Abstract

This paper provides a comparative study of epenthetic vowels in varieties of Arabic. It states that these varieties can be grouped into four categories on the basis of this criterion: a) those with i-epenthesis, b) those with a-epenthesis, c) those with schwa epenthesis, and d) those which use a combination of more than one epenthetic vowel, depending on the phonological context. With a close analysis of Classical Arabic, the paper shows that, while /i/ is regularly inserted to satisfy phonotactic constraints, there are vestiges of epenthetic /a/ in the language, which probably used to be a productive process in an earlier stage. On the basis of this suggestion, the paper concludes that this process could have been regular in the ancestor of the extant varieties of Arabic.

Key words: vowel epenthesis; Classical Arabic; Arabic dialects; phonotactic constraints
1 Introduction

In Semitic in general, while there are a lot of similarities between languages in the constraints governing syllable structure, both word-internally and across words, there is some variation in the nature of vowel epenthesis used to repair illegal consonant clusters that result from morphological processes (cf. Moscati et al., 1980). Within Arabic, too, modern dialects studied so far exhibit this type of variation even among those varieties which still maintain more or less the same constraints on syllable structure as Classical Arabic (CA), like some Eastern varieties. These dialects vary as to whether their typical epenthetic vowel is /i/ or /a/, while others have assimilatory processes that require the epenthesis of more than one type of vowel. In yet another group of dialects, as is the case in North African varieties, the schwa is inserted to break illicit clusters.

Although vowel epenthesis in Arabic has been studied amply, both in separate varieties and from a comparative perspective, studies have usually focused on syllable types and the positions where the epenthetic vowel is inserted (cf. Broselow, 1992; Kiparsky, 2003; Watson, 2007, Farwaneh, 2009, Haselwood and Watson, 2013). To my knowledge, however, no study so far has addressed the diachronic relation between the epenthetic vowels used in these varieties. Therefore, the present paper will provide a comparative study in view of reconstructing vowel epenthesis in Old Arabic (OA), the ancestor language from which all known varieties of Arabic descend\(^1\).

This paper will be constituted of three main sections. In Section 2, the different types of epenthetic vowels in Arabic dialects will be described on the basis of available data. Section 3 will focus on data from CA in search for vestiges of archaic epenthetic vowels. Section 4 will discuss the historical development of modern Arabic dialects and CA in relation to OA. A conclusion will finally summarize the main ideas and findings of the study.

2 Vowel epenthesis in Arabic dialects

As pointed out above, Arabic dialects vary as to which vowel is epenthesized to break illegal consonant clusters. Some of these dialects use /i/, others use /a/, a third group uses a schwa, while a fourth group uses /i/, /a/ or /u/, depending on the phonological environment. Because of some complex assimilatory processes, the fourth group can be very heterogeneous.

Among the four strategies, i-epenthesis is the most well-known and probably the most widespread. A good example of a dialect that uses this repair strategy is Egyptian Arabic. In this variety, as in CA, complex onsets and complex codas are banned. Therefore, a /i/ vowel is inserted to break consonant clusters whenever they rise. The following examples illustrate this fact (cf. Watson, 2002, p. 64)\(^2\):

\[
\begin{align*}
\text{(1) } /\text{ult } + \text{lak/} & \text{?ult[i]lak ‘I told you m.s’} \\
& /\text{kunt } + \text{hina/} \text{kunt[i]hina ‘I/you m.s was/were here’}
\end{align*}
\]

In these examples, the first word ends in a CC-cluster and the second has an initial C. The syllabification of the final consonant of the first word either as part of the preceding coda or the following onset will result in unwanted structures. Therefore, the insertion of /i/ creates a new syllable and, thus, avoids the rise of a complex coda or a complex onset. The epenthesis of this vowel is found

\(^1\)It should be pointed out that OA is a theoretical construct only postulated to account for the similarities between Arabic varieties, including CA. The comparative method in historical linguistics reconstructs previous stages of a language where differences between related languages or dialects are assumed to descend from one and the same ancestor language, generally referred to as a "proto-language". However, no precise information can result from the comparative method itself about the speakers of the reconstructed variety or the factors that lead to its change.

\(^2\)Here, as in the other cases listed below, the examples provided serve only for illustration. For more data, the reader is referred to the works cited.
similarly in Levantine, Iraqi, Najdi and a number of Jordanian dialects, although with some variation in the domains in which the epenthesis occurs. For instance, while Egyptian Arabic allows complex codas in word final positions, Iraqi Arabic does not and, consequently, requires the insertion of /i/ to break word final clusters. (Compare Egyptian ‘bint’ with Iraqi ‘binit’ “girl”). However, we need not be concerned with these details since our primary interest is the nature of the epenthetic vowel per se rather than the domains of epenthesis.

In comparison, some other dialects require the insertion of the low vowel /a/ in response to similar syllable constraints. Among these, the available literature cites Makkan (Abu-Mansour, 1991), Sanʿani (Watson, 2002) and Sudanese (Ali, 2014; Hamid, 1984). In Makkan Arabic, for example, Abu-Mansour (1991) points out that, while CVCC and CVVC can occur in word-final positions, the two types of syllables are banned word-medially and, therefore, invite vowel epenthesis as a repair strategy. But unlike the dialects illustrated above, Makkan Arabic inserts /a/, as in the following examples:

(2) /kitaab + ha/ kitaab[a]ha ‘her book’
/katabt + ha/ katabt[a]ha ‘I wrote it (F)’

As in the previous cases, complex onsets and complex codas are avoided by the insertion of this vowel. Similarly, super-heavy CVVC syllables are avoided by parsing the final C of such syllables as the onset of the epenthetic vowel. Abu-Mansour (1991) notes that the same vowel is used also as a prosthetic vowel, as in the imperative form ‘ʔaktub’ (Write!) where in other varieties, including CA, we generally find /i/, which harmonizes with the stem vowel (viz. ʔilʕab (Play!) but ʔuktub (Write!)).

In a third group of dialects, while one vowel may be unconditioned, the others can also be epenthized in specific environments. This is the case in many rural Jordanian dialects, for instance, where /i/ is generally the most preferred epenthetic vowel but where /a/ and /u/ can also be inserted under specific phonological conditions. In Bani Hassan Arabic (Irshied, 1984), one of these dialects, /i/ is inserted to break final consonant clusters only when these are not in accordance with the Sonorant Hierarchy, as a comparison between the first and the second pairs of examples shows:

(3) ħilm “dream” dafin “burial”
jamʕ “collecting” himil “load”

In the right-hand examples, the clusters [fn] and [ml] violate the Sonorant Hierarchy in that the second phoneme is more sonorous than the first. Irshied (1984, pp. 54–67) argues that this is the reason why /i/ is epenthized between the two consonants, unlike the left-hand cases in which the clusters do not require epenthesis because the Hierarchy is observed. Other dialects exhibiting a similar behavior include Wadi Ramm Arabic (Al-Mashaqba, 2015), Maʕani Arabic (Rakhieh, 2009) and other Jordanian dialects mentioned by Al-Sughayer (1990).

But among many of these dialects, /a/ and /u/ can also be inserted for the same phonotactic reasons. In Bedouin varieties in particular (Blanc, 1970), gutturals cannot appear in a coda position and, when they do, they are usually parsed as onsets of an epenthetic vowel. In these environments, the epenthetic vowel is typically /a/, as in the following:

(4) gahawa “coffee” naʕajah “ewe”
ʔaʕama “blind” sahal “plain”

This phenomenon is known as the gahawah syndrome (de Jong 2007). When the stem vowel is /u/, some varieties require that the epenthetic vowel be harmonized with it, as in these examples cited by Al-Mashaqba (2015):
But epenthetic /u/ is not always the result of vowel harmony. Al-Sughayer (1990) points out that some Jordanian varieties do insert a back vowel in CaCC in the environment of a labial or a pharyngeal sound. The following items illustrate this fact:

(6) /gaṣr/ [gaṣur] “palace”
    /ḥarf/ [ḥaruf] “letter”

Thus, this group of dialects uses the three short vowels in contexts requiring vowel epenthesis. It seems, however, that /i/ enjoys a wider distribution, at least as far as types of phonological contexts are concerned. As to frequency, it will depend on the frequency of the constraining environments. In other words, the insertion of /a/ and /u/ will depend on how frequent gutturals, labials and pharyngeals are in the lexicon.

A final remark concerns the fourth group dubbed by Kiparsky (2003) as the C-dialects. Although these dialects tolerate CC-clusters, they do require the insertion of a vowel to break longer strings. In many of these dialects, the epenthetic vowel is a schwa. For example, in most North African varieties, where CCC clusters are banned, a schwa is usually inserted either before the last C or after the first C. The following examples are from Moroccan Arabic (cf. Boudlal, 2001; Dell and Elmedlaoui, 2002):

(7) xḍm “he worked”
    xḍma “work.f”

In the first item, the schwa is inserted between the second and the third Cs, but when the third C serves as onset to the following feminine marker “a”, the epenthetic vowel occurs between the first and the second Cs. The status of this vowel within Arabic dialectology is not clear, however. Since these varieties have been in close contact with non-Arabic languages such as Berber, it is not obvious whether the schwa, and indeed the whole syllable structure, is inherited from early Arabic dialects or was borrowed from the contact languages (cf. Bensoukas and Boudlal, 2012a,b; Dell and Elmedlaoui, 2002). For this reason, this group will be discarded, though the possibility that the schwa served also as an epenthetic vowel in old Arabic dialects remains to be explored in future studies. This vowel is attested also in some Middle Eastern dialects such as the Syrian (cf. Cowell, 1964) and the Palestinian (cf. Palva, 1965) varieties. It is also attested in other Semitic languages such as those spoken in Eastern Africa (cf. Moscati et al., 1980; Ullendorff, 1955).

The issue that should be tackled after this brief survey of vowel epenthesis in modern Arabic dialects relates to the diachronic development of the process. In particular, the following questions need to be addressed: a) How can the different epenthetic vowels be traced back to an original process in Old Arabic? b) How do modern Arabic dialects relate to CA as far as epenthesis is concerned? c) Can vowel epenthesis provide a reliable criterion for a plausible classification of Arabic varieties, including CA? But before an answer to these questions is attempted, a study of epenthesis in CA will be conducted in the following section with the objective, first, to compare this variety with modern dialects and, second, to search for traces of archaic epenthesis strategies which may serve as a basis for the diachronic reconstruction.

3 Vowel epenthesis in CA

Given that CA is the oldest variety of Arabic of which a thorough description has been handed down to us by the Arabic grammatical tradition, it could serve as a good example of an intermediate stage
between Old Arabic and the modern dialects. This does not necessarily mean that each and every modern dialect must be derived directly from CA; it is very likely that many of them descended from other varieties of Arabic which were spoken in the peninsula and other parts of the Middle East but which differed from CA in significant ways (cf. Owens, 2015). But given its archaic nature, CA is likely to have preserved some traces of old epenthesis strategies if these can be uncovered from deeper layers of the language. Indeed, the main objective of this section is to provide some evidence that /a/ used to be a regular epenthetic vowel in CA before it was replaced by /i/ for unknown reasons, but a brief discussion of /i/ epenthesis will be provided first.

3.1 i-Epenthesis

As was already pointed out, CA has a limited inventory of syllable types, namely, the light CV and the heavy CVC and CVV (Holes, 1994). As in many dialects, CA also bans consonant clusters either in onsets or in codas, just like other cases of super-heavy syllables. In pausal positions, however, a super-heavy syllable can be exceptionally tolerated. Often, the short vowels marking case and mood help avoid such syllables, but when they occur as a result of morphological processes or in connected speech where case and mood markers do not appear, /i/ is inserted as a repair strategy. The following examples illustrate this fact:

(8) /qatal/ (Form I) “kill” /qtatal/ [ʔiqtatal] (Form VIII) “kill each other”
    /ʔakalat + lbint/ ʔakalat[i]lbint “the girl ate”

In the first example, Form VIII of the verb inserts the infix /-t-/ after the first root consonant, which results in a complex onset. The insertion of [i] parses the first consonant as a coda and, thus, avoids the illegal CC-cluster word-initially. (The glottal stop is, in turn, inserted to avoid an onsetless syllable, also banned in the language). Similarly, the epenthesis of the same vowel in the second example parses the final consonant of the first word as an onset and the initial one in the second as a coda in order to avoid either a complex coda or a complex onset in connected speech.

Unlike underlying /i/, epenthetic /i/ is subject to vowel harmony. More precisely, this vowel can be realized as [u] if the following root vowel is of the same quality, as a comparison of the following examples illustrates:

(9) /rmi/ [irmi] “Throw!”
    /ʕab/ [iʕab] “Play!”
    /dxul/ [udxul] “Come in!”

While i-epenthesis operates in the first two cases, whose root vowel is /i/ or /a/, it would yield an incorrect output if applied to the third example, whose root vowel is /u/. Therefore, /u/ is inserted instead to satisfy a requirement on the harmony between the root and the epenthetic vowels. This fact suggests clearly that i-epenthesis is the default strategy.

The alternation of i/u-epenthesis occurs also in coda positions, though only variably. Sibawayh (1988, 173–4, Vol. 4) reports that /i/ or /u/ can be epenthesized in some dialects of CA to break word-final CC-clusters. Examples of this include:

(10) /haaðaa + Bakr/ [haaðaabakur] “This is Bakr”
    /haaðaa + ʕidl/ [haaðaabʕidil] “This is an equivalent”

/u/ is inserted in the first example and /i/ in the second. This type of epenthesis, however, seems to be conditioned both by grammatical and phonological factors. More specifically, the /u/ in the first case seems to be a nominative marker that has shifted its position as an instance of metathesis, apparently
to satisfy the ban on consonant clusters. The proof is that “Bakir” is also attested, always according to Sibawayh, in genitive contexts; e.g. min Bakir “from Bakr”. By opposition, the force of vowel harmony overrides that of the grammatical factor in the second example and /i/ is inserted where we would expect /u/ if the same reasoning is followed. In other words, the epenthetic /i/ in this example would be an underlying nominative marker that has assimilated to the preceding front vowel. This phenomenon, however, is not very relevant to the main issue at hand and, accordingly, will not be discussed any further.

3.2 a-Epenthesis

In comparison with i/u, the epenthesis of /a/ occurs only marginally, but the very few cases in which it is attested are quite suggestive. In this section, a set of arguments will be advanced to support the claim that some instances of this vowel are epenthetic and that these are probably vestiges of an archaic productive process.

3.2.1 Definite article

Traditional grammars of CA state that the definite article in this language is /al-/ as in these examples:

(11) bint “girl” al-bint “the girl”
rağul “man” al-rağul “the man”

In the first example, the article is usually realized as [ʔal] in utterance-initial positions. Besides, the /l/ does not assimilate to the following labial consonant, just like all non-coronal consonants. In the second example, however, the /l/ assimilates to the following /r/, as it does with all coronal phonemes, and is consequently realized as [ʔar].

Synchronically, the vowel of the definite article behaves like an epenthetic element. More specifically, while it shows up in utterance-initial positions, it deletes when preceded by a vowel. What is more is that even when preceded by a consonant within a sentence and an illegal CCC-cluster would rise without vowel epenthesis, /i/ rather than /a/ is epenthesized. The following cases illustrate these facts:

(12) al-wald-u kataba “the boy(Nom.) wrote”
kataba l-walad “the boy wrote”
katabat[ɪ] l-bint “the girl wrote”

Unlike the first example where /a/ occurs, the second example shows that this vowel is not needed when the syllable well-formedness is warranted by other underlying segments. In fact, even when epenthesis is required, as in the third case, it is not /a/ but rather the regular /i/ that is appealed to. If /a/ was underlying, we would expect it to occur at least when preceded by a consonant so that no other vowel would be needed to avoid CCC-clusters. Therefore, the most economical way to account for the formal variation in the definite article is to postulate that its underlying representation is /l-/ and that its vocalic part is only epenthetic to avoid complex onsets. This idea is not a new one; it has been advanced by the majority of traditional grammarians of Arabic as well as by a number of modern researchers (cf. Testen, 1998 for a review). A problem, however, remains as to why this epenthetic /a/ should occur only utterance-initially; an explanation will be suggested later.

The special status of /a/ of the definite article is also reflected in the dialects. While it becomes /i/ in many dialects in which this vowel is used productively in epenthesis, it is still maintained in more conservative varieties. Thus, in Egyptian Arabic, for example, we find /i/ not only in the definite article, but also in relative pronouns, which may be argued to derive diachronically from a combination.
of an article and a demonstrative (Haywood and Nahmad, 1965). In comparison, many varieties of Jordanian Arabic have /a/ in these contexts. The following data illustrate the point:

(13)  
irragil illi biyṣalli  \( \rightarrow \) “the man who is praying” (Mitchell, 1962, p. 102)  
aḏḏyf alli jāʔ al-bāriḥ  \( \rightarrow \) “the guest who came yesterday” (Al-Mashaqba, 2015, p. 214)

The first example is from the Egyptian variety whereas the second is from a Jordanian dialect spoken in the Wadi Ramm valley. As can be noticed, this Bedouin variety retains /a/ where Egyptian has /i/. In fact, it retains this vowel even sentence-medially where CA would insert /i/ (viz. “al-bāriḥ” in the second example). For this reason, Wadi Ramm Arabic can be considered even more conservative than CA itself. In comparison, the Egyptian variety seems to be more innovative in that it regularizes all occurrences of /a/ that are apparently exceptional to the epenthesis rule.

In brief, both synchronic and diachronic analyses of the definite article in Arabic seem to support the claim that its underlying form is /l/ and that /a/ is epenthetic. Given that this vowel changes into /i/ in some varieties but not in others, we can advance that these varieties differ as to the stage of the regularization of vowel epenthesis. This is particularly relevant to those in which i-epenthesis is productive. While dialects like Egyptian have adopted this process completely, others still retain some vestiges of a-epenthesis. In this respect, CA is not necessarily the most conservative among them.

3.2.2 The preposition “min” (from)

A number of prepositions and particles in CA end in a consonant (e.g. min “from”, bal “rather”, man “who”). For reasons that should be obvious by now, these are subject to vowel epenthesis when they are followed by nouns marked for definiteness, given that these strings have an initial CC-cluster.

But unlike most C-final prepositions and articles, “min” requires the epenthesis of /a/ rather than the productive i-epenthesis. This is clear from the following illustrative examples:

(14)  
/min + manzilin/  \( \rightarrow \) “from a house”  
/min + l-manzil/  \( \rightarrow \) “from the house”  
ʕan + l-walad/  \( \rightarrow \) “about the boy”

In the first example, “min” is followed by CV and, consequently, no epenthesis is required, a fact that indicates that this preposition has no underlying final vowel. In the second example, however, a cluster of three consonants results from the combination of the preposition and its complement, a situation which is remedied by the insertion of /a/. In comparison, a similar situation in the third example is remedied by the more productive i-epenthesis.

Thus, although “min” and “ʕan” are fairly similar, they behave differently in relation with vowel epenthesis. In this regard, the occurrence of epenthetic /a/ with “min” is quite exceptional and, for that reason, it did not escape the notice of traditional Arab grammarians. Sibawayh (1988, p. 154), for example, notes in this regard that “some Arabs are reported to say “minillāh” [from God], thus regularizing it by inserting /i/” instead of /a/. In other words, although a-epenthesis with “min” was widespread in CA varieties, there was a tendency in others to use the more productive i-epenthesis.

At this point, the question as to why a-epenthesis should linger longer with the preposition “min” than with other items needs to be addressed. In this connection, we may appeal to the effect of frequency which has been shown to contribute significantly to the entrenchment of forms, leading thus to the maintenance of irregular forms when these are very frequent (Bybee, 1985). Sibawayh (1988) also appeals to frequency and ease of pronunciation as possible factors, though he couches these notions only in impressionistic terms. For him, the low vowel requires less effort in pronunciation than the
high vowel and this explains why /a/ is maintained with highly frequent items such as the preposition “min”. Surprisingly, modern computerized corpora seem to support this explanation. The Quran corpus (corpus.quran.com), for example, puts this preposition at the top of the most frequent items with 3226 tokens. Among the other C-final prepositions and particles, the relative pronoun “man” (who) comes second, with 606 tokens only, and “ʕan” (on) in the third place, with 465 tokens, while the others occur less frequently. In Modern Standard Arabic (MSA) as well, “min” is still highly frequent, outranked only by “fī” (in), the definite article and the coordinator “wa” (and) (Buckwalter and Parkinson, 2011). Obviously, the Quran cannot be considered representative of the language but, together with corpora of MSA, it points to a crucial effect of frequency.

Frequency seems also to stand behind the exceptional behavior of the definite article. In the preceding sub-section, we argued that the low vowel that occurs with this article in utterance-initial positions is also epenthetic but no explanation was attempted there. In the Quran, it occurs 8377 times; in Buckwalter and Parkinson (2011)’s frequency dictionary, it is ranked first with more than five million occurrences in a thirty-million corpus of MSA. Of course, this fact does not by itself explain why the occurrence of /a/ should be limited to utterance-initial positions only while /i/ replaces it in other contexts, as explained above. If VSO was the only word order allowed in the language, we would not expect utterance-initial nouns to be frequent enough to entrench irregular patterns, but CA does allow alternative word orders as well as nominal sentences in which topics occur very frequently in initial positions. Besides, it is very likely that nouns used to occur often independently in speech, a context which may have continued to favor the archaic epenthetic vowel for very long.

3.2.3 The preposition “l” (to)

Another piece of evidence in favor of the existence of an epenthetic /a/ in CA comes from the behavior of the preposition “l” (to). This preposition takes an /i/ vowel when followed by a noun but /a/ when followed by a pronoun, as in these examples:

(15) l + lbint/ libint “to the girl”
l + bintī/ libintī “to my girl”
l + hā/ lahā “to her”

As is clear from these examples, there seems to be no phonological conditioning for the occurrence of either /i/ or /a/. Therefore, other explanations for the alternation between the two vowels must be sought.

In the above examples, the preposition is represented by the single consonant /l/ on the assumption that the following vowel is epenthetic. This assumption, however, is not evident and further analysis is required. Synchronically, since the i~a alternation cannot be accounted for by phonological conditioning, we must suppose that it occurs at the interface between phonology and syntax. In other words, the preposition must have two allomorphs, “li” and “la”, and the rule selecting one or the other must make reference to the syntactic category of the object of the preposition, i.e. whether it is a noun or a pronoun.

However, irrespective of whether this synchronic analysis has any psychological reality, it fails to take into account the alternation between /a/ and /i/ noted elsewhere in the language. If we assume that the preposition does not have an underlying vowel, an epenthetic vowel will be needed whenever the preposition is followed by a C-initial word in order to avoid complex onsets; e.g. *lhā “to her”. To put it differently, the a~i alternation in this preposition bears a lot of resemblance to what is noticed in the definite article and the preposition “min” discussed earlier. Given that Arabic nouns and pronouns

---

3I would like to thank Marijn van Putten for providing me with this information
are typically C-initial, it is very likely that the vowels occurring after “l” were not diachronically underlying and that they were inserted only for phonotactic reasons.

The behavior of this preposition in modern Arabic dialects seems to support the hypothesis that it is constituted underlyingly of /l/ only. In Syrian Arabic, for example, the tendency is to delete the vowel except when illegal consonant clusters would result, in which case a schwa is inserted or the archaic vowel is maintained. This is illustrated by the following examples (Cowell, 1964, pp. 476–84):

(16) lǝl-walad “to the boy”
žībī-lna “bring (f.) … for us”
bžǝb-lak “I will bring … for you (m.)”

In accordance with the phonotactics of the language, a schwa is inserted between the preposition and the definite article in the first example in order to avoid the initial cluster of three consonants. In the second example, however, the epenthetic vowel is not required and the preposition is directly attached to the pronoun affix. As to the third example, /a/ appears between the preposition and the second person pronoun, for otherwise a CCC-cluster would result. This vowel is unlikely to be underlying given that the /-ak/ form of the pronoun occurs only when /k/ cannot be parsed as a syllable coda, as a comparison of “ʔǝxt-ak” (your sister) and “ʔaxūk” (your brother) indicates. While in the second case the /k/ is immediately preceded by a vowel and, consequently, can serve as a coda, in the first it is preceded by two consonants, which leaves no room for it in the cluster. In this regard, the /a/ functions very much like its CA equivalent with pronouns considered above except that, in Syrian Arabic, it behaves more clearly as an epenthetic vowel. Why should /a/ be inserted in this context, instead of the more regular schwa, is an issue that seems to have only a diachronic explanation.

In more conservative varieties such as Najdi Arabic, internal variation suggests that the preposition “l” is in an intermediary stage. In this variety, while “l” is followed consistently by /i/ before a noun, both /i/ and /a/ can occur before pronouns. Besides, the two vowels can appear before or after the preposition in response to syllabification requirements. The following examples are taken from the transcribed texts in Ingham (1994):

(17) li-l-malāhi “to the Fun Fair” p. 150
li-na “to us” p. 151 li-h “to him” p. 153 la-ha “to her” p. 153
gil-t-il-ha “I told her” p. 152 gul-t-al-ha “you told her” p. 152
twajjih-ū-l-hum “you direct them” p. 157 ni-ntibih-li-hum “we pay attention to them” p. 158

In the first example, /i/ occurs between the preposition and the definite article, as is the case in most Arabic varieties, including CA. In the second line, however, both /i/ and /a/ show up with pronouns. There seems to be no decisive factors behind the choice of one or the other, but the neighboring vowels may have some effect. This is more obvious in the third line, where /i/ is preferred when the preceding vowel is of the same quality while /a/ occurs with a preceding round vowel. Unlike in the previous examples, /i/ and /a/ occur before the preposition here. The two examples in the last line indicate that the position of the two vowels depend on the syllabification of the preceding segments. More precisely, when /l/ is preceded by a vowel, it serves as a coda and no vowel is needed following it (e.g. twajjih-ū-l-hum). If it is preceded by VC and followed by CV, however, a vowel will necessarily follow it to avoid a complex onset in the subsequent syllable (e.g. ni-ntibih-li-hum). Likewise, when it is preceded by VCC and followed by CV, a vowel will appear after the second segment of the VCC cluster and the preposition will serve as a coda of that syllable (e.g. gul-t-al-ha). This shifting of positions strongly supports the hypothesis that these vowels are epenthetic and that the reason for their insertion is mainly to satisfy syllable well-formedness conditions.

In brief, the Arabic data indicate that the /a/ that occurs after the preposition “l” (to) is most likely
epenthetic. In some varieties such as CA, this vowel is limited to pre-pronominal positions and survives merely as a vestige of a previous diachronic stage. But in others like Najdi Arabic just considered, there is also synchronic evidence pointing to its epenthetic character.

3.2.4 ʔinna, ʔanna, kaʔanna and lākinna

Vestiges of epenthetic /a/ have apparently survived also in a small set of particles that have both a “weak” and a “strong” form. Among these are “ʔinna” (verily/indeed), “ʔanna” (that), “kaʔanna” (as if), and “lākinna” (but). These strong forms correspond to the weak forms “ʔin” (indeed), “ʔan” (to), “kaʔan” (as if), and “lākin” (but), respectively. The weak and the strong forms may have different semantic nuances, but they are still similar enough for us to assume that they are mere variants of the same lexemes. Indeed, this is how traditional grammars of CA treat them.

For illustration, we will consider first “kaʔan(na)”. Both the strong and the weak forms of this particle can be rendered in English as “as if” or “as though”. The two can occur interchangeably, but there are contexts in which only one would be appropriate or even correct. In the following Quranic verse (Q31:7), for example, both forms are used:

(18) wallā mustakbir-an kaʔan lam yasmaʕ-hā
Turn.3ms arrogant-Acc as Neg heard.3ms-3fs
kaʔanna fī ʔuðunay-hi waqr-an
as in ears.Gen-3ms deafness-Acc

(He turns away arrogantly as if he had not heard them, as if there was in his ears deafness)

In the first instance, the weak form is followed by a verbal clause introduced by the negative marker “lam”, whereas in the second, the strong form is followed by a nominal clause introduced by a prepositional phrase. In the nominal clause, the topic occurs after the prepositional phrase, but the accusative case indicates that it is within the domain of “kaʔanna”, given that this particle assigns the accusative to its topic and the nominative to the comment. In comparison, the weak “kaʔan” can be followed by a nominal clause only in archaic use, and when it does, traditional grammarians are in disagreement whether it can assign case, like the strong form, or not. In less archaic texts, as in MSA, the weak form occurs exclusively with verbal sentences and the strong form with nominal sentences. In the Quran, for example, the strong form occurs 29 times, 28 of which are followed by a pronoun or the particle “mā”, which apparently has no semantic content. It seems that the main function of this particle is to enable the strong form to be followed by a verbal sentence (e.g. kaʔannamā qatala l-nās-a ġamīʕan “As if he had killed all people”). This is further evidence that the two forms are mere variants.

A similar point can be made in relation with “lākin(na)”. Both the strong and the weak forms of this particle express contrast with the preceding clause and, therefore, can be rendered in English as “but”. As is the case with “kaʔan(na), the strong form occurs typically with nouns and pronouns to which it assigns the accusative case while the weak form occurs typically with verbs. The weak form can occur with nouns, though rarely, and when it does, it has no effect on the case assigned to those nouns. Ryding (2005, p. 427) notes that “[i]n written text, it is almost impossible to tell the difference between these two particles, except that “lākin” may be followed by a verb.” In other words, the difference between the two forms is more one of syntactic distribution than of semantic content. In the Quran, for example, the weak form occurs 65 times of which 54 are part of verbal sentences. Of the remaining 11 instances, 6 are followed by nouns with the definite article while in the other 5 cases the nouns begin with a CV-string. In comparison, the strong form occurs in the same text 56 times of which 22 are followed by the definite article and only 4 by a pronoun. In all the 56 instances, however, the strong form is followed by a noun phrase functioning either as a topic of a nominal sentence or a preposed subject.
In spite of the complementary distribution between the two forms, there is some sense in which “lākin” can be considered as more basic. In terms of syntactic distribution, “lākinna” is more restricted than “lākin”, at least in principle, if not in practice. More specifically, while “lākin” can occur with both verbs and nouns, “lākinna” can co-occur only with nouns. In fact, the strong form can be replaced by the weak form in practically all contexts with the only condition that the following noun must take the nominative instead of the accusative case. Consider, as examples, the following two Quranic verses:

(19) lākin llāh-u yašhad-u (But Allah witnesses that …) (Q4:166)
walākin lā ta-ʕlamūn (But you do not know) (Q7:38)

In both examples, the weak form is used. When it is followed by a definite noun, as in the first example, an epenthetic /i/ is inserted to avoid a CCC-cluster that would result when “lākin” and the following definite noun are combined. If the weak form is replaced by the strong form in such cases, the resulting sentences are synonymous; (compare: lākinna llāh-a yašhad-u). But while it is possible to replace the weak form by the strong form in the first verse, it is not possible in the second. This shows that all occurrences of the strong form can be replaced by the weak one, but not vice versa.

The two forms of “ʔan(na)” also exhibit complementary distribution, with the strong form occurring before nouns and the weak form before verbs. Both of them function as subordinators. Whether one or the other will occur depends largely on the matrix verb, as the following examples illustrate:

(20) ʔarāda ʔan yaðhab-a “He wants to leave”
iʕtaqada ʔanna l-bint-a ðāhibah “He thought the girl was leaving”

In the first example, the subjects of the main and the subordinate clauses are co-referential. If a different interpretation is intended, an object noun or pronoun must be inserted after the main verb “ʔarād” (want). In comparison, the object noun or pronoun must occur after the subordinator “ʔanna” in the second example, even when the object is co-referential with the subject of the main verb “iʕtaqad” (think). Different verbs require different positions for their objects. (For more on this issue, see Person, 1999). What is crucial for us is that the strong form, as in the cases considered previously, is typically followed by a noun while the weak forms can only be followed by a verb. This fact implies that the strong form is likely to occur more frequently with nouns and pronouns. Indeed, of the 362 instances of “ʔanna” that occur in the Quran, only 31 are followed immediately by a prepositional phrase and in many of these cases, the topic is postposed only for stylistic reasons. As to the 578 instances of “ʔan” that occur in the same text, they are all followed by a verb or a marker of mood/modality, negation or similar categories. Therefore, the “–na” of the strong form seems to be dictated by constraints related to the noun/pronoun context, though the nature of these constraints remains to be determined.

The case of “ʔin(na)” as well does not seem to be very different. The strong form can be used either as a modal particle to intensify the truth of a proposition or as a subordinator in reported speech (cf. Ryding, 2005). As to the weak form, it is often used as a conditional conjunction but also, though only in archaic usage, as a modal particle or a negative marker. Whether the conditional conjunction is historically related to the particle under consideration is not directly relevant to the issue, though it would provide further support to our argument if the two can be shown to descend from the same form. As to the other two functions, grammarians seem to be at a loss; it was probably not clear to them how to interpret examples such as the following:

(21) ʔin qatalta la-musliman
ʔin killed.2ms la-Muslim-Acc
(You killed a Muslim/ You killed none but a Muslim)

The two English translations reflect the grammarians’ uncertainty about the function of “ʔin”. Some view it as a modal expressing certainty, though to a lesser degree than that expressed by the strong
form; this use requires the presence of some particles on the object such as “la”, as in this example. Others, however, consider it as equivalent to negation particles such as “mā” in which case “la” would be a form of “ʔillā” expressing exception. Under this interpretation, the above example would be equivalent to “mā qatalta ʔillā musliman” (you killed none but a Muslim (for a review, see Ibn Yaʕīš, 2001, Vol.4). This controversy indicates that, when the grammatical tradition started, this use of the weak form “ʔin” had perhaps already become obsolete. But since there were instances of it in the Quran that could not simply be overlooked, grammarians had to guess its meaning. Judging from the cases they cite, it seems that this weak form occurs both in verbal and in nominal sentences, though its meaning in the two types may be slightly different. For this reason, and given the parallel behavior with the particles discussed above, we can claim that the weak form is basic and that the strong form developed as a consequence of processes resulting from its interaction with the following nouns or pronouns.

To recapitulate, the four particles considered in this sub-section, namely, “ʔinna”, “ʔanna” “kaʔanna” and “lākinna”, have a weak and a strong form. Since the relation between the two forms cannot be overlooked, an explanation is needed for the formal difference between them. In particular, we need to find out whether the final “-na” is added to the strong form or deleted from the weak one. Given that in all the four cases the weak form is noted to have a wider syntactic distribution than the strong form, we have argued that the strong forms are derived from the weak forms through the epenthesis of /a/ probably as a repair strategy to avoid illicit consonant clusters. It is also possible that this epenthetic /a/ was reanalyzed as part of the particles as their strong forms started to form a special category as a result of their close association with nouns and their case assigning behavior. As to the germination of the preceding nasal, it may have been caused by some constraint on word prosody the details of which need to be investigated in future research.

3.2.5 Other particles

As a follow-up, we may also compare other particles and function words that typically precede verbs. Although these may not have a strong and a weak form like the ones considered earlier, there is a tendency for these items not to have a final /a/, unlike those which co-occur typically with nouns or pronouns. Although this cannot constitute in itself direct evidence for the epenthetic character of the final /a/ of the particles discussed in the preceding sub-section, given that no aspect of synchronic variation can support such an analysis, it is diachronically suggestive, nonetheless.

Examples of particles that co-occur exclusively with verbs include the negative particles “lam” and “lan”, the relative pronoun “man”, the time adverb “ʔið”, and the conditional conjunction “law”. Because of their syntactic-semantic properties, these words are always followed by a verb, and even when it is logically possible for some of them, like the negative particles, to operate on a nominal sentence, a different particle would be used, as is illustrated by the following examples:

\[
\begin{align*}
\text{(22)} \quad \text{lays-a l-walad-u marīḍ-an} & \quad \text{“the boy is not sick”} \\
\text{lam yakun l-wala-du marīḍ-an} & \quad \text{“the boy was not sick”} \\
\text{lan yakūn-a l-wald-u marīḍ-an} & \quad \text{“the boy will not be sick”}
\end{align*}
\]

In the three examples, more or less the same proposition is negated; the only difference resides in time reference. Since the copula has no lexical content when in the present tense (cf. Fassi-Fehri, 1993), as in the first example, the sentence is constituted of a topic and a comment. The use of “lam” or “lan” in such a case is not allowed and “laysa” is used instead, unlike in the other two examples where the copula is a lexical verb. In brief, the four words are always followed by a verb.

Although these are all constituted of a single closed syllable, just like the weak forms of the particles discussed above, they never show the allomorphic variation exhibited by those particles, i.e. they do
not have equivalent strong forms. This fact may be a matter of coincidence, but it may as well be a consequence of their restricted distribution to preverbal positions. In other words, these words did not develop a strong form probably because they cannot occur before nouns or pronouns, unlike the particles which can. If this is indeed the case, it would lend more support to the morpho-phonological analysis advanced earlier.

This analysis predicts that all particles that can co-occur with both verbs and nouns will develop an allomorphy between a weak and a strong form. However, a set of particles similar in form to the previous four do not support this prediction. This set includes the question words “hal” and “kam” (how much/many), the subordinator “bal” (rather), and the coordinators “ʔaw”/ “ʔam” (or). According to grammarians, these can be used equally correctly with verbs as with nouns, but in either case, their form remains unchanged, as in these examples:

(23) hal waṣal-a l-ʔawlād “Have the children arrived?”
    hal l-ʔawlād-u waṣal-ū “Have the children arrived?”

Obviously, the second example occurs more frequently in MSA than in CA usage, but it would certainly not be ruled out by the grammar rules of CA. However, when a CA corpus such as the Quran is considered, the instances in which “hal” is followed by a verb far outnumber those followed by a noun. This word occurs 93 times in the Quran, but apart from a few cases followed by independent pronouns or prepositions, all the others are followed by a verb. This fact suggests that “hal” did not develop a strong form in CA probably because conditions rarely rose which required the repair strategies discussed above. A similar remark holds also for the other items. Like the case of “hal”, none of the 20 occurrences of “kam” precedes a noun. Likewise, “ʔam” occurs 137 times only 4 of which are followed by a definite noun, and “bal” occurs 122 times but only 10 tokens are followed by a definite noun. As to “ʔaw”, 280 instances of it are attested in the Quran, most of which are followed by a verb. Among the nouns which occur after this conjunction, only six take the definite article.

In brief, it seems that only particles which occur frequently before nouns in CA, as exemplified by the Quran, have developed strong forms. In comparison, those which occur typically before verbs or prepositions have only a weak form. This fact supports the analysis advanced above according to which the “-a” part of the strong forms discussed earlier is in fact epenthetic.

4 Conclusions: Vowel epenthesis in Old Arabic

Previously, the following questions were raised: a) How can the different epenthetic vowels be traced back to a single process in Old Arabic? b) How do modern Arabic dialects relate to CA as far as epenthesis is concerned? c) Can a plausible classification of Arabic varieties, including CA, be advanced on the basis of vowel epenthesis? It should be recalled that Arabic varieties have been classified into four major groupings, depending on their default epenthetic vowel. One group uses /i/; another uses /a/; a third one uses both vowels, depending on the phonological context, while a fourth group inserts a schwa. Since we have decided to ignore schwa epenthesis for the time being, we are left with two main options: to postulate /a/ or /i/ as the original epenthetic vowel from which all the attested strategies of epenthesis can be derived diachronically. For reasons of economy, a proto-system in which more than one epenthetic vowel is used will simply be discarded; we assume that, barring complications caused by language change, languages will normally use a single epenthesis strategy to repair illicit structures.

Based on the analysis presented in this paper, we can conclude that a-epenthesis was once a productive process. This conclusion follows from the fact that vestiges of this process can still be attested in some of these varieties, especially the conservative ones among them. Such conservative varieties
include classical ones, like CA, and modern ones, like Najdi Arabic. Therefore, an ancestor of such varieties can be postulated in which /a/ was regularly epenthesized when phonotactic constraints required epenthesis. Accordingly, varieties like Sanʕani, Makkan and some Jordanian dialects, in which epenthesis regularly involves only /a/, are more conservative than others in which this vowel has been replaced by a high front one. However, it is not clear why and how this replacement happened. It could have originated in vowel harmony still noticed in some of the modern dialects discussed in this paper or in /a/-raising generally known as “imāla”. Particularly in the Levant, /a/ raises to /e/ or even to /e/ in some contexts, a process that was recorded by early Arab grammarians and is reflected in some Quranic readings. It is possible that inserted /i/ resulted from the raising of epenthetic /a/.

The geographical distribution of dialects, as grouped by vowel epenthesis, can be suggestive. In particular, i-epenthesis is found Levantine, Egyptian and Iraqi dialects while a-epenthesis is attested mainly in parts of Jordan, the Arabian Peninsula and the Sudan. If indeed a-epenthesis is more conservative, the innovative epenthesis of /i/ must have originated in the north, probably in the Levant. Since Arab tribes witnessed large scale movements after the expansion of the Muslim empire, however, it cannot be stated with certainty which tribes used the innovation and which were merely influenced by contact with conservative dialects. More research is needed to elucidate this point as well as others relating to epenthetic vowels in other Semitic languages.

References


