# The Arabic definite article does not assimilate

BARRY HESELWOOD[a], JANET C. E. WATSON[b]
[a]*The University of Leeds, UK; b.c.heselwood@leeds.ac.uk*
[b]*The University of Leeds, UK; j.c.e.watson@leeds.ac.uk*

**Abstract**

In this paper we argue that the popular account of the phonology of the Arabic definite article in terms of assimilation to a following coronal consonant is not justified. The accepted account holds that the definite article has an underlying phonological form /ʔal/ (or /l/ in some versions) which surfaces as [ʔal] when the following word begins with a non-coronal consonant, but when the following word begins with a coronal consonant the /l/ completely assimilates to the coronal resulting in a geminate coronal consonant: compare for example *al-bint* [ʔalbint] 'the girl' and *al-zaffa* [azːafːa] 'the wedding procession'. We present theoretical and empirical grounds for rejecting the assertion that the /l/ of the definite article assimilates to a following coronal consonant in any synchronically meaningful sense of 'assimilation'. We argue that for something to count as synchronic assimilation it must be optional, meaning that an unassimilated pronunciation must also be allowed by the grammar. In the absence of counter-evidence, we also assume that optionality implies gradience. Historical assimilation, by contrast, admits neither optionality nor gradience, only unsystematic token-to-token variation in the realisation of the product of a historical process of change.

Using illustrative acoustic and electropalatographic data, the situation with the /l/ of the definite article when followed by a coronal consonant is compared to within-word sequences of /l/ + coronal consonant in *alzam* /alzam/ 'most necessary', *alṭaf* 'most kind', the form I doubled verbal coronal geminate in *hazza* /hazza/ 'to shake', and the optional assimilation of word-final /l/ to word-initial /r/ in *ḥabil rafī*ʿ 'a thin rope'. It is also compared with the optional assimilation of the definite article /l/ to a following dorsal stop in Cairo Arabic.

Electropalatographic and acoustic data are presented to support the argument that forms such as [azːafːa] should be regarded synchronically not as assimilation but as a type of 'true' or 'lexical' geminate resulting from phonologically-determined allomorphy.

## 1 The Arabic definite article

The definite article in Arabic forms a syntactic word with the noun or adjective which it defines (Watson 2002: 61–2), and in context forms a phonological word with the preceding syntactic word (ibid). Standard Arabic and the majority of the dialects outside the south-west of the Arabian Peninsula show six allomorphs (Haywood and Nahmad 1965: 22): it has the phonological form /ʔal/ or /ʔil/ when the noun or adjective is utterance-initial and begins with a non-coronal consonant or vowel; it has the form /l/ when the noun or adjective begins with a non-coronal consonant and follows a vowel-final word; when the noun or adjective begins with a coronal consonant, however, the consonant of the article must be that same coronal consonant, giving /ʔaC/ or /ʔiC/ in utterance-initial position and /C/ in utterance-medial position following a vowel-final word. In traditional Arabic grammar, the fourteen non-coronal consonants are known as *al-ḥurūf al-qamarīya* ('letters of the moon') (/b ʤ k q ʔ f χ ʁ ħ ʕ h m w j/), and the fourteen coronal consonants as *al-ḥurūf al-šamsīya* ('letters of the sun') (/t ṭ d ḍ θ ð ṯ s ṣ z ʃ l n r/).[1] Table 1 provides some examples from Standard Arabic.

| Gloss | Non-coronal initial consonant | Gloss | Coronal initial consonant |
|---|---|---|---|
| *the moon* | ʔal-qamr | *the sun* | ʔaʃ-ʃams |
| *the boy* | ʔal-walad | *the figs* | ʔat-tiːn |
| *the girl* | ʔal-bint | *the journey* | ʔas-safar |
| *the big book* | ʔal-kitaːbu l-kabiːr | *the long river* | ʔan-nahru ṭ-ṭawiːl |
| *the name* | ʔal-ism | *the wedding procession* | ʔaz-zaffa |

Table 1: Arabic definite article before a selection of non-coronal and coronal consonants

Many dialects in the south-west of the Arabian Peninsula, including dialects of southern Saudi Arabia and Yemen, do not exhibit the /l/ definite article in any environment. Several dialects of far northern Yemen and the coastal plain, including Saudi Rijāl Almaʿ (Asiri 2009) and Yemeni Minabbih (Behnstedt 1987: 85) exhibit /m/∼/am/∼/im/ with no (complete) assimilation to any following consonant, as in: *am-safar* 'the journey', *am-qamar*

[1] Emphatic consonants are represented with a subscript dot; /ʤ/ was historically a non-coronal /g/ although it now behaves as a coronal consonant in many modern varieties with respect to the definite article.

'the moon'. A selection of dialects in northern Yemen exhibit an /n/ definite article, which again shows no assimilation to any following consonant, as in northern Yemeni Majz *in-ṣa'bah* 'the female donkey foal', *in-šams* 'the sun'. Finally, dialects scattered throughout the western Yemeni mountain range, and some dialects of southern Oman, exhibit an article which involves gemination of any nominal-initial consonant, as in: *ab-bēt* 'the house', *ag-gamar* 'the moon', *ih-hōd* 'the wedding' attested, for example, in Rāziḥīt, Jiblah, Ġamar and Xawlān (cf. Behnstedt 1987: 85).

## 2 Review of accounts that take an assimilatory view

Most linguistic accounts of the definite article in dialects which exhibit the [ʔVl]∼[Vl]∼[l] ∼[ʔVC]∼[VC]∼[C] variants establish an underlying form /al/ (or /l/) with feature-spreading rules, or gestural phasing, to the left of a nominal with an initial coronal (Salib 1981, Watson 2002, Woidich 2006, Youssef 2013). Adopting feature geometry models, Watson (2002) and Youssef (2013) propose that definite article assimilation is motivated by a lexical violation of the Obligatory Contour Principle on the [coronal] tier—both /l/ and the adjacent coronal consonant are marked on the place tier as [coronal]. Assimilation to a following coronal consonant is total—manner and voice, and also place (e.g. ranging from dental /θ ð/ to postalveolar /ʃ/). Youssef (2013: 26) further proposes that assimilation of /l/ of the article to a following coronal results in a false geminate rather than a true geminate. What is interesting about these analyses is the fact that with the exception of /l/ of the definite article, /l/ in Arabic only ever assimilates productively and totally to a following coronal sonorant (Wensinck 1931, Watson 2002, Youssef 2013), rather than to all coronal consonants, as in Baghdadi /baddal-na/ [baddanna] 'we changed', /staʕmal raff/ [staʕmar raff] 'he made a shelf' (Youssef 2013).[2]

[2]Sibawayh discusses examples where /l/ may assimilate in Classical Arabic to dentals and interdentals, but states that assimilation of /l/ is less frequent before interdentals than dentals and less frequent before dentals than apical sonorants (Sibawayh II: 416–417, cited in Testen 1998: 151–152).

## 3 A non-assimilatory account

In this section, we begin by defining terms we use in this paper: 'true' geminate, 'false' geminate, 'fake' geminate, and assimilation. We will then consider examples of different kinds of phonetic accommodation in Arabic involving /l/ followed by a coronal consonant, and also different kinds of geminates involving coronal consonants, namely: within-word coarticulation, coarticulation across a word boundary, and 'true' geminates. Then we will consider which of these phenomena is most like what we observe when the definite article is followed by a nominal beginning with a coronal consonant, and by a nominal beginning with a velar stop in Cairene Arabic. Accommodations will be described and analysed in terms of articulatory gestures and gestural phasing (see for example Gafos 2002), illustrated with some electropalatographic (EPG) and spectrographic data from two male speakers speaking Modern Standard Arabic. Impressionistic transcriptions of the auditory qualities of the sequences will also be given. Table 2 presents the test words used for this part of the study and the number of tokens collected from each of the two speakers.

| Word | Gloss | Speakers A | Speakers B | Number of tokens |
|---|---|---|---|---|
| *alsan* | 'most eloquent' | 6 | 6 | 12 |
| *alzam* | 'most necessary' | 6 | 6 | 12 |
| *alṭaf* | 'most kind' | 6 | 6 | 12 |
| *χassa* | 'to be mean' | 6 | 6 | 12 |
| *hazza* | 'to shake' | 6 | 6 | 12 |
| *ḥaṭṭa* | 'to put' | 6 | 6 | 12 |
| *al-sahm* | 'the arrow' | 6 | 6 | 12 |
| *al-zaffa* | 'the wedding procession' | 6 | 6 | 12 |
| *al-ṭabaq* | 'the cover' | 6 | 6 | 12 |

Table 2: Test words and number of tokens.

### 3.1 Terminology

Relevant terms here are geminate types, coarticulation and assimilation. Regarding geminates, we distinguish 'true' geminates, 'false' geminates, and

'fake' geminates. 'True' geminates are typically defined as geminates which are 'monomorphemic and non-derived' (Davis 2011: 880, fn.7). They are also known in the literature as 'lexical' geminates because of their assumed non-derived status in the lexicon (e.g. Oh and Redford, n.d.). However, it is problematic to identify a geminate as 'true' or 'lexical' in a language exhibiting root-and-pattern morphology such as Arabic in which every occurring major class lexical item is said to be derived from an abstract consonantal root by the application of vowelling patterns, each pattern being the phonological form of at least one morpheme. Our suggestion is therefore to define a 'true' geminate as one which is obligatory, contrasts (at least potentially) with singletons, and displays 'geminate inseparability' (Gafos 2002: 274) which means that the articulation cannot be released until the end of the geminate; we avoid the term 'lexical' geminate. 'False' and 'fake' geminates emerge on concatenation: 'false' geminates through the concatenation of two identical consonants (cf. Oh and Redford, n.d.), e.g. *bad dog*, and 'fake' geminates through assimilation of one consonant to another (usually following) consonant, e.g. *bad boy* pronounced as [bab bɔɪ]. 'False' geminates may also be referred to as concatenative geminates, while 'fake' geminates may be referred to as assimilatory geminates.

We will use the term 'coarticulation' as a generic term for overlapping gestures whether within a word or across a word boundary. The term 'assimilation' will be used for cases of coarticulation where the realization shows a complete change of phonetic category due to the total displacement of one gesture by another, for example when the English phrase *in bits* is pronounced [ɪm bɪts] with no evidence of a coronal gesture during the pronunciation of *in*. Non-assimilatory coarticulation is when there is what Jones (1972: 217–21) describes as 'similitude', which is due to gestural overlap across the boundary between two segments such that one segment influences the production of the other, or they both influence each other.

### 3.2 Within-word coarticulation – *alzam, alṭaf*

In the word *alzam* /ʔalzam/ 'most necessary', the /l/ is the first radical of the root which is brought into contact with the second radical as part of the manner of forming elatives on the /ʔafʕal/ (أَفْعَل) pattern (Hayward and Nahmad 1965: 88–90). The first syllable, /ʔal/, is therefore phonologically the same as the form of the definite article when prefixed to a nominal begin-

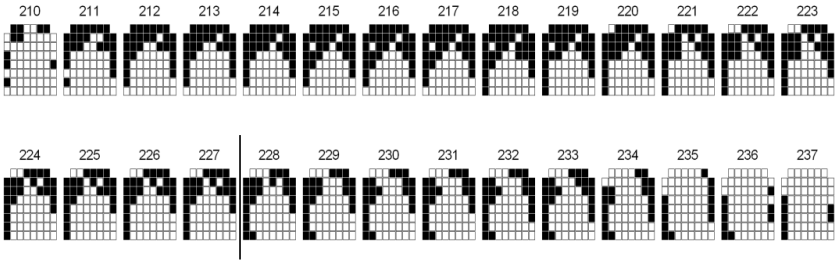



Figure 1: EPG frames of the /–lz–/ sequence in *alzam* – auditory impression [–lz–]; line indicates segment boundary.

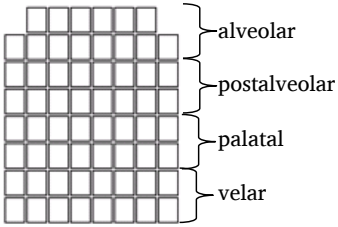



Figure 2: An enlarged blank EPG frame showing the articulatory zones. Right and left correspond to right and left in the speaker's mouth.

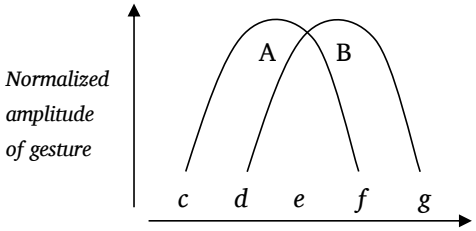



Figure 3: Gestures phased so as to overlap in time.

ning with a non-coronal consonant, e.g. *al-bint* /ʔalbint/ 'the girl'. Figure 1 presents EPG frames of the /–lz–/ sequence in *alzam* produced by a male Libyan speaker from Tripoli speaking Modern Standard Arabic (Figure 2 shows the articulatory zones of an EPG frame). These EPG frames, and those in other figures below, are sampled at 10 ms intervals. In Figure 1 they reveal a gradual reduction in the amount of tongue contact during the /l/ from the frame of maximum contact (frame 218 with 31 contacted electrodes) such that by frame 227 the contact is almost identical, with 24 contacted electrodes, to the contact pattern at the onset of /z/ in frame 228 with 23 contacted electrodes; the difference is that a central channel starts to open up for the fricative in frame 228. During the realization of /z/, perseveration of the lateral pattern of contact can be seen at the back righthand side where there is the same gap for lateral airflow as can be seen during the realization of /l/. The articulation of /z/ here can be symbolized as [ɮ] (IPA 1999: 188), denoting simultaneous lateral and central airflow, although we must assume that lateral airflow is minimal given that the auditory impression is of [z].

In gestural terms, Figure 1 shows that the articulatory configurations for realizing /l/ and /z/ overlap to some extent in a real-time dynamic relationship of mutual influence which is strongest at the segment boundary. Figure 3 models coarticulation by representing the phasing of gestures such that gesture B begins (point *d*) before gesture A has been completed (point *f*). The interaction of the two gestures is greatest at point *e* where the amplitude of gesture B equals that of gesture A; we can identify point *e* in Figure 1 between frames 227 and 228.

In addition to accommodations of the primary articulation gestures involving the tongue tip and lateral margins, there is auditory and acoustic evidence for accommodation of the tongue body as well. The /l/ is realized with a 'light' timbre before /z/ in *alzam* but a 'dark' timbre before the emphatic /ṭ/ in *alṭaf*. The influence of the /ṭ/ can be heard in both vowels which have the low back [ɑ] quality found in realisations of /a/ in the environment of emphatics in Arabic, in contrast to the front [a] quality in other contexts. The emphatic pharyngealization gesture for the realization of /ṭ/ is coarticulated with the gestures for the vowels and for /l/ with acoustic consequences in the form of prominent low-frequency resonances which can be seen in the righthand spectrogram in Figure 4. It is well-reported in instrumental studies of Arabic emphatics that the secondary articulation be-

gins earlier and finishes later than the primary articulation (see e.g. Watson 1999).

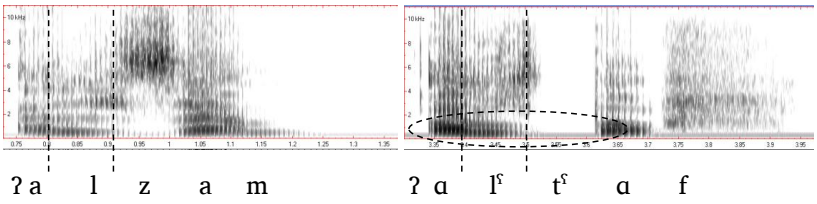



Figure 4: Spectrograms showing a front [a] vowel and clear [l] in *alzam* (left) and a back [ɑ] vowel and dark (pharyngealized) [lˤ] in *alṭaf* (right); dotted ellipsis picks out the thick band of low-frequency resonance responsible for the 'dark' timbre.

### 3.3 Coarticulation across a word boundary — *ḥabil rafī'*

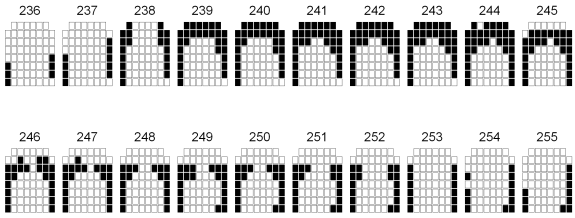
Arabic displays accommodation of manner of articulation among lingual sonorants such that those lower down the sonority hierarchy optionally assimilate across a word or morpheme boundary to those higher up the hierarchy to form what we label 'fake' geminates (see also examples and discussions in Garbell 1958: 326–7, Watson 2002: 237–9, Heselwood, Howard and Ranjous 2011: 63–6). The relevant part of the sonority hierarchy is shown in (1).

(1) n l r j

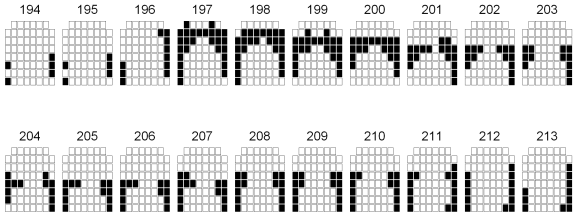
*Increasing sonority* →

An example of coarticulation across a word boundary is provided in Figure 5 where EPG frames of the /–l#r–/ sequence in three productions of the phrase *ḥabil rafī'* /ħabil rafiːʕ/ 'a thin rope' are shown. The speaker is a female from Syria, speaking the standard Damascus dialect of Syrian Arabic.[3] The first production (a) exhibits the same kind of coarticulation as we saw

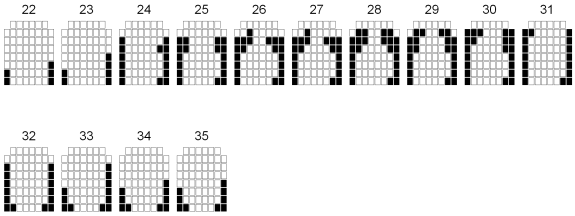
[3]Data reproduced from Heselwood et al. (2011: 84).




(a) Unassimilated /–l#r–/ with a short gestural overlap showing coarticulation[4] in frames 244–247 — auditory impression [–l ɹ–]




(b) Unassimilated /–l#r–/ but with earlier inception of the gesture for /r/ — auditory impression [–ˡɹ–]




(c) Totally assimilated /–l#r–/ with no gesture for /l/ — auditory impression [ɹː]

Figure 5: EPG frames showing a) some coarticulation, b) more coarticulation, and c) complete assimilation, in realisations of the sequence /–l#r–/ in *ḥabil rafī‘*

[4]Jones' (1972) 'similitude'.

above in *alzam*. More extensive coarticulation is found in the second one (b), while the third one (c) is an example of complete assimilation resulting in a geminate central approximant [ɹː]. Each will be described now in some detail.

In (a), frames 239–243 exhibit full alveolar closure and a gap at the back righthand side for the realization of /l/. In frames 46–252 we can see full lateral closures and a central open channel for an approximant realization of /r/; frames 244–247 show the transition from /l/ to /r/ where the two gestures overlap and influence each other. As in the case of *alzam*, the two adjacent consonants overlap but maintain perceptual distinctness. In (b), the lateral configuration for /l/ rapidly becomes more /r/-like through frames 197–200 while maintaining lateral openings for airflow until frame 204. Gestural overlap can account for this pattern in the same way as we saw above, the difference being that the gesture for /r/ begins earlier resulting in a realization of the sequence in which perception of the lateral is less clear. In terms of the diagram in Figure 3, point *d* is closer to point *c* in example (b) than in example (a).

The situation is rather different in 5(c) though. In this token, there is no perceptual or articulatory presence of a lateral consonant. The evidence is of complete assimilation with only a [ɹ]-gesture between the final vowel in *ḥabil* and the first vowel in *rafiʿ*. In terms of the diagram in Figure 3, there is no A gesture at all, and thus no point *e*.

The *ḥabil rafiʿ* examples show a range of degrees of accommodation of the final /l/ to the following initial /r/, from relatively short gestural overlap in 5(a) with a perceptually robust [l], through more extensive overlap in which [l] is less robustly present in articulatory and perceptual terms, to total assimilation in which there is no discernable [l]-segment in production or perception. The result of this total assimilation is a long [ɹː] extending across the word boundary, our 'fake' or 'assimilatory' geminate. An important observation in relation to these variable productions is that they demonstrate gradience and optionality. That is to say, the phonology of Arabic allows speakers to use any of these variants at fake geminate sites. A further very important point is that the possibility of releasing the word-final consonant before forming the following word-initial consonant decreases as the degree of gestural overlap increases, reaching zero in the case of total assimilation. A pronunciation such as *[ħabiɹ✓ ɹafiːʕ] is not attested in Arabic and we

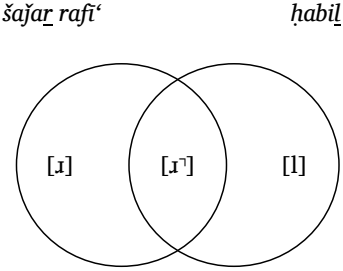



Figure 6: The main realizational possibilities for final /r/ in *shaǰar rafī‘* (left circle) and final /l/ in *ḥabil rafī‘* (right circle).

have what Gafos (2002: 274) calls ‘geminate inseparability’.[5] Similarly in English, if there is assimilation of /n/ in *in bits*, it cannot be pronounced *[ɪm✓ bɪts]. A crucial question concerning tokens which exhibit total assimilation is whether the final consonant of the first word has changed its phonological identity: has the /l/ of *ḥabil* become /r/, or can we justifiably and coherently say that the /l/ is realized as unreleased [ɹ˺] in this context? If /l/ has become /r/, then the realizational possibilities should be the same as for a word-final /r/ followed by a word initial /r/ as in *šaǰar rafī‘* /ʃaʤar rafiːʕ/ ‘a thin tree’: that is to say, if it is claimed that /l/ has become /r/, then it should behave exactly as /r/ behaves. The fact is that it does not. In *šaǰar rafī‘*, although there is the option of not releasing the [ɹ] of *šaǰar*, and thus forming a ‘false’ (concatenative) geminate, there is also the option of releasing it which demonstrates that a ‘false’ geminate does not exhibit geminate inseparability. ‘Fake’ or ‘assimilatory’ geminates do not have this option because the assimilated consonant must be unreleased or remain unassimilated. Following Heselwood, Howard and Ranjous (2011: 95–6), we can represent the difference between ‘false’ and ‘fake’ geminates using Venn diagrams, as in Figure 6.

Although final /r/ and final /l/ can both be realized as unreleased [ɹ˺], they nevertheless both have unique realizations in their respective sets of possible realizations and are thus disambiguated. This means that while neutralization of the /r/–/l/ opposition may occur phonetically in individual

[5] We use the tick ✓ to explicitly denote release of an articulatory constriction, whether audible or not.

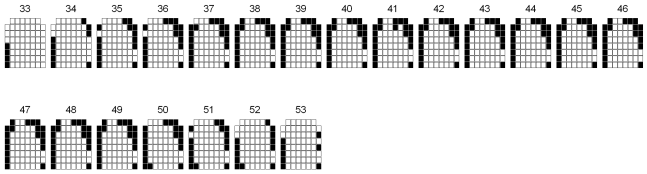



Figure 7: EPG frames showing the /zz/ coronal articulation in *hazza* — auditory impression [z:]

utterances, it does not occur in the phonological system because it is not obligatory. That is to say, in Saussurean terms, not all neutralizations which can be found in *parole* are neutralizations in *langue*.[6] To ensure that we do not misinterpret them to be so, we must take account of whole sets of possible realizations, not just those we observe on particular occasions.

### 3.4 'True' geminates — *hazza*

'True' geminates are geminates that exhibit geminate inseparability and contrast, at least potentially, with singletons (cf. above). Figure **??** presents EPG frames showing the close approximation constriction for the realization of the 'true' geminate /zz/ in a token of the Modern Standard Arabic form I doubled verb *hazza* /hazza/ 'to shake' (same speaker as in Figure 1).

The frames in Figure **??** show a central channel for airflow narrowing transversely in the alveolar region. The contact pattern remains stable (see frames 38–49), indicating that a single articulatory gesture is executed and maintained between the two vowels.

### 3.5 Definite article plus coronal consonant — *al-zaffa*

EPG frames for the so-called 'assimilated' /l/ of the definite article plus coronal /z/ in *al-zaffa* [az:af:a] 'the wedding procession' can be seen in Figure **??**. The speaker is the same as in Figures 1, 4 and **??**.

The patterns of lingual–palatal contact in Figures **??** and **??** are almost identical, and the auditory impressions are of the same long [z:]. This is of

[6] For Saussure's langue–parole distinction see Culler (1976: 29–34).

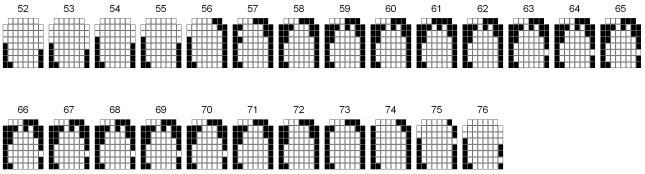



Figure 8: EPG frames for the underlined part of *al-zaffa* — auditory impression [zː].

course to be expected—nobody suggests that any trace of a lateral articulation can be observed in pronunciations of the definite article when it is followed by a word beginning with a non-lateral coronal consonant such as /z/. The question is therefore not whether there is evidence of [l] in *al-zaffa*, but whether the observed long [zː] should be analysed phonologically as containing an underlying /l/ which totally assimilates to the following /z/ in manner and aspect of articulation (in the case of words beginning with voiceless /t ṭ θ s ṣ ʃ/, assimilation of glottal state would also have to be postulated, and in the case of /θ ð ḍ ʃ/ assimilation of place of articulation), or whether it makes more sense to regard the phonological form of *al-zaffa* as /azzaffa/. Where we find clear examples of assimilation such as in *ḥabil rafī‘* pronounced as [ħabiɹ˺ ɹafiːʕ], we also find the possibility of unassimilated forms exhibiting various degrees of coarticulation consistent with real-time dynamic relationships between adjacent elements also seen between /l/ and /z/ in *alzam* in Figure 1, and between /l/ and /ṭ/ in *alṭaf* in Figure 3. Total assimilation resulting in ‘fake’ geminates across a word boundary, like in *ḥabil rafī‘*, is simply the limiting case of the element in the more dominant position exerting its maximum influence in real time but, crucially, without compromising the phonological identity of the assimilated element: /l/ remains /l/ even when realized as [ɹ˺] because of the other members of the set of possible realizations. In words such as *hazza* with ‘true’ geminates, however, the geminate is not a member of a set of alternative non-geminate realizations, and the same is true of *al-zaffa* and all other so-called definite article plus coronal consonant assimilations. Our contention is that they are not assimilations at all, but ‘true’ geminates which occur as phonologically determined allomorphs of the definite article.

Setting up /ʔal/ (or /l/) as the underlying form of the Arabic definite article, and deriving the geminated forms which occur in the context of a following coronal consonant from it through a process of assimilation, fails to account for the lack of optionality and gradience which real-time dynamic accommodations exhibit. There is no evidence at all of a real-time process of accommodation of /l/ to a following coronal of the kind we can observe in 'false' gemination when a speaker selects an item starting with a certain consonant and places it after an item ending in that consonant, or a consonant which can assimilate to it as in the example of *ḥabil rafi'* so as to produce a 'fake' geminate. Phonetic analysis of constructions such as *al-zaffa* strongly suggests that speakers do not select /ʔal/ (or /l/) and /zaffa/ and then put them into construction in real time. Rather, it suggests either that they select the syntactic element /azzaffa/ with its geminate /zz/ 'ready-made', or that speakers choose the definite article allomorph according to the initial segment of the defined word.

### 3.6 Definite article plus dorsal stop in Cairo Arabic

In modern Cairo Arabic, when a word beginning with /k/ or /g/ is preceded by the definite article, for example *il-kalb* 'the dog', *il-gaṛas* 'the bell', speakers have the option of pronouncing it either with /l/ ([ʔilkalb], [ʔilgaṛas]) or with a geminated stop ([ʔikːalb], [ʔigːaṛas]) (see Watson 2002: 217–22).[7] This optionality could be the same kind as the optionality attending accommodations across a word boundary as in *ḥabil rafi'*, that is to say an optionality in which a range of variants dependent on gestural phasing is implicated, or it could be a straight binary choice between a form with /l/ on the one hand, and a 'true' geminate form on the other. This question can in principle be answered by analysis of EPG data to see if there is any evidence of a gesture for /l/. Unfortunately, we do not yet have such data available.

## 4 Brief review of the historical development of the definite article

Part of the motivation for the assimilation analysis of definite article plus coronal consonant sequences is probably acceptance of the view that histor-

[7] In Cairo Arabic, as in many other dialects, the vowel in the definite article is /i/, not the /a/ of Modern Standard Arabic.

ically the /l/ assimilated to following coronals and that the underlying form of the article maintains an original /l/ in all contexts, an argument that is supported by Arabic orthography that represents the definite article as ⟨al⟩ (ال) in all contexts. Voigt (1998) and Testen (1998) adopt this position, arguing respectively for /al/ deriving from a demonstrative or an asseverative * *l* particle. There are, however, reasons for doubting this view, and several Semitic philologists have contested the claim that the definite article in Arabic is underlyingly /al/ from a historical point of view.

Kuryłowicz (1972: 131–2) argues that the historically recorded definite article with the principal allomorph /al/ is a relatively recent innovation. This is pursued by Zaborski (2000), who advances the intriguing proposition that the original article had three allomorphs according to number and gender: an *n*-based article for masculine singular, a *t*-based article for feminine singular, and an *l*-based article for plural. Since both /n/ and /t/ regularly assimilate, /l/ was selected as the orthographic form for the Arabic article as the consonant least prone to assimilation.

On the basis of historical reconstruction and discussion of the definite article in various Semitic languages, Wensinck (1931) and Ullendorff (1965) present cases for regarding the /ʔal/ form of the definite article as resulting from a historical process of dissimilation of non-coronal geminates: the Hebrew definite article is /ha/ with gemination of the nominal-initial consonant except where this is a laryngeal or pharyngeal; before laryngeals and pharyngeals, the form is /hā/ followed by a non-geminate. According to Ullendorff, the basic form of the Hebrew article is lengthening or tenseness with degemination and vowel lengthening before laryngeals and pharyngeals; according to Wensinck, the basic form of the Hebrew article is *hā-*. In the (vowel-less) Dadanite and Lihyanite inscriptions, the definite article is /h/ before all but laryngeals and pharyngeals, and is assumed (Ullendorff 1965: 635) to have induced gemination of the initial consonant. Before laryngeals and pharyngeals, gemination is dissolved by insertion of either /n/ or /l/, as in: *hlḥmq* and *hlḥmy* (Jaussen II, 474, no. 158, cited in Ullendorff 1965: 636; cf. also Macdonald 2000: 40). Noting further that dialects in Oman have been identified as optionally geminating initial labials and uvular consonants (b, f, q) (Rhodokanakis 1908–11), and that several dialects geminate velar consonants (cf. Watson 2002 and others), Ullendorff argues that the definite article in Hebrew and Arabic may be more closely related than previously thought, that definiteness was originally expressed by gem-

ination or tensing of the initial consonant, with degemination of glides, laryngeals and pharyngeals through /l/.

Ullendorff's argument for an original geminate definite article was later dismissed by Wagner (1993) on the basis of evidence from Modern South Arabian. Wagner argued that Mehri showed no morphological gemination, and therefore that gemination could not have been an exponent of definiteness. Work has, however, shown that both Omani Mehri and its sister languages Śḥerēt and Ḥarsūsī do exhibit gemination of the initial C in definite nouns and adjectives, particularly where this C is voiceless and non-emphatic, as in Mehri: *tōmar* 'dates' > *(a)t-tōmar* 'the dates', *xīl* 'uncle' > *(a)xxaylī* 'my uncle' (Watson 2012: 20–22).

## 5 Conclusion

We have argued that the geminates which occur in definite article plus coronal consonant constructions are not the result of synchronic assimilation and should instead be regarded as 'true' geminates, not assimilatory geminates. Our illustrative articulatory and acoustic data indicate that the geminate [z:] in *al-zaffa* is no different from that in *hazza*, and very different from within-word coarticulations involving /l/ and from 'fake' geminates resulting from assimilation of /l/ across a word boundary. Whether the Arabic definite article began historically with a ubiquitous /l/ which then assimilated to coronals, or whether /l/ was part of an alternation /n/∼/t/∼/l/ based on gender and number, or whether the marker of definiteness began as gemination and then developed an /l/-form in a process of historical dissimilation, makes no difference to our analysis, although of course it does make a difference to an account of how the current state of affairs arose.

The non-assimilatory account of the phonology of the Arabic definite article we have presented is better accommodated in a theoretical approach which does not assume invariant underlying phonological forms of morphemes from which observable variants are derived. The assumption found i.a. in generative phonology that a single morpheme must at some ultimate abstract level be instantiated by a single phonological form can be characterized as an example of a reification fallacy in which a single item in morphology is required to correspond to a single item in phonology. Setting up an underlying form with /l/ is very likely motivated by one or more of three factors. First is the fact that in written Arabic it is represented with

the letter corresponding to /l/ in all orthographic contexts, as mentioned above. Second is the widely-held view that the original form of the article had /l/ in all phonological contexts. Third is the fact that, in formulating a phonological rule to derive surface forms, it is relatively simpler to derive the geminate variants from an underlying /l/ because they can all be specified by the single feature-value [+coronal] with the [l] variant occurring 'elsewhere' (see Kiparsky 1973). Taking each of these motivations in turn: regarding the first point, it is often risky to base phonological analysis on orthographic evidence; regarding the second point, the original form of the definite article may or may not have had an /l/, the arguments for and against remain inconclusive; as for the third point, there is little if any justification for assuming that the simplest rule we can come up with bears any relation to what language users actually do in real time when speaking. To our knowledge, there is no evidence from available data that Arabic speakers, when saying *al-zaffa*, begin with /ʔal-zaffa/ and then assimilate the /l/ to the /z/. If that was the case, we would expect to find real-time dynamic influences of the kind seen in *alzam*, *alṭaf* and *ḥabil rafi'* as the articulatory gestures for realizing /l/ adapt to the local circumstances. In our opinion, the facts of the definite article in Arabic are best accounted for in terms of phonologically conditioned allomorphy, not by derivation from a single invariant form.

There is an interesting further consequence of our analysis for the notion of a 'true' geminate which we only have space to touch on here, but which we would nevertheless like to point out. It relates to the point made above about the problematic nature of the definition of a 'true' geminate as monomorphemic and non-derived. In a case such as *al-zaffa* the definite article morpheme has been prefixed to a noun, resulting in an obligatory, non-gradient and inseparable geminate [z:]. It is clearly not monomorphemic, and is clearly derived in the sense of coming about due to prefixation. Similarly, in a case such as *kassara* 'to smash (s.th.)' in which the obligatory non-gradient inseparable geminate [s:] is classed as a 'true' geminate, a morpheme with the function of intensifying a verb has been affixed to *kasara* 'to break (s.th.)'. *kassara* is thus grammatically derived from *kasara* by addition of a morpheme. There are therefore strong grounds for regarding these two geminates as belonging to the same type in Arabic. Our contention is that they are both 'true' geminates in the revised typology of geminates presented in Table 3, which does not appeal to derivational relations as a criterion.

| | TRUE | FALSE | FAKE | PSUEDO |
|---|---|---|---|---|
| | *hazza kassara al-zaffa* | *šaǰar rafiʿ* Eng. *bad dog* | *ḥabil rafiʿ* Eng. *bad boy* | Coll. *hazz* (?) |
| Inseparable | ✓ | ✗ | ✓ | ✓ |
| Obligatory | ✓ | ✗ | ✗ | ✓ |
| Potential contrast with singleton | ✓ | ✓ | ✓ | ✗ |

Table 3: A revised typology of geminates

If *kassara* contains a true geminate, then so do *al-zaffa, al-šams, al-tīn, al-nahr* and all words containing the definite article followed by a coronal consonant.

The category of 'pseudo' geminate covers those final long consonants in colloquial forms of Arabic resulting from vowel apocope, giving e.g. /hazz/ compared to MSA /hazza/. If there is no possibility of commutation with a singleton */haz/ because of syllable weight requirements for monosyllabic words, then the long consonant represents a neutralization between geminate and singleton for which 'pseudo geminate' may be an appropriate and useful term. However, further research is needed to confirm whether such neutralization does in fact occur.

## References

Behnstedt, P. (1987). *Die Dialekte der Gegend von Ṣaʿdah (Nord-Jemen)*. Wiesbaden: Harrassowitz.

Culler, J. (1976). *Saussure*. Fontana Paperbacks.

Davis, S. (2011). Geminates. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds), *The Blackwell Companion to Phonology Volume II: Suprasegmental and prosodic phonology*. Malden MA: Wiley-Blackwell, pp. 873–97.

Gafos, A. (2002). A grammar of gestural coordination. *Natural Language and Linguistic Theory*, **20**, 269–37.

Garbell, I. (1958). Remarks on the historical phonology of an East Mediter-

ranean Arabic dialect. *Word*, **14**, 303–37.
Haywood, J. A. and H. M. Nahmad (1965). *A New Arabic Grammar*. London: Lund Humphries.
Heselwood, B., S. Howard and R. Ranjous (2011). Assimilation of /l/ to /r/ in Syrian Arabic. In Z. M. Hassan and B. Heselwood (eds), *Instrumental Studies in Arabic Phonetics*. Amsterdam: John Benjamins, pp. 63–98.
IPA (1999). *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press.
Jones, D. (1972). *An Outline of English Phonetics*, 9th ed. Cambridge: Cambridge University Press.
Kiparsky, P. (1973). 'Elsewhere' in phonology. In S. Anderson and P. Kiparsky (eds), *A Festschrift for Morris Halle*. New York: Holt, Rinehart & Winston. pp. 93–106.
Kuryłowicz, J. (1972). *Studies in Semitic Grammar and Metrics*. Wroclaw: Ossolinuem.
Macdonald, M. (2000). Reflections on the linguistic map of pre-Islamic Arabia. *Arabian Archaeology and Epigraphy*, **11**, 28–79.
Oh, G. E. and M. A. Redford (n.d.) The production and representation of fake geminates in English. http://pages.uoregon.edu/redford/Papers4download/Oh&Redford_v2.pdf [accessed 28th November 2013].
Rhodokanakis, N. 1908–11. *Der Vulgärarabische Dialekt im Dofār (Zfār)*. Wien: ÖAW.
Sibawayh, Abū Bišr 'Amr b. 'Uṯmān b. Qanbar 1965/1317. *Kitāb Sībawayhi*. Beirut: al-Maṭba'ah al-Kubrā.
Testen, D. (1998). *Parallels in Semitic Languages*. Leiden: Brill.
Ullendorff, E. (1965). The form of the definite article in Arabic and other Semitic languages. In G. Makdisi (ed.), *Arabic and Islamic Studies in Honor of H. A. R. Gibb*. Leiden: Brill, pp. 631–7.
Voigt, R. (1998). Der Artikel im Semitischen. *Journal of Semitic Studies*, **43**, 221–58.
Wagner, E. (1993). Gedanken zum Verb des Mehri aufgrund der neuen Materialien von Johnstone. *Zeitschrift für Arabische Linguistik*, **25**, 316–39.
Watson, J. C. E. (1999). The directionality of emphasis spread in Arabic. *Linguistic Inquiry*, **30**, 289–300.
Watson, J. C. E. (2002). *The Phonology and Morphology of Arabic*. Oxford: Oxford University Press.

Watson, J. C. E. (2012). *The Structure of Mehri*. Wiesbaden: Harrassowitz.
Wensinck, A. J. (1931). The article of determination in Arabic. *Mededeelingen der Koninklijke Akademie van Wetenschappen* Deel 71, Serie A.
Youssef, I. (2013). *Place assimilation in Arabic: Contrasts, features, and constraints*. Ph.D. Thesis, University of Tromsø.
Zaborski, A. (2000). Inflected article in Proto-Arabic and some other West Semitic languages. *Asian and African Studies*, **9**, 24–35.