On the status of derived affricates in Arabic dialects

Dua’a Abu Elhija and Stuart Davis
Indiana University

Arabic dialects vary as to the presence of affricates in their phonemic inventory. Many dialects, such as San’ani Arabic, have only the voiced palato-alveolar affricate /dʒ/; others, like Baghdadi have two (/dʒ/ and /tʃ/), and still other dialects, such as Cairene, lack affricate phonemes altogether. Although dialects differ on the presence of underlying affricates, many have derived affricates. These arise when the alveolar stop /t/ and the fricative /ʃ/ come together over a morpheme boundary or as a result of vowel deletion. In this paper we explore the phonological patterning of the derived sequence [t+ʃ] as a single affricated segment or as a bisegmental sequence. We examine the evidence from three dialects: San’ani, Cairene, and Iksali (a Palestinian variety) and show that its patterning differs among the dialects.

Keywords: affricates, monosegmental, bisegmental, Cairene Arabic, San’ani Arabic, Iksali Arabic

1. Introduction

Arabic dialects vary as to whether they include affricates as part of their phoneme inventory. One can find three major varieties with respect to the presence of affricates. There are many dialects such as San’ani (Yemen) Arabic (as well as Classical Arabic) that only have one affricate phoneme, namely the voiced palato-alveolar affricate /dʒ/. There are dialects such as Baghdadi Arabic (Blanc, 1964; Youssef, 2013) and many Levantine rural varieties that have both /dʒ/ and its voiceless counterpart /ʃ/. The development of a phoneme /ʃ/ comes about through language-internal mechanisms (e.g. k-palatalization in the environment of high front vowels plus paradigm uniformity), often reinforced by borrowings through language contact (especially contact with Turkish in the former Ottoman Empire). Finally, there are the many dialects of North Africa that have no affricate phonemes. In the dialects that range from Libya to Morocco the historical *dʒ is
realized as /ʒ/ under French influence; in Cairene Arabic (and in other varieties of Egyptian Arabic) the /dʒ/ of Classical Arabic has the reflex of a voiced velar stop /g/. Additionally, in the standard variety of Sudanese Arabic described in Bergman (2002) there are no affricates; here, the /dʒ/ of Classical Arabic has the reflex of a voiced palatal stop. While dialects differ as to the presence and number of affricate phonemes, many dialects have derived affricates. These arise when, for example, the voiceless alveolar stop /t/ and the voiceless palato-alveolar fricative /ʃ/ come together, often over a morpheme boundary or as a result of vowel syncope. Two examples from Cairene Arabic are given in (1), the first showing a derived [tʃ] through morpheme concatenation and the second showing [tʃ] as [t] and [dʒ], respectively, when it is important to show that the affricate is patterning as a sequence or when it is vague as to whether it patterns as a sequence or a single segment. The transcription [tʃ] and [dʒ] (with the ligature) is used to indicate a monosegmental interpretation of the affricate. (As a note on the transcription, we will transcribe the affricate /tʃ/ and /dʒ/ as [tʃ] and [dʒ], respectively, when it is important to show that the affricate is patterning as a sequence or when it is vague as to whether it patterns as a sequence or a single segment. The transcription [tʃ] and [dʒ] (with the ligature) is used to indicate a monosegmental interpretation of the affricate.)

(1)  
   a. /ma-darasit-f/ → [madarasitʃ] ‘she did not study’  
   b. /bi-ti-ju:f/ → [bitʃu:f] ‘you (m.s.) see’

Examples like (1) are commonplace in many dialects. The question that we explore in this paper is whether the derived sequence of [t + ʃ] acts as a single affricated segment with respect to the phonology or whether it retains its bisegmental status. While we will not explore the implications of our findings beyond Arabic, we do note that this question can bear on general theoretical issues regarding affricates, such as the nature of their feature geometric representation (Lin, 2011) and whether affricates constitute a natural class (van de Weijer, 2014). It also may have consequences for the validity of the notion of segment as a phonological entity as raised by Port and Leary (2005).

Before examining the question of the status of derived affricates in Arabic dialects, we note that underlying affricates in Arabic dialects pattern as single obstruent segments. This is not surprising with respect to /dʒ/ since dialects that have /dʒ/ lack /ʒ/ as a separate phoneme. Consequently, it would be odd if /dʒ/ displayed biphonemic behavior if there is no separate phoneme /ʒ/. Further, as Youssef (2013) makes clear for Baghdadi Arabic, phonemic /tʃ/ patterns as a single segment even though the dialect has both /t/ and /ʃ/ as separate phonemes. The evidence comes from phonotactic patterning. Baghdadi words can begin with (at most) two consonants, regardless of their sonority sequencing. This is shown in (2), taken from Youssef (2013).

(2)  
   a. [smi:n] ‘fat’  
   b. [tıra] ‘he bought’  
   c. [t-baddal] ‘he changed’  
   d. [la:n] ‘tongue’

© 2016, John Benjamins Publishing Company  
All rights reserved
Affricates can cluster in the same fashion, either as the first member of the cluster as in (3a-b) or as the second member of the initial cluster as in (3c–d).

(3) a. [dʒnu:b] ‘south’ b. [ʃru:x] ‘wheels’
   c. [ldʒa:m] ‘bridle’ d. [ʃuːʃ] ‘jaws’

Since there are no other word initial clusters of three consonants, it is quite suggestive that both of the affricates pattern as monosegmental. Further, Baghdadi words cannot end in a sequence of two distinct obstruents (Blanc, 1964). The only possible exceptions are words like (3d) that end in an affricate; again this implies the single segment status of underlying affricate phonemes. Similar arguments for affricate phonemes as monosegmental can be given from other dialects, but we do not show this here.

Given that underlying affricates in Arabic dialects tend to pattern as single segments, the focus of this article is on the monosegmental vs. bisegmental status of derived affricates, like those shown in (1) for Cairene Arabic. This is an intriguing question and one on which dialects vary. (The question is intriguing because phonetically identical utterances – sequences of [t] followed by [ʃ] – can sometimes be phonologically interpreted as one segment and sometimes as two.) As we will show, it is often the case that a number of arguments can be brought to bear on the question and usually dialects are quite consistent in treating derived affricates as either a single segment or a bisegmental sequence. We will suggest a preliminary position based on our study that if a dialect does not have any underlying affricates as phonemes, then it will treat derived affricates as a sequence of two phonemes. But if a dialect has one or more underlying affricates then it can either treat the derived affricate as monosegmental or bisegmental depending on the dialect. We will focus on the patterning of derived affricates in three dialects: San’ani Arabic, Cairene Arabic, and Iksali Arabic, which is a Palestinian variety spoken in the Lower Galilee in Israel. The discussion of derived affricates in San’ani will be based on Watson (2002), whereas the discussion of Cairene and Iksali will be largely based on our own work. These three dialects present an interesting comparison for two reasons. First, as mentioned, Cairene has no underlying affricates, San’ani has just /dʒ/, while Iksali Arabic has both the phonemes /dʒ/ and /ʃ/. Second, all three dialects have the negative verbal suffix /ʃ/, which attaches to the very end of a (non-future) verb form. Because the negative suffix can be immediately preceded by a person marker ending in /t/, there are many instances of potential derived /t+ʃ/ sequences. In Section 2 of this paper we consider the status of derived affricates in San’ani Arabic. Here we review the arguments put forward by Watson (2002) that derived [ʃ] patterns as a single segment. In Section 3 we consider the status of [ʃ] in Cairene Arabic. Youssef (2013) and Watson (2002) have disagreed on its status in this dialect: Youssef suggests
that it is a single segment while Watson maintains that it is bisegmental. However, neither Youssef nor Watson presents the full range of evidence bearing on the issue for Cairene Arabic. We will present a wide variety of phonological evidence that makes clear its bisegmental character. In Section 4 we consider Iksali Arabic where we show a difference between underlying affricates and derived affricates in that underlying affricates pattern as monosegmental while derived sequences of /tʃ/ pattern as bisegmental. We conclude by discussing the relationship of the segmental status of derived affricates to the existence (or non-existence) of underlying affricates in the phonemic inventory of Arabic dialects.

2. Affricates in San’ani Arabic

The status of the derived affricate [tf] as monosegmental or bisegmental in San’ani Arabic (henceforth, SA) has been discussed by Watson (2002, p. 59–61). There is no underlying voiceless palato-alveolar affricate /tʃ/ in SA, but it does have the voiced palato-alveolar affricate /dʒ/ in its phonemic inventory. Similar to what was mentioned earlier for Baghdadi Arabic, the affricate /dʒ/ must certainly be a single segment and not a sequence since SA does not have [ʒ] as a phoneme. It would be odd for /dʒ/ to pattern as a sequence if there is no phoneme [ʒ]. Nonetheless, the issue of the single segment versus bisegmental analysis of affricates in SA does come up with the status of derived [tf], which comes about through morpheme concatenation. Because SA employs suffixal /ʃ/ as part of its system of verbal negation, there are many cases of derived instances of [tf] (from /tʃ/) in San’ani given that the negative marker can occur immediately after the person suffix, which can sometimes be [-t] (as in /daras-t/ ‘I/you m.s. learned’) or end in [-at] (as in /daras-at/ ‘she learned’).

Watson (2002, p. 59–61) gives two arguments that support the status of derived [tf] as a single segment: the unexpected phonotactic patterning of /tʃ/ in word-final and word-initial position and the behavior of the sequence [tf] in loanwords from English. We consider the patterning of /tʃ/ in word-final clusters first. As seen by the data items in (4), SA words can end in a cluster of two consonants.

\[
\begin{align*}
\text{4.} & \quad \text{a. [gambart] ‘I/you (m.s.) sat’} \\
& \quad \text{b. [darast] ‘I/you (m.s.) learnt’} \\
& \quad \text{c. [ma: ka:n-ʃ] → [ma: kanʃ] ‘he was not’} \\
& \quad \text{d. [ma: yihibb-ʃ] → [ma: yhib] ‘he doesn’t like’}
\end{align*}
\]

The data items in (4c–d) show the effect of the negative suffix [-ʃ]. The suffix can form a cluster with a single final consonant, but will trigger vowel shortening (4c) if the prior vowel is long or degemination (4d) if the immediately preceding
consonant is a geminate. Now consider the data in (5) that involve underlying sequences of three different consonants.

(5) a. /gult + lih/ \(\rightarrow\) [gultalih] ‘I told him’ (Watson, 2002, p. 64)
b. /ma: kunt-ʃ/ \(\rightarrow\) [ma: kunctʃ] (\([ma: kunctʃ]\) ‘I/you (m.s.)
  was/were not’ – *[ma: kuntaʃ]
c. /ma: gult-ʃ/ \(\rightarrow\) [ma: gultʃ] (\([ma: gultʃ]\) ‘I/you (m.s.)
  didn’t say’ – *[ma: gultaʃ]

The item in (5a) shows that when a sequence of three consecutive consonants comes about through morphological concatenation, the vowel [a] is inserted between the last two consonants in the sequence. In this light, the data items in (5b) and (5c) are of interest. In each of these forms, the base verb with the person marking ends in a consonantal sequence: [nt] in (5b) and [lt] in (5c). When the negative suffix /-ʃ/ is added, one would expect vowel epenthesis to occur as in (5a), so as to avoid the sequence of three consonants. That is, the expected, but non-occurring output forms for (5b) and (5c) are *[ma: kuntaʃ] and *[ma: gultaʃ], respectively. But instead there is no epenthesis and the surface forms are [ma: kunctʃ] and [ma: gultʃ] as shown in (5b) and (5c), respectively. Watson observes that forms like (5b) and (5c) where [t] and [ʃ] cluster at the end of a word after another consonant constitute the only sequences in SA where three consonants are allowed. However, given that the [tʃ] can be considered as an affricate and given that SA has the affricate phoneme /dʒ/, it makes sense to view the derived [tʃ] sequence in (5b) and (5c) as the single segment [tʃ]. We take the data in (5b) and (5c) as providing strong evidence for the monosegmental analysis of derived /t + ʃ/ in SA.

The phonotactic behavior of /t + ʃ/ as monosegmental is consistent with an observation in Watson (2002) that the verbal prefix ti/ta- for Form V and VI (and Form II quadriliterals) can appear without its vowel before stems that begin with /ʃ/. This is shown in (6).

(6) a. [ta-ḥa:ka:] ‘to talk’
b. [ti-gabbab] ‘to become bitter’
c. [t-farʃaf] ‘to wear the shirshaf’

The lack of a vowel after the prefixal /t-/ in (6c) is consistent with /t + ʃ/ patterning as the single affricate [tʃ], although we would not consider it to be definitive evidence since there is also no vowel after the prefixal [t] with verb roots beginning in other sibilants as well, as in [t-zawweedʒ] ‘to get married’. While SA does not normally allow for word-initial clusters, it could be that obstruent-plus-sibilant clusters are allowed marginally; thus, a data item like (6c) is not compelling evidence for a single segment analysis of derived affricates, but is consistent with it. This matter is in need of further study.
A stronger piece of evidence for the single segment status of the derived affricates in (5b) and (5c) comes from the behavior of word final /tʃ/ in English loanwords as noted by Watson (2002). Consider the observation that the English word ‘sandwich’ is borrowed into SA as [sandawíːtʃ] with stress on the final syllable. The unexpected lengthening of the final vowel seems best explained if the final consonant sequence of /tʃ/ is being interpreted as the monosegmental affricate [tʃ]. This is because, as discussed by Watson (2002), a word that ends in two consonants would naturally attract stress to the final syllable even when the vowel is short. On the other hand, if a SA word ends in a single consonant, then stress can only fall on that final syllable if the vowel is long. If the final sequence of /tʃ/ in the borrowed word ‘sandwich’ ([sandawíːtʃ]) were interpreted as two consonants there would be no need to lengthen the vowel of the final syllable since that syllable would receive primary stress anyway. Consequently, the lengthening of the final vowel is consistent with the interpretation of /tʃ/ as a single consonant. This monosegmental interpretation of the final [tʃ] sequence of the borrowing ‘sandwich’ in SA is made clear by the comparison with how the same word is borrowed into Cairene Arabic. Here, according to Badawi and Hinds (1986), it is borrowed from English as [sandawíːʃ] with final stress, but without the vowel lengthening. As we will elaborate on in the next section, this is highly suggestive that the sequence [tʃ] has a bisegmental interpretation in Cairene Arabic, but a monosegmental one in SA.

To conclude, as Watson (2002) argues, there is consistent evidence that the derived instances of [tʃ] in San’ani Arabic pattern phonologically as a single palato-alveolar affricate segment [tʃ] and not as a bisegmental sequence. This is interesting because San’ani does not have /tʃ/ as part of its phonemic inventory, but it does have the voiced affricate /dʒ/ as a monosegmental phoneme. We suggest that there is systemic pressure for sequences of [tʃ] to pattern as a single segment; given that the voiced affricate [dʒ] occurs, [tʃ] would constitute its voiceless counterpart.

3. Affricates in Cairene Arabic

The issue of whether derived sequences of /tʃ/ in Cairene Arabic comprise a single segment or not is interesting because Cairene Arabic does not have any underlying affricates as phonemes. Crucially it has ‘giim’ (i.e. /g/) where Classical Arabic and many other dialects have the affricate ‘jiim’ (i.e. /dʒ/). Furthermore, the occurrence of /tʃ/ is common in derivation in Cairene Arabic (henceforth, CA) so one can examine its phonological patterning. Youssef (2013) and Watson (2002) disagree on the status of such sequences in CA. Youssef asserts that it is a single segment without offering any real phonological argumentation while Watson
On the status of derived affricates in Arabic dialects

maintains that it is bisegmental based on the patterning of the suffix \([-ʃ]\) as part of the negation process. However, neither Youssef nor Watson presents the full range of evidence bearing on the issue for CA. In this section we present six arguments providing evidence for the bisegmental status of \(/t+ʃ/\) sequences. While virtually all the evidence from CA is consistent in supporting the bisegmental character of derived affricates in CA, we will also discuss one piece of evidence that seems more consistent with a monosegmental analysis.

The first argument bearing on the analysis of whether a derived sequence of \(/t+ʃ/\) patterns as monosegmental or bisegmental comes from the well-known stress pattern of CA. As made known from Hayes (1995, p. 67–71), in CA words in which the final syllable has a short vowel, the final syllable receives stress only if it ends in two consonants, but not if it ends in one consonant. This can be seen in the comparison of \([\text{katabt}]\) ‘I/you (m.s.) wrote’ where stress is on the final syllable (indicated by the underscore) with \([\text{kabt}]\) ‘she wrote’ where stress is on the initial syllable. But if the word \([\text{ktabt}]\) is made negative by the suffixing of \([-ʃ]\) (which is part of the negation process) so that the negative form ends in a derived [tʃ] sequence (i.e. \([\text{makabtʃt}]\) ‘she did not write’), the stress shifts to the final syllable. This strongly suggests that the sequence transcribed as \([tʃ]\) and pronounced like the affricate \([t͡ʃ]\) is really two segments; otherwise the stress would not have shifted to the final syllable (or, the final vowel would have lengthened, \(*[\text{makatabiːtʃ}]\) ‘she didn’t write’, so as to have a final stress). Consequently, the stress attracting nature of final syllables that end in [tʃ] is strongly supportive of its bisegmental character.

The second argument bearing on the analysis of whether a derived sequence of \(/t+ʃ/\) patterns as monosegmental or bisegmental also relates to the suffixing of \(/ʃ/\) as part of the negation process. Watson (2002, p. 61) observes that CA words can end in any two consonants, but no word can end in three consonants; moreover, words cannot have three consecutive consonants. When such sequences arise, epenthesis occurs between the 2nd and 3rd of the three consecutive consonants. Examples are shown in (7).

\[
\begin{align*}
(7) & \quad a. \ /\text{bint-na/} & \rightarrow & [\text{bintina}] \text{ ‘our daughter’} \\
& \quad b. \ /\text{ʃuft-na/} & \rightarrow & [\text{ʃuftina}] \text{ ‘you (m.s.) saw us’}
\end{align*}
\]

Since the negation marker \(/-ʃ/\) can be suffixed to any past tense verb, we are able to observe what happens when it is suffixed after a cluster ending in /t/. Consider the data in (8).

\[
\begin{align*}
(8) & \quad a. \ /\text{ma-kabat-ʃ/} & \rightarrow & [\text{makabatʃ}] \text{ ‘he didn’t suppress’} \\
& \quad b. \ /\text{ma-kunt-ʃ/} & \rightarrow & [\text{makuntʃ}] \text{ ‘I/you (m.s.) was/were not’} \\
& & & (*[\text{ma-kuntʃ}])
\end{align*}
\]
In (8a), we see an example of /-ʃ/ attaching to a verb that ends in a single consonant. No epenthesis occurs. The key example, as pointed out by Watson (2002, p. 61), is (8b). Here, the verb before the suffixation of the negative marker /-ʃ/ ends in the cluster /-nt/. When the suffix /-ʃ/ is added, epenthesis occurs between /t/ and /-ʃ/, resulting in [makuntʃ]. This strongly suggests that the /t +ʃ/ sequence is being treated as bisegmental. The CA form [makuntʃ] should be compared with its cognate in San’ani Arabic in (5b), [ma: kuntʃ], where epenthesis fails to apply in the underlying cluster. In the San’ani case, the derived /t +ʃ/ cluster is patterning as monosegmental, whereas in CA it is clearly being treated as bisegmental.

A third argument for the analysis of the derived sequence of /t +ʃ/ in CA as a cluster comes from syncope. In Cairene Arabic, an unstressed (non-final) high vowel deletes in an open syllable when preceded by an open syllable. Examples of the syncope process are given in (9) where all forms have final stress because of the superheavy syllable (CVCC or CVVC).

(9)  a. /ma-fihim-ʃ/ → [maf.himʃ] ‘he didn’t understand’
    b. /bi-ći-naːm/ → [bit.naːm] ‘you (m.s.) sleep’
    c. /bi-ći-ʃuːf/ → [bit.ʃuːf] ‘you (m.s.) see’

The output form in (9c), [bit.ʃuːf] ‘you (m.s.) see’, is interesting because, through syncope, the consonants /t/ and /ʃ/ have come together. However, the derived sequence of [tf] in [bitʃuːf] does not act as a monosegmental affricate. This becomes clear when we consider the form in (10a) in which [bitʃuːf] is negated through the prefixing of [ma-] and the suffixing of [-ʃ].

(10) a. /ma-bitʃuːf-ʃ/ → [ma.bitʃuːfʃ] ‘you (m.s.) don’t see’
    *[mabʃuːfʃ], *[mabtʃuʃʃ]
    b. /ma-bitktib-ʃ/ → [mab.tik.tibʃ] ‘you (m.s.) don’t write’

What the form in (10a) indicates is that the sequence [tf] of [bitʃuːf] is being treated as a sequence. In the output for /ma-bitʃuːf-ʃ/ the high vowel /i/ fails to delete since the result, *[mabʃuːfʃ], would have a sequence of three consonants. If the sequence [tf] were a single segment then the environment for syncope should be met and the predicted output would be *[mab.tʃuʃʃ]. The fact that no syncope occurs in (10a) strongly suggests that the /t +ʃ/ sequence in CA is not a singleton phonologically. The form in (10b) is there to show that the high vowel in the inflectional syllable /bi/ can delete when occurring after the prefix /ma/ as long as the result is not a cluster of three consonants. The comparison of the vowel deletion in (10b) with the lack of deletion in (10a) provides strong evidence that the /t +ʃ/ sequence is being treated as bisegmental in CA.
On the status of derived affricates in Arabic dialects

The fourth argument for the bisegmental analysis of the /t + f/ sequence in CA comes from vowel shortening: In CA, a long vowel in a final syllable will shorten if before two consonants but not before a singleton. An example is given in (11) that does not involve the derived affricate.

(11) a. [na:m] ‘he slept’
   b. /ma-na:m-f/ → [manamf] ‘he didn’t sleep’

In (11b), the long vowel of /na:m/ shortens when the /-f/ suffix is added. Now consider the example in (12) where the verb [ʔare:t] ‘I/you (m.s.) read (past)’ is negated with the suffix /-ʃ/ so that the form ends in the derived sequence of [tʃ].

(12) a. /ʔare:t/ → [ʔare:t] ‘I/you (m.s.) read (past)’
   b. /ma-ʔare:t-ʃ/ → [mataritʃ] ‘I/you (m.s.) didn’t read’

In (12b), we see that the underlying long vowel of [ʔare:t], shortens to [i] when the negation marker [-ʃ] is suffixed. The vowel shortening provides strong evidence that the word final [tʃ] is patterning as a sequence. If it were patterning as monosegmental, then no shortening should occur. Thus, the observation that derived [t +ʃ] triggers vowel shortening is consistent with its bisegmental status.

The fifth argument for the bisegmental status of derived [tʃ] sequences in Cairene Arabic comes from the minimal word constraint. Watson (2002, p. 88–89), among others, observes that CA has a strong minimal word condition requiring content words to be bimoraic. Since a word-final consonant is considered extrametrical in CA, a monosyllabic content word cannot be [CVC]. It needs to be minimally [CVCC] (where the CC can be a geminate) or the vowel needs to be long (i.e. [CVVC]). Watson points to a number of content words in CA that underlyingly are subminimal (/CVC/). When such words surface without any suffix attached, they undergo final consonant gemination as shown in (13):

(13) a. [ʔab-u:k] ‘your (m.s.) father’
   b. /ʔab/ → [ʔabb] ‘father’
   c. /ʔax-u:k/ → [ʔaxu:k] ‘your (m.s.) brother’
   d. /ʔax/ → [ʔaxx] ‘brother’

The gemination of the final consonant in the surface forms of (13b) and (13d) adds an extra mora to the word so that the surface forms meet the strong bimoraic minimal word constraint of CA. Now, consider the form in (14):

(14) /xadʃ/ → [xatʃ] scratch *[xatʃtʃ]

The form [xatʃ] ‘scratch’ in (14) is revealing because the word ends in a [tʃ] sequence that is derived through voicing assimilation (i.e. devoicing of the /d/
before a voiceless consonant) rather than through morpheme concatenation or syncope. That the /d/ is underlyingly present in (14) is evidenced by the broken plural form [xuduːʃ]. The issue that arises is whether this derived sequence of [tʃ] in (14) patterns as monosegmental or bisegmental. If it were a monosegmental affricate the word would have the shape [CVC] and one would expect gemination as in (13b) and (13d), or vowel lengthening so that the word would be minimally bimoraic. However, the fact that no augmentation of [xatʃ] occurs in (14) suggests that the word is already bimoraic having the shape [CVCC], not [CVC]. This thus provides evidence that the [tʃ] sequence is phonologically acting as bisegmental and not as a monosegmental affricate.

The final evidence discussed here for the bisegmental character of [tʃ] sequences in Cairene Arabic comes from the patterning of [tʃ] in loanwords borrowed from English. (All loanword data reported here are taken from Badawi & Hinds, 1986 and Reynolds, 2013.) First, we observe that loanwords that end in a sequence of two final consonants are borrowed into CA without change and stress occurs on that final syllable in accordance with the stress rules of the dialect. Some examples are given in (15). (An acute accent over a vowel in the final syllable is shown to make clear that the final syllable receives stress if there is more than one syllable in the CA form.)

(15)  
a. [film] ‘film’
b. [bank] ‘bank’
c. [biríns] ‘prince’

On the other hand, a monosyllabic source word ending in C_oVC typically has the final consonant geminated so that the minimal word condition can be met with stress on that syllable, as in (16a–c). While the last two words in (16) are bisyllabic in CA, final gemination occurs on these words so as to preserve the final stress of the source word; as previously mentioned, a final syllable that is CVC would not receive stress in CA.

(16)  
a. [watˁtˁ] ‘watt’
b. [ʔabb] ‘up’ (as in ‘Seven Up’)
c. [duʃʃ] ‘douche’
d. [ʔistúkk] ‘stock’
e. [filúʃʃ] ‘flush’

Words with final affricates are borrowed without gemination. Three examples are shown in (17).

(17)  
a. [kilátʃ] ‘clutch’
b. [siwítʃ] ‘switch’
c. [sandawítʃ] ‘sandwich’
The loanwords in (17) pattern exactly like the words in (15) that end in two consonants. There is no gemination of the final consonant, which would occur if the English word-final affricate were interpreted as a single segment when borrowed into CA (as shown by the forms in (16)). The word in (17c), [sandawî:tʃ], is of particular interest since, as mentioned earlier, the word is borrowed with a long vowel on the last syllable in San’ani Arabic, [sandawî:tʃ]. Both dialects want to maintain a perceived final source stress from the English loanword (albeit, in the English pronunciation of the word ‘sandwich’ if there is a stress on the final syllable it would be a secondary stress), but because the final [tʃ] is interpreted as a single segment in San’ani, vowel lengthening must occur in order to maintain final syllable stress. No such lengthening needs to occur in CA since the final [tʃ] sequence is interpreted as bisegmental. As noted, CA words that end in two consonants always have final stress. Thus, the lack of gemination in (17) provides support for the bisegmental analysis of the sequence [tʃ] in CA.

So far, we have presented six arguments for the bisegmental status of derived [t+ʃ] sequences in CA. The evidence overwhelmingly supports a view that such sequences are treated as bisegmental in CA unlike San’ani Arabic. Youssef, however, has pointed out (personal communication) that the word for the country name ‘Chad’ is often pronounced in CA as [tʃa:d] with no “epenthetic” vowel either before the cluster or between the two consonants. Since CA typically does not allow words to begin with two consonants, the occurrence of [tʃ] unmodified at the beginning of the word [tʃa:d] ‘Chad’ is suggestive that the initial [tʃ] sequence is patterning as a monosegmental affricate. However, we can give three arguments that even in the word [tʃa:d] the [tʃ] at the beginning of the word is acting as a sequence. First, CA has at least one other word that begins in a sequence of consonants, namely [kwayyis] ‘good’, which begins with [kw]. This is a very common word and it is understood to begin with a sequence, though in careful speech there may be a short epenthetic [u] between the two consonants (Islam Youssef, personal communication). Thus, the fact that ‘Chad’ can be pronounced with the initial sequence [tʃ] is not totally an aberration in the dialect. It is known from the literature on loanword phonology that loanwords can introduce new sequences into a language (see Kang, 2011). Second, when preceded by a word ending in a single consonant, the initial [tʃ] in [tʃa:d] (or its adjectival form [tʃa:di] ‘Chadian’) may be split up by epenthesis, which inserts a high vowel after the second of three consecutive consonants (and can occur over a word boundary within a phrase). This is observed in (18).

(18)  a. [ʔalam tʃa:di] ‘Chadian pen’
    b. [ʔalam tʃa:di] ‘Chadian pen’

The form in (18b) shows evidence that word-initial [tʃ] behaves as a cluster since it can be broken up by epenthesis. If it were a monosegmental affricate, then
epenthesis should be impossible. The third argument for the bisegmental nature of the initial [tʃ] cluster in [tʃa:d] concerns the comparison with the pronunciation of the country name (or adjective) ‘Czech’. This is shown in (19).

\[(19) \ [tʃiːki]\ ‘Czech’ \*\[tʃiːki]\]

With the word for ‘Czech’, as shown in (19), a vowel occurs between the /t/ and the /ʃ/. Here, a word-initial /tʃ/ in the source form is broken up by a vowel. Thus, while the occurrence of the initial [tʃ] in [tʃa:d] is suggestive of a monosegmental affricate, a closer examination leads us to conclude that it behaves as a very marginal initial consonant cluster in CA.

To sum up regarding the status of derived sequences of [t + ʃ] in Cairene Arabic, Youssef (2013) maintained that such sequences were monosegmental while Watson (2002) had suggested their bisegmental nature. Neither researcher, though, considered the full range of evidence. In this section we have presented six arguments that when taken together provide a strong case that derived sequences of [t + ʃ] in Cairene Arabic pattern as bisegmental and not as a single affricate segment [tʃ]. Perhaps this is not surprising given that CA lacks affricates entirely in its phonemic inventory.

4. Affricates in Iksali Arabic

So far in this paper we have shown that derived affricates in Arabic dialects can either pattern as a single segment as in San’ani or as a bisegmental sequence as in CA. We suggested that the difference between these two dialects might be related to the fact that San’ani Arabic has the affricate /dʒ/ in its phonemic inventory whereas Cairene Arabic lacks phonemic affricates. As a consequence, there may be systemic pressure in San’ani Arabic to treat derived [t + ʃ] sequences as monosegmental so that on the surface there would be both voiced and voiceless affricates. Because CA lacks an affricate phoneme, there is no systemic pressure to treat derived [t + ʃ] sequences as monosegmental. An interesting question arises as to whether such derived sequences are treated as monosegmental in dialects that have both /tʃ/ and /dʒ/ in their phonemic inventory.

In this section we consider Iksali Arabic, a Palestinian variety spoken in the village of Iksal (about 16,000 inhabitants) in the lower Galilee area of Israel. We know of no previous study of this dialect and the data are based on the native intuitions and observations of one of the co-authors. This dialect includes both /tʃ/ and /dʒ/ as part of its underlying phonemic inventory. As we will show, there seems to be a difference between underlying affricates and derived affricates in that underlying affricates pattern as monosegmental while derived sequences of /t + ʃ/ pattern...
as bisegmental. In examining the patterning of the underived affricates /tʃ/ and /dʒ/ , we show that these underlying affricate phonemes pattern as monosegmental. First, with respect to /dʒ/, one would expect it to pattern as a single segment since Iksali Arabic lacks /ʒ/ as a phoneme. This makes it odd for /dʒ/ to pattern as the bisegmental sequence /d + ʒ/ if there would be no other occurrence of /ʒ/ except after /d/. This type of argument, though, has no bearing on whether /tʃ/ patterns as monosegmental or bisegmental since the dialect has both /t/ and /ʃ/ as phonemes. A second argument that underlying affricates pattern as monosegmental comes from the patterning of word-final consonant sequences. In Iksali Arabic, generally speaking, words are not allowed to end in a consonant cluster other than a geminate. However, there are a handful of exceptions involving words that end in a cluster where the second element of the final cluster is /dʃ/ and the first one /t/. Examples are provided in (20).

(20) a. [mardʒ] ‘field’
   b. [sardʒ] ‘saddle (for a horse)’
   c. [hardʒ] ‘lap’

The fact that such final clusters can end in /dʒ/ is certainly evidence that /dʒ/ patterns as a single segment; otherwise such words would end in a sequence of three consonants, which is not otherwise attested in Iksali Arabic.

While the two arguments mentioned above apply to the analysis of /dʒ/, there is also phonotactic patterning evidence supporting the monosegmental status of /tʃ/. A salient observation about Iksali Arabic is that words in this dialect cannot end in a sequence of two distinct obstruents. Yet, as seen in (21), words ending in /tʃ/ are common.

(21) a. [biratʃ] ‘pools’ *(birafʃ]
   b. [samatʃ] ‘fish’
   c. [faratʃ] ‘rubbed’
   d. [rabutʃ] ‘your (f.s.) father’
   e. [rimmitʃ] ‘your (f.s.) mom’

This observation would be odd under a bisegmental analysis since it would mean that the only exceptions to the constraint against words ending in two distinct obstruents would be just those obstruent sequences that are analyzable as affricates. Along these lines, we note from data like those in (20) that /dʒ/ also patterns as a single segment or else the data items in (20) would end in a sequence of obstruents.

Another piece of evidence that underived /tʃ/ patterns as a single segment comes from the minimal word constraint. As with Cairene Arabic, it seems that content words in the Iksali dialect cannot be [CVC], they need to be minimally
[CVCC] (where the CC can be a geminate) or [CVVC]. We observe that in Iksali Arabic there are [CVCC] nouns that can end in a final geminate obstruent including an affricate. This is shown in (22).

(22) a. [fatʃʃ] ‘jaw’ * [fatʃ] *[fatʃʃ]
b. [libb] ‘pulp’

If /tʃ/ were really the bisegmental sequence /t +ʃ/, then the gemination in (22a) would be unexpected. Instead, one might expect either the form * [fatʃ], or, more likely, * [fatʃʃ] since the dialect will frequently have a vowel between two consonants at the end of the word where other dialects do not, as in [milik] ‘possession’, where many other dialects have [milk]. The gemination in (22a), as opposed to the vowel insertion in * [fatʃʃ], constitutes strong evidence for the monosegmental analysis of underived /tʃ/; that is, /tʃ/ cannot be broken up (as in * [fatʃʃ]) because it is a single segment and /tʃ/ must geminate at the end of a /CVC/ word so as to meet the minimal word length requirement.

The data in (20)–(22), taken together, present strong evidence that both affricate phonemes /tʃ/ and /dʒ/ pattern as singleton consonants. Given that underlying affricate phonemes pattern as single consonants, one might expect derived affricates to pattern as such. Consider the data from Iksali Arabic in (23).

(23) a. /ma-darastuʃ/ → [madarastuʃ] ‘you (pl.) didn’t study’
b. /ma-katabʃ/ → [makatabʃ] ‘he didn’t write’ (compare with CA [makatabʃ])
c. /ma-naːmatʃ/ → [manaːmatʃ] ‘she didn’t sleep’ * [manaːmatʃ]
d. /ma-katbatʃ/ → [makatbatʃ] ‘she didn’t write’ * [makatbatʃ]

The data item in (23a) shows the suffixing of /-ʃ/ after a vowel-final verb. As in other dialects, the suffix triggers the lengthening of an immediately preceding vowel. The data item in (23b) shows the suffixing of /-ʃ/ with a verb ending in a single obstruent. Because /ʃ/ is an obstruent, and the dialect does not allow words to end in a sequence of two obstruents, what we find in (23b) for the underlying form /ma-katab-ʃ/ is the output [makatabʃ] with an inserted vowel between the two final obstruents so as to prevent the word from ending in a sequence of obstruents. Given this, the forms in (23c) and (23d) are revealing. We focus on (23c), /ma-naːmat-ʃ/. In this form, the suffix /-ʃ/ is attached to a verb that ends in the single consonant /t/, thus constituting a derived sequence of /t + ʃ/. Hypothetically, this sequence could pattern as a single affricate segment [tʃ], especially given that /tʃ/ is a phoneme in the dialect. So, one might expect, then, that the output for the form in (23c) should be * [manaːmatʃ] with a single affricate at the end, since, as
we observed in (21), words in Iksali Arabic can end in such an affricate. However, the actual output as shown in (23c) is [manaːmatʃ] with epenthesis between the /t/ and /ʃ/. This means that a derived sequence of /t+ʃ/ is being treated as bisegmental. Since the dialect does not allow words to end in two obstruents, epenthesis occurs in (23c) (and (23d)), just like in (23b). We take this as strong evidence that the dialect avoids treating /t+ʃ/ sequences as a monosegmental affricate.

To summarize, the preliminary analysis of affricates in Iksali Arabic suggests that they are underlying single segments, but potential derived instances of /t+ʃ/ pattern as a bisegmental sequence.

5. Conclusion

In this paper we have examined the monosegmental vs. bisegmental status of derived affricates in Arabic dialects. Derived sequences of /t+ʃ/ are common in Arabic dialects that have the verbal negation marker /-ʃ/ as a suffix. Since Arabic dialects possess very clear means of phonologically distinguishing between a single consonant and a cluster, it then becomes an interesting question as to whether a dialect treats sequences of /t+ʃ/ as a monosegmental affricate or as a biconsonantal cluster. In examining three dialects, we found that in San’ani Arabic such derived sequences behave as a monosegmental affricate while in Cairene Arabic and Iksali Arabic the same sequences pattern as if they were bisegmental. One question that arises, then, is whether it can be predicted if a given dialect will treat /t+ʃ/ sequences as monosegmental or bisegmental. Based on our small survey, we would like to put forward one claim, namely that if a dialect does not have any affricates in its phonemic inventory then its derived affricates will pattern as a cluster, not as a single segment. The evidence in Section 3 from Cairene Arabic, a dialect lacking phonemic affricates, is unequivocal: in every meaningful phonological way, derived /t+ʃ/ sequences pattern in Cairene Arabic as bisegmental. While we make the claim that Arabic dialects lacking phonemic affricates do not have derived monosegmental affricates, we are aware that this is ultimately an empirical question and dialects such as those of North Africa and the Sudan, which lack phonemic affricates, need to be examined systematically.

On the other hand, it is hard to make concrete predictions regarding the status of derived /t+ʃ/ sequences in dialects that have affricates. We saw that in San’ani Arabic, which has underlying /dʒ/ but not /tʃ/, there is systemic pressure for derived /t+ʃ/ sequences to pattern as monosegmental, so that there would be a voiceless counterpart to the voiced affricate. However, in Iksali Arabic, which has both /dʒ/ and /tʃ/ as phonemic affricates, there was evidence that the dialect avoids treating derived /t+ʃ/ as a single segment. We do not know if the patterning...
reflected in these dialects is common or atypical and call upon other researchers to examine the patterning of both derived and phonemic affricates in the dialects with which they are familiar.

As a final matter that we will not pursue here, the patterning of affricates in Arabic dialects may have broader implications for phonological theory that go beyond Arabic. As one example, Port and Leary (2005) have argued against the validity of the notion of a segment (phoneme) in phonology. However, the fact that the same sound sequence (i.e. [tʃ]) can either behave as one element or two, can present a problem for a theory that argues against the validity of the segment. Other issues that the Arabic affricate data may bear on are representational issues (Lin, 2011) and whether affricates can constitute their own natural class (van de Weijer, 2014). We leave such matters for future research.

Acknowledgements

We wish to acknowledge James Dickens, Marwa Ragheb, Janet Watson, Jeroen van de Weijer, and Islam Youssef for comments on various aspects of this paper. We also wish to thank the two anonymous reviewers and the editors for their detailed and helpful comments.

References