (as in Iraqi).

Both the global and the directional approaches predict that the position of an inserted vowel in a four-consonant sequences is the same in both dialects, since in such sequences, there will always be two unsyllabifiable consonants under simultaneous consideration. While these approaches are couched in derivational frameworks, both assume that the minimal number of vowels is inserted to accommodate all consonants in the syllable structure (a principle that is made explicit in constraint-based accounts like Optimality Theory, which penalize each deviation from the underlying representation; see Mester and Padgett 1994 for a recasting of the directional approach in a parallel Optimality Theory framework).

With respect to peripheral CC sequences, the global and directional approaches both correctly predict the position of the inserted vowel in final CC in rhyme dialects (as in Iraqi [bini̲t] ‘girl’); where the final unsyllabified consonant is a coda (on the global approach), or is one of two consonantal moras ([n], [t]) considered as a pair in a right-to-left scan (on the directional approach). A potential problem for the onset/rhyme approach is the prediction that an onset dialect (CCC > CC_C) should, if it fails to tolerate final CC, repair this sequence by insertion of a vowel following the two consonants. Makkan appears to be a counterexample, since it has insertion in CC_C but in final C_C sequences of rising sonority, e.g., [katabta̲ha ‘I wrote it (fem.)’ but [ʔibi̲n] ‘son’ (Watson 2007: 347). Conversely, a potential problem for the directional approach is posed by the treatment of initial CC in rhyme dialects (Broselow 1992). Whereas the placement of a vowel before initial _CC (e.g., Mitchell’s (1993: 60) Jordanian careful pronunciation [ʔi̲k.taab] ‘book’) represents the assignment of a single unsyllabified consonant to coda position in the onset/rhyme approach, right-to-left syllabification wrongly predicts that single-consonant mora will be syllabified as default CV, yielding the wrong form *[ki̲.taab]. However, Farwaneh (1995) argues for an alternative directional approach in which a segment “associates with the first available and suitable position in the template it encounters. In

RL mapping, the first available position is the postvocalic position" (Farwaneh 1995: 38). Data from loanword adaptation, which provides contexts for vowel insertion that may not occur in native vocabulary, provide additional puzzles (Broselow 2015). Thus, the question of whether any single approach can present a unified treatment of vowel insertion in all contexts is still not fully resolved.

3.4. Vowel Deletion

The vowel shortening and vowel insertion processes above serve to avoid trimoraic syllables and/or syllables with complex margins, thereby reducing the complexity of individual syllables. Many dialects also have a metrically conditioned process of vowel deletion (syncope) which, though not obviously motivated by restrictions on possible syllables, is constrained by these restrictions, in that a vowel can be deleted only when the output of deletion is syllabifiable.

This pattern is illustrated by Cairene Arabic, where unstressed short [i] is deleted in an open syllable. However, deletion is possible only in the context VC_CV: precisely the context in which the onset to the deleted vowel can be reassigned to the preceding syllable (e.g., /ʃirib+u/ > [ʃir. bu] 'they drank'). Cairene vowel deletion, like vowel insertion, takes the entire phrase as its domain, and an inserted vowel may serve as part of the context for vowel deletion. For example, insertion of a vowel following a word ending in CC triggers vowel deletion in the following word (Broselow 1976, 1992):

(12) Cairene Vowel Insertion and Deletion

a.	bint	'girl'
b.	kibiira	'big'
c.	bin. tik. bii. ra	'a big girl'
	(bint kibiira > bint̲i kibiira > binti k_biira)	

Cases like these illustrate the less than perfect alignment of word edges and syllable edges in this dialect discussed above; this pattern of 'backwards' resyllabification is unusual and leads to problems among English speaking learners of Arabic, who make frequent errors in segmenting Arabic phrases into words (Broselow 1984).

Some dialects allow vowel deletion not only in VC_CV but also in postpausal C_CV. In San'ani Arabic, for example, any short unstressed vowel may optionally delete in this context, giving rise to variant forms such as [katabt] ~ [ktabt] 'I wrote' (Watson 2003: 73). Consistent with the restriction that the output of vowel deletion must be syllabifiable according to the restrictions of the dialect, vowel deletion in this context is permitted only in dialects that allow phrase-initial CC.

4. FUTURE DIRECTIONS

Much research in Arabic syllable structure has been typological in nature, with the goal of finding implicational relationships among various surface properties of the dialects and formal properties to explain these relationships. The initial division of dialects into onset vs. rhyme dialects (or left-to-right vs. right-to-left dialects) on the basis of differences in the position of inserted vowels (Selkirk 1981, Broselow 1992, Itô 1989, Farwaneh 1995) has been extended to encompass a number of logically independent structural properties.

Perhaps the most ambitious attempt to construct a typology related to differences in

allowable syllabic structure is that of Kiparsky 2003, which divides the dialects into three groups: CV (onset) dialects (associated with Cairo and vicinity); VC (rhyme) dialects (associated with the Levant, Turkey, Eastern Libya, and some Bedouin groups); and C dialects (associated with North Africa). This division is based on a number of properties: the tolerance of consonant sequences within words and at word and/or phrase edges; the position of vowels in medial and/or peripheral clusters; the tolerance of phrase-initial geminates; the glottalization or desonorization of final consonants in CC; the tolerance of non-final CVVC vs. the occurrence of closed syllable shortening; the metathesis of CiCC to CCiC (/ji+ktib+u/ > [jikitbu] 'they write'); the deletion of a high vowel following a geminate (/ji+kallim+u/ > [jikalmu] 'they speak'); and the invisibility of inserted vowels for stress. Kiparsky finds implicational relationships among various of these properties (e.g., metathesis is confined to VC dialects, closed syllable shortening to CV dialects) and derives the constellation of patterns from a single parameter, the tolerance of semisyllables. In Kiparsky's system, CV dialects do not tolerate semisyllables; VC dialects tolerate semisyllables at the word level but not at the phrasal level, where the semisyllable is incorporated into a syllable; and C dialects tolerate semisyllables at all levels.

Watson (2007), considering an expanded database of dialects, presents evidence that not all dialects conform to the implicational relationships proposed by Kiparsky's semisyllable typology. She points out that some of the dialects classified by Kiparsky as CV dialects exhibit mixed behavior: San'ani Arabic, for example, patterns with CV dialects in the position of epenthetic vowels in clusters of three consonants (/bint+naa/ > [bint<u>a</u>naa] 'our daughter') but patterns with VC dialects in tolerating word-internal CVVC syllables (Watson 2007: 348). To accommodate such dialects, she argues convincingly for the addition of a fourth category to Kiparsky's three groups (CV, VC, and C dialects): Cv dialects, which (like CV dialects) prohibit semisyllables at both the lexical and postlexical levels, accounting for the assignment of an unsyllabifiable consonant to syllable onset, but which (like VC dialects) permit a consonant to share a mora with a preceding long vowel, accounting for the tolerance of CVVC syllables. She also notes the possible need for additional subdivisions, pointing out that Libyan Tripoli Arabic, for example, shares some characteristics of C and VC dialects. These typological proposals offer an invaluable framework for investigating the limits of cross-dialectal variation and the relationships between surface facts and more abstract analyses of the mental representation of linguistic structure.

The study of Arabic has been enormously influential in the cross-linguistic study of syllable structure and prosody. But while the past decades have seen an incremental increase in our knowledge and understanding of the many varieties of Arabic, the beauty and complexity of the phonological systems of the Arabic varieties and the wealth of cross-dialectal variation that Arabic offers are bound to occupy researchers for decades to come.

RELATED TOPICS
Stress, accent, metrical structure, syllable weight
Phonotactics
Geminates
Vowel insertion, epenthesis

REFERENCES

Abu-Mansour, M., 1991. Epenthesis in Makkan Arabic: Unsyllabified consonants vs. degenerate syllables. *In*: B. Comrie and M. Eid, eds. *Perspectives on Arabic Linguistics II.* Amsterdam: John Benjamins, 167–191.

Abu-Mansour, M., 1992. Closed-syllable shortening and morphological levels. *In*: E. Broselow, M. Eid, and J. McCarthy, eds. *Perspectives on Arabic Linguistics IV*. Amsterdam and Philadelphia: John Benjamins, 47–75.

Al-Ani, S. and May, D. 1973. The phonological structure of the syllable. *In:* S. Al-Ani, ed. *Readings in Arabic Linguistics*. Bloomington: Indiana University Linguistics Club, 113-125.

Angoujard, J.-P., 1990. Metrical structure of Arabic. *Publications in Language Sciences 35*. Dordrecht/Providence: Foris Publications.

Birkeland, H., 1952. *Growth and structure of the Egyptian Arabic dialect*. Oslo: I kommijon hos Jacob Dybwad.

Blevins, J., 1995. The syllable in phonological theory. *In*: J. Goldsmith, ed. *The Handbook of Phonological Theory*. Oxford: Blackwell, 204–266.

Brame, M., 1971. Stress in Arabic and generative phonology. *Foundations of Language*, 7, 556–591.

Broselow, E., 1976. *The phonology of Egyptian Arabic*. Thesis (PhD). University of Massachusetts at Amherst.

Broselow, E., 1980. Syllable structure in two Arabic dialects. *Studies in the Linguistic Sciences*, 10, 13–24.

Broselow, E., 1984. An investigation of transfer in second language phonology. *International Review of Applied Linguistics* 22, 253–269.

Broselow, E., 1992. Parametric variation in Arabic dialect phonology. *In*: E. Broselow, M. Eid, and J. McCarthy, eds. *Perspectives on Arabic Linguistics IV*. Amsterdam and Philadelphia: John Benjamins, 7–46.

Broselow, E. 2015. The typology of position-quality interactions in loanword vowel insertion. *In*: Y. Hsiao and W. Lee, eds. *Capturing Phonological Shades: Papers in Theoretical Phonology*, 292-319.

Broselow, E., Chen, S., and Huffman, M., 1997. Syllable weight: convergence of phonology and phonetics. *Phonology*, 14 (1), 47–82.

Broselow, E., Chen, S., and Huffman, M., and Hsieh, R., 1995. The timing structure of CVVC syllables. *In*: M. Eid, ed. *Perspectives on Arabic Linguistics VIII*. Amsterdam and Philadelphia: John Benjamins, 119–138.

Clements, N., 1990. The role of the sonority cycle in core syllabification. *In*: J. Kingston and M. Beckman, eds. *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech.* Cambridge: Cambridge University Press, 283–333.

Cowell, M., 1964. *A Reference Grammar of Syrian Arabic*. Washington, D.C.: Georgetown University Press.

Dell, F. and Elmedlaoui, M., 2002. *Syllables in Tashlhiyt Berber and in Moroccan Arabic*. Dordrecht: Kluwer.

Erwin, W., 1963. *A Short Reference Grammar of Iraqi Arabic*. Washington: Georgetown University Press.

Farwaneh, S., 1995. *Directionality Effects in Arabic Dialect Syllable Structure*. Thesis (PhD). University of Utah.
Farwaneh, S., 2009. Toward a typology of Arabic dialects: The role of final consonantality. *Journal of Arabic and Islamic Studies*, 9, 82–109.
Fischer, W. and Jastrow, O., 1980. Handbuch der Arabischen Dialekte. Wiesbaden: Harrassowitz.
Harrell, R., 1962. Consonant, vowel, and syllable in Moroccan Arabic. *In*: A. Sovijarvi, ed. *Proceedings of the Fourth International Congress of Phonetic Sciences,* The Hague: Mouton, 643–647.
Hayes, B., 1995. *Metrical stress theory: principles and case studies*. Chicago: University of Chicago Press.
Itô, J., 1989. A prosodic theory of epenthesis. *Natural Language and Linguistic Theory*, 7 (2), 217–260.
Kenstowicz, M., 1986. Notes on syllable structure in three Arabic dialects. *Revue Québéecoise de Linguistique*, 10 (1), 101–128.
Kenstowicz, M. and Abdul-Karim, K., 1980. Cyclic stress in Levantine Arabic. *Studies in the Linguistic Sciences*, 10, 55–76.
Khattab, G. and Al-Tamimi, J., 2014. Geminate timing in Lebanese Arabic: the relationship between phonetic timing and phonological structure. *Laboratory Phonology*, 5 (2), 231–269.
Kiparsky, P., 2003. Syllables and moras in Arabic. *In*: C. Fery and R. van de Vijver, eds. *The Syllable in Optimality Theory*. Cambridge: Cambridge University Press, 147–182.
Lowenstamm, J., 1996. CV as the only syllable type. *In*: J. Durand and B. Laks, eds. *Current Trends in Phonology: Models and Methods*. Salford: ESRI, 419–441.
McCarthy, J., 1979. On stress and syllabification. *Linguistic Inquiry*, 10 (4), 443–465.
Mester, A. and J. Padgett., 1994. Directional syllabification in Generalized Alignment. *In*: J. Merchant, J. Padgett, and R. Walker, eds. *Phonology at Santa Cruz 3*, 79-85.
Mitchell, T. F., 1956. *An introduction to Egyptian Colloquial Arabic.* Oxford: Clarendon Press.
Mitchell, T. F., 1993. *Pronouncing Arabic 2*. Oxford: Clarendon Press.
Nespor, M. and Vogel, I., 1986. Prosodic phonology. Dordrecht: Foris.
Selkirk, E., 1981. Epenthesis and degenerate syllable in Cairene Arabic. *In*: H. Borer and J. Aoun, eds. *Theoretical Issues in the Grammar of Semitic Languages. MIT Working Papers in Linguistics 3*. Cambridge, MA: Department of Linguistics and Philosophy, MIT, 209–232.
Selkirk, E., 1984. On the major class features and syllable theory. *In*: M. Aronoff and R. Oehrle, eds. *Language Sound Structure: Studies in Phonology*. Cambridge: MIT Press, 107–136.
Shaw, J., Gafos, A., Hoole, P., and Zeroual, C., 2009. Syllabification in Moroccan Arabic: evidence from patterns of temporal stability in articulation. *Phonology*, 26, 187–215.
Shaw, J., Gafos, A., Hoole, P., and Zeroual, C., 2011. Dynamic invariance in the phonetic expression of syllable structure: a case study of Moroccan Arabic consonant clusters. *Phonology*, 28 (3), 455–490.
Versteegh, K. 1977. *The Arabic Language*. New York: Columbia University Press.
Watson, J., 2002. *The Phonology and Morphology of Arabic*. Oxford: Oxford University Press.
Watson, J., 2007. Syllabification patterns in Arabic dialects: long segments and mora sharing. *Phonology*, 24 (2), 335–356.

FURTHER READING

Mitchell, T. F., 1993. *Pronouncing Arabic 2*. Oxford: Clarendon Press.
A descriptive overview of the sound systems of a number of different varieties of Arabic, comparing phoneme systems, phonotactics, word stress, and intonation.

Watson, J., 2002. *The Phonology and Morphology of Arabic*. Oxford: Oxford University Press.
A discussion of Arabic phonology within a generative framework, focusing mainly on the comparison of San'ani and Cairene.

Kiparsky, P., 2003. Syllables and moras in Arabic. *In*: C. Fery and R. van de Vijver, eds. *The Syllable in Optimality Theory*. Cambridge: Cambridge University Press, 147–182.
A proposal for a typology of the syllable structure-related differences among various Arabic dialects.

BIOGRAPHICAL NOTE

Ellen Broselow is Professor of Linguistics and Associate Dean at Stony Brook University (State University of New York) and a Fellow of the Linguistic Society of America. She is co-editor of two volumes of *Perspectives on Arabic Linguistics* and a member of the editorial boards of *Linguistic Inquiry*, *Natural Language and Linguistic Theory*, and *Phonology*.