The Phonology of the Definite Determiners of Tihami Arabic

Mohammed J. Al-Ariqy
University of Utah

1 Introduction

Unlike in Standard Arabic and other Arabic dialects, the definite determiner in Yemeni Arabic dialects occurs in different forms and shows much more variety than its counterparts in other Arabic dialects. This paper is about the phonology of the definite determiners in three Yemeni Arabic dialects spoken within the governorate of Hajjah. The theme of this article was inspired by an atlas of the dialects of north Yemen created by Peter Behnstedt (2016), which briefly presents a few examples of the different definite determiners attested across some northern Yemeni dialects. This paper focuses on the definite determiner of a specific region of the north of Yemen and brings more data to investigate further the phonology of the definite determiner (DET) used in this region using an OT analysis.

The dialects addressed in this paper are spoken in Hajjah governorate, a north-western city part of the Tihama coastal line of Yemen. Its population is around 2 million. This paper analyzes data from three dialects in this province. The dialects are collectively known as Tihami Arabic. In two of these dialects, the determiner is /m-/, and in the third, it is /b-/, but I will analyze one of the /m-/ dialects and the /b-/ dialect. This paper mainly investigates two of these dialects, which show some interesting phonological alternations of the definite determiner. The definite determiner of one of these dialects is /m-/, and it will be referred to as the *OCP m-dialect.* I use the term OCP with this dialect because the Obligatory Contour Principle (OCP, Will Leben 1973) plays a major role in the analysis of this dialect. The *OCP m-dialect* is spoken in the mountains of Hajjah, and data was collected from two participants who live in the Kashar area. The definite determiner of the other dialect is /b-/, so this dialect will be referred to as the *b-dialect*. The data for the *b-dialect* was collected from two speakers from WashHa, particularly the Bani Hani tribe. The 4 informants for these dialects are between the age of 30-40 years old. They are all males. They all have a bachelor's degree at least.

This paper is organized as follows. Section 2 introduces geminates and the approach I use to analyze initial geminates in Tihami dialects. Word-initial gemination is a common property between these dialects that results from the assimilation of the determiner into the following word-initial labial. Section 3 presents data from the *OCP m-dialect* and the *b-dialect* and analyzes the phonology of their determiners using optionality within OT. It also goes over interesting cases of DET in the b-dialect. Section 4 discusses the findings and concludes.

2 Geminates

I analyze geminates in this article based on theories arguing for the moraic weight representation of geminates proposed by Hayes (1989, 1995). Geminates are assumed to contribute a mora to the syllable weight based on this approach. Davis (1999), Topintzi (2008), and Topintzi and Davis (2017) argue that initial geminates are moraic, and I adopt their approach. We will see evidence for this in Tihami Arabic in sections 2.1 and 2.2. See the representation of the definite form [f-faa.nuus] 'the lantern' below.

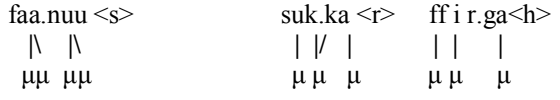
(1) Onset Geminates in Tihami Arabic

* I wish to thank the AMP 2022 participants for their insightful comments on this work. I am grateful to Aaron Kaplan for his feedback on this project.

Proceedings of *AMP 2022*

So, word-initial geminates contribute a mora to the syllable weight in these dialects. This type of geminates is attested in the Tihami dialects analyzed here only in the definite form of the nouns as part of the process of constructing the determiner phrase as will be shown in subsequent sections. They are limited to bilabial consonants only.

As for syllable weight, Watson (2007, 2011) argues that syllables in Yemeni Arabic are maximally bi-moraic, which is also the argument of this paper with these two Yemeni dialects. She draws a distinctness between non-geminate consonants, which are assigned a mora through Weight-by-Position, and segments, which are underlyingly moraic; geminates and vowels in this case. In this paper, I make the same distinction between CVC syllables (light), on the one hand, and CVV and CVG syllables (heavy), on the other. So, the first syllables in [f-fir.gah] 'the band' and [w-wardah] 'the rose' carry two moras based on this analysis, the initial geminate contributes the first moras for each, and the second mora is contributed by the vowels. Word-final codas are assumed to be extrasyllabic in these dialects following Watson (2007, 2011). See the representation of moras for [faanuus] 'lantern' and [sukkar] 'sugar' below.



(2) Mora representation in Tihami Arabic

3 Definite Determiners of Tihami Arabic

3.1 *The OCP m-dialect*

The definite determiner in this dialect surfaces as [m-] in all environments except before labials, and I, therefore, assume it is /m-/ underlyingly. See (3) for some examples.

(3)	a.	tees	'sheep'	m-tees	'the sheep'
	b.	ħanaʃ	'snake'	m-ħanaʃ	'the snake'
	c.	ʔasˁli	'original'	m-ʔasˁli	'the original'
	d.	milħ	'salt'	m-milħ	'the salt'
	e.	χubz	'bread'	m-χubz	'the bread'
	f.	ʔajtaam	'orphans'	m-ajtaam	'the orphans'
	g.	raas	'head'	m-raas	'the head'
	h.	naas	'people'	m-naas	'the people'
	i.	ʔiʃaarah	'sing'	m-ʔiʃaarah	'the sign'

As seen in (3) the definite determiner creates onset clusters in the definite forms. The definite determiner surfaces faithfully as [m] before non-labials and before [m]. With words that start with the labials [b, m, w, f] the definite determiner displays optionality in the output except before [m]. See some examples in (4).

(4)	a.	basˁal	'onion'	m-basˁal	~ b-basˁal	'the onion'
	b.	firgah	'band'	m-firgah	~ f-firgah	'the band'
	c.	wardah	'roses'	m-wardah	~ w-wardah	'the roses'
	d.	matˁar	'rain'	m-matˁar		'the rain'

The definite determiner either fully assimilates to the following word-initial labial or surfaces faithfully as a labial nasal. It is worth pointing out that these two options lead to the same result when the following sound is [m], hence the lack of optionality in (4d). This optionality is assumed to be triggered by the OCP-LAB constraint, which penalizes adjacent labial consonants, such as partial geminates like [mf..], on the boundaries of the prosodic word; I am assuming the definite determiner to be outside the prosodic word. So, the /m-f/, which are from separate morphemes, when they assimilate to [ff], are fused into one [f] (which is weighted with two moras). I specify that this constraint only applies to the labial sequences on the

boundaries of the prosodic word, the left boundary of the prosodic word because labial sequences within the prosodic word, such as [ha**fw**ah] 'mistake' and [la**bw**ah] 'lioness' in this dialect are unaffected.

The assimilation triggered by this OCP-LAB constraint comes at the expense of preserving the input features of the determiner in the output. The assimilation of a nasal to a non-nasal labial violates IDENT-SONORITY (IDENT-SON), which penalizes changes to the sonority of the input segments. This constraint penalizes the increase or decrease of the sonority of the sonorant segments, so a nasal becoming glide or vice versa is penalized by this constraint because that means an increase or a decrease on the sonority scale of sounds. I am assuming the sonority hierarchy: vowels > glides > liquids > nasals > obstruents, with vowels having the highest sonority and obstruents having the lowest (Clements 1990, Kenstowicz 1994, Smolensky 1995, Parker 2011).

The form in (4b) surfaces optionally as [mfirgah] or [ffirgah]. The former violates OCP-LAB by having the [mf] sequence, the latter, however, satisfies this constraint through assimilation but violates IDENT-SONORITY instead. This optionality will be analyzed using the Partial Orders (PO) model of optionality proposed by Anttila (2007). The model assumes partially ordered constraints to produce more than one actual output. So, if X and Y are both possible attested outputs for an input Z, and if the ranking A >> B >> C gives us only X and A >> C >> B gives only Y, then to get both outputs we posit that constraints B and C are not ranked with respect to each other, and in any tableau a ranking is chosen at random: A >> B or A >> C. Adopting this model, IDENT-SONORITY and OCP-LAB will be partially ordered in relation to each other to produce both outputs in (5). Let the analysis start with these two major constraints in (5).

(5)

/m - firgah/	ID-SONORITY	OCP-LAB
☞a. mfir.gah		*(!)
☞b. ffir.gah	*(!)	

The output in (5a) violates OCP-LAB because of the two adjacent labials in its onset cluster. (5b) satisfies OCP-LAB but does so by changing the identity of the underlying determiner from nasal to obstruent through assimilation, thus violating ID-SONORITY. The dotted line represents the partial order. So, the ranking ID-SONORITY >> OCP-LAB produces (5a), and the opposite ranking produces (5b). Putting the exclamation mark in round brackets indicates the fatality of the violation under one of the two rankings, so the violation of candidate (5a) is fatal under the OCP-LAB >> ID-SONORITY ranking, while the violation of (5b) is fatal under the opposite ranking. These two constraints produce the variation in (3b). We can expand the tableau in (6) by considering more candidates and constraints. See (6).

(6)

/m - firgah/	ID-SONORITY	OCP-LAB	DEP-μ
☞a. mfir.gah		*(!)	
☞b. ffir.gah	*(!)		*
c. ʔ**a**m.fir.gah		*	*!*
d. ʔ**a**f.fir.gah	*		**!
e. bfir.gah	*(!)	*(!)	

The candidates in (6c, d) contain epenthetic syllables before the determiner. These two candidates are relevant here because this kind of epenthesis is attested elsewhere in this dialect, as we will see below. They satisfy *COMPLEX-ONSET, which bans onset clusters and is assumed to be low-ranking, at the expense of DEP-μ, which penalizes epenthetic moras. The segments which introduce the epenthetic moras in (6c, d) are in bold in the tableau. DEP-μ is low-ranking. I bring the candidates in (6c, d) only to show that they would not surface given the grammar of this dialect where DEP-μ outranks *COMPLEX-ONSET unless they satisfy a higher-ranking constraint. DEP-μ is violated by vowel insertion, gemination of the word-initial onsets, and the assignment of the determiner to the coda position, codas are assumed to contribute to the weight of the syllable in the dialects presented in this paper.

So, the candidates in (6c, d) violate DEP-μ. (6b) and (6d) also violate this constraint because they contain geminates which are not present in the input, the word-initial geminate in (6b) and the intervocalic geminate in (6d). The geminate created by assimilation contributes a mora to the weight of the first syllable in (6b). Onset geminates are assumed to be moraic in this dialect as mentioned in the geminates section

earlier. Candidate (6c) violates OCP-LAB because it contains a sequence of two labials on the boundary of the prosodic word in [ʔa**m.f**ir.gah], the boundary is in bold, and the sequence happens on the left boundary of the prosodic word once the definite determiner which is labial attaches to the word that starts with a labial. Candidate (6d) violates ID-SONORITY by changing the determiner from a nasal to an obstruent through assimilation. So, candidates (6c, d) are ruled out by DEP-μ given either PO rankings, and candidate (6e) is ruled out by violating both the high-ranking partially ordered constraints.

As (7) shows, a high-ranking IDENT-PLACE which preserves the place features of the output segments in the output is necessary.

(7)

/m - firgah/	ID-PLACE	ID-SONORITY	OCP-LAB	DEP-μ
☞a. mfir.gah			*(!)	
☞b. ffir.gah		*(!)		*
c. ʔam.fir.gah			*	*!*
d. ʔaf.fir.gah		*		**!
e. bfir.gah		*(!)	*(!)	
f. mθir.gah	*!			

This constraint is necessary because one way to satisfy OCP-LAB is by changing the place features of the prosodic word-initial labial when it attaches to the labial determiner. A candidate like *[m-θirgah] for the input /m-firgah/ in (7f) satisfies OCP-LAB by changing the place features of the prosodic word-initial labial instead of assimilation of the determiner, and it would be penalized by the high-ranking constraint IDENT-PLACE then. The winning candidate would then be [ffirgah]; one that preserves the place features of the prosodic word. This constraint is high-ranking in the grammar of this dialect. I will not show IDENT-PLACE or candidates that violate it in subsequent tableaux. See (8) for another example of optional labial assimilation. This tableau gives an example of a change to the sonority of the determiner on the sonority scale from nasal to glide.

(8)

/m - wardah/	ID-SONORITY	OCP-LAB	DEP-μ
☞a. mwar.dah		*(!)	
☞b. wwar.dah	*(!)		*
c. ʔuw.war.dah	*		**!
d. ʔam.war.dah		*	*!*

The faithful output (8a) keeps the nasal determiner and thus violates OCP-LAB. The assimilating output in (8b) satisfies OCP-LAB through assimilation. However, the sonority of the definite determiner is not preserved as glides are more sonorous than nasals, so this candidate violates ID-SONORITY. The word-initial geminate resulting from assimilation violates DEP-μ because onset geminates are moraic. Both candidates are optional, and the partial order of these two constraints produces them both. The other candidates are harmonically bounded and, therefore, cannot win under any ranking. The losing candidate in (8c) has the same violations of (8b) but with one more violation of DEP-μ caused by the epenthetic vowel. The same DEP-μ violation applies to the losing candidate in (8d) compared to the winning candidate in (8a).

If the root-initial syllable is heavy, the gemination of the determiner must be accompanied by epenthesis because onset geminates cannot be created in a syllable that is already heavy without the geminate. Epenthesis makes the geminates intervocalic in this case. Consider the data in (9).

(9)

a.	faanuus	'lantern'	m-faanuus	~ ʔaf.faa.nuus	'the lantern'
b.	fiiraan	'mice'	m-fiiraan	~ ʔaf-fiiraan	'the mice'
c.	waadi	'valley'	m-waadi	~ ʔuw-waadi	'the valley'
d.	baamija	'okra'	m-baamija	~ ʔab-baamija	'the okra'
e.	miizaan	'scale'	-----	ʔim-miizaan	'the scale'

So, a form like *[ffaanuus] where the onset geminate is not allowed because the onset geminate is created in a syllable that is already heavy without the geminate. Notice that when assimilation happens, word-initial geminates created in heavy syllables are resolved by epenthesis in (9a-d). When the word-initial consonant is [m] as in (9e), an epenthetic vowel is inserted before the determiner to avoid word-initial nasal geminates.

Compare this data to that given in (4), where the initial geminates created by the definite form are allowed because they do not make their syllables super heavy.

The ban on onset geminates in heavy syllables raises the question of why they are allowed in (4) but blocked in (9). The answer to this question is that initial geminates are moraic as mentioned earlier in section 2. If they fall in syllables that are already heavy, they make them super heavy. This violates the maximum syllable weight in this dialect which is bi-moraic.

So, a weight constraint like *μμμ penalizes superheavy syllables, which explains why word-initial geminates are banned in the definite forms in (9). To prevent superheavy syllables, the definite determiner either surfaces faithfully to create partial geminates [mf-, mb-] or the first part of the geminate is assigned to the coda of an epenthetic syllable after assimilation. See (10) for the evaluation of the partial geminate option.

(10)

/m - fiiraan/	*μμμ	ID-SONORITY	OCP-LAB	DEP-μ
☞a. mfii.raan			*	
b. ʔaf.fii.raan		*!		**
c. ffii.raan	*!	*		*
d. ʔam.fii.raan			*	*!*

The faithful output in (10a) violates OCP-LAB, which is satisfied by violating ID-SONORITY in (b) and (c). However, candidate (10c) is ruled out by the high-ranking *μμμ which penalizes superheavy syllables. Output (10b), the other licit optional output, also violates ID-SONORITY but satisfies *μμμ through epenthesis, thus violating the low-ranked DEP-μ constraint. Epenthesis in (10b) resolves the ban on word-initial geminates in superheavy syllables by syllabifying the first part of the geminate as the coda of the epenthetic syllable, in other words by making the geminate intervocalic. ID-SONORITY and OCP-LAB must be partially ordered in relation to each other to yield both optional outputs in (10). Like the winning output in (10a), candidate (10d) violates OCP-LAB but is ruled out by DEP-μ. So, the PO ranking in (10) gives the partial-geminate variant. See (11) for the other PO ranking that produces the geminate-and-epenthesis variant.

(11)

/m - fiiraan/	*μμμ	OCP-LAB	ID-SONORITY	DEP-μ
a. mfii.raan		*!		
☞b. ʔaf.fii.raan			*	**
c. ffii.raan	*!		*	*
d. ʔam.fii.raan		*!		**

So, to get the geminate-and-epenthesis output, the ranking of the constraints must be OCP-LAB >> ID-SONORITY, as shown in (11). One thing to notice here is that (11d) is ruled out by OCP-LAB under this ranking while it is ruled out by DEP-μ under the first PO ranking in (10). This candidate was competing with the partial-geminate attested output in (10) but is competing with the geminate-and-epenthesis output in (11).

Now that a grammar is developed to account for the labial cases in this dialect, let us test the grammar on one of the examples from (3) where the definite determiner surfaces faithfully, and no optionality is attested. See (12).

(12)

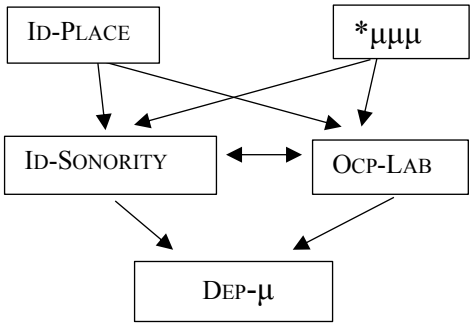
/m - tees/	ID-PLACE	*μμμ	ID-SONORITY	OCP-LAB	DEP-μ
☞a. mtees					
b. ʔam.tees					*!*
c. btees			*!		
d. ntees	*!				

Epenthesis in candidate (12b) is penalized by DEP-μ, and nasal dissimilation in (12c) is penalized by ID-SONORITY. The possible candidate in (d), where the determiner assimilates to the place feature of the following alveolar consonant, is assumed to be ruled by a high-ranking identity constraint such as ID-PLACE, which penalizes change to the input place features in the output. The faithful winning candidate does not violate any of these constraints. The constraint violated by the winning output is *COMPLEX-ONSET which is generally ranked low in the grammar of this dialect. So, the grammar developed so far can

account for all the facts about the determiner in this language. See the ranking of the constraints in (13), which is presented in a Hasse diagram in (14).

(13) ID-PLACE, *μμμ >> ID-SONORITY, OCP-LAB >> DEP-μ

(14) OCP m-dialect grammar



The Hasse diagram in (14) shows the relationship between the constraints that produce optionality in this grammar which are ID-SONORITY and OCP-LAB. While the ranking between ID-PLACE and *μμμ does not matter, the ranking between ID-SONORITY and OCP-LAB matters but is variable, indicated by the double-pointing arrow, these two variable constraints must be outranked by ID-PLACE, and they must outrank DEP-μ. Next, we will look at the other Tihami dialect whose definite determiner is [b-]. The definite determiner in this dialect shows more variation than the *OCP m-dialect*.

3.2 *The b-dialect*

The definite determiner in this dialect is /b-/. See (15) for some examples.

(15)	a.	θoor	'ox'	b-θoor	'the ox'
	b.	garjah	'village'	b-garjah	'the village'
	c.	ʁanam	'goats'	b-ʁanam	'the goats'
	d.	tˁullaab	'student'	b-tˁullaab	'the students'
	e.	ðahab	'gold'	b-ðahab	'the gold'
	f.	lisaan	'tongue'	b-lisaan	'the tongue'
	g.	zeet	'oil'	b-zeet	'the oil'
	h.	diraasah	'study'	b-diraasah	'the study'
	i.	jaasamiin	'jasmine'	b-jaasamiin	'the jasmine'
	j.	tees	'sheep'	b-tees	'the sheep'
	k.	kammuun	'cumin'	b-kammuun	'the cumin'

This analysis assumes /b-/ to be the underlying form of the determiner in this dialect. It occurs in more environments than the other derived forms of the determiner which, as we will see soon, occur in more restrictive environments. Complex onsets are attested in this dialect as the examples above show. There are, however, some environments where the definite determiner does not surface faithfully in this dialect. We will first go over these environments one by one with some examples. Like the *OCP m-dialect*, the definite determiner in this dialect fully assimilates to the following word-initial labial in (16). However, it is obligatory in this dialect but optional in the *OCP m-dialect*.

(16)	a.	faham	'understanding'	f-faham	'the understanding'
	b.	balaħ	'dates'	b-balaħ	'the dates'
	c.	bet	'house'	b-bet	'the house'
	d.	matˁar	'rain'	m-matˁar	'the rain'
	e.	wasˁijjah	'will'	w-wasˁijjah	'the will'

If the word is mono-syllabic and its syllable structure is CVCC, both the geminate-only and geminate-and-epenthesis outputs are optional. With such forms, assimilation of the determiner is obligatory in this dialect, while it is banned if not accompanied by epenthesis in the *OCP m-dialect*. See (17).

(17)

a.	wagt	'time'	w-wagt ~ ʔuw-wagt	'the time'
b.	waħʃ	'beast'	w-waħʃ ~ ʔuw-waħʃ	'the beast'
c.	ward	'roses'	w-ward ~ ʔuw-ward	'the roses'

However, if the word is mono-syllabic and its syllable structure is CVG where G stands for geminate, then only the geminate-and-epenthesis output is attested. The geminate-only output is banned. See (18).

(18)

a.	fann	'art'	ʔaf-fann	'the art'
b.	burr	'wheat'	ʔab-burr	'the wheat'
c.	bunn	'coffee'	ʔab-bunn	'the coffee'

Word-initial geminates in heavy syllables of the type CVV are attested in this dialect as opposed to the m-dialect which bans them in this environment. See (19).

(19)

a.	waadi	'valley'	w-waadi ~ ʔuw-waadi	'the valley'
b.	faatuurah	'bill'	f-faatuura ~ ʔif-faatuura	'the bill'
c.	baab	'door'	----- ʔib-baab	'the door'
d.	miizaan	'scale'	m-miizaan ~ ʔim-miizaan	'the scale'

This is one key difference between the two dialects regarding the determiner before labials. So, either onset geminates are not moraic in this dialect which then justifies their occurrence in superheavy syllables, or, like the OCP m-dialect, they are moraic but *μμμ is low-ranking. I assume the latter possibility in this analysis. In other words, the syllable weight requirement in the b-dialect differs from that of the m-dialect. In both dialects, initial geminates are moraic, but the syllable weight limit in these dialects differs. For the purpose of this paper, I will hold to this assumption and leave the issue for future research.

This dialect also displays another different behavior of the definite determiner in another context. The definite determiner becomes nasal when followed by nasals in a word-initial position. See (20).

(20)

a.	naħl	'bees'	m-naħl	'the bees'
b.	milħ	'salt'	m-milħ ~ ʔim-milħ	'the salt'

Preceding the pharyngeals and laryngeals [ʕ, ħ, ʔ, h], the definite determiner surfaces as [m-]. See (21).

(21)

a.	ʔibra	'needle'	m-ʔibra	'the needle'
b.	ħuut	'whale'	m-ħuut	'the whale'
c.	ʕasal	'honey'	m-ʕasal	'the honey'
d.	hunuud	'Indians'	m-hunuud	'the Indians

These are all the environments where the definite determiner does not surface faithfully in the b-dialect. In what follows, we will go over the analysis of each of these cases one by one.

The assimilation in (15) is triggered by the same constraint that bans adjacent labials we saw in the OCP m-dialect, which is OCP-LAB. This is one similarity between both dialects. Such a constraint can be satisfied by changing the place features of the labial determiner or the onset, e.g., *[θ-faham], which will be a violation for the high-ranking constraint ID-PLACE a constraint that penalizes change to the place features of the input. Epenthesis, e.g., *[ʔab.faham], does not satisfy OCP-LAB, and the epenthetic segments will be penalized either by DEP-μ, which penalizes epenthetic moraic segments, or by DEP, which penalizes epenthesis in general. In (22) is an example from the data set in (16). This example does not display optionality. Only one output is attested in the data. The same constraints used in the OCP m-dialect are imported here. The ranking is different, however.

(22)

/b - faham /	ID-PLACE	OCP-LAB	DEP-μ
☞a. ffa.ham			*
b. ʔif.fa.ham			**!
c. bfa.ham		*!	
d. θfa.ham	*!		

To produce this winning output only, DEP-μ must be low-ranking. This means that the ranking of the constraints that produce the geminate output must be ID-PLACE, OCP-LAB >> DEP-μ. Candidate (b) is ruled out by DEP-μ, the faithful candidate (c) is ruled out by OCP-LAB, and (d) is ruled out by ID-PLACE because the place features of the determiner changed.

While epenthesis was ruled out by the grammar in (22), it is attested in the data given in (17), which show two optional outputs for ([w-waħʃ], [ʔuw-waħʃ]) and ([w-wagt] ~ [ʔuw-wagt]), for example. The underlying form of the determiner is obstruent, and when it assimilates to the following labial nasal and glide, e.g., [w-waħʃ], its sonority changes which then violates ID-SONORITY, a constraint that penalizes the change to the sonority of the input segments on the sonority scale, as mentioned earlier. Let's start with the assimilating form [w-waħʃ] using the ranking from the previous tableau. See (23).

(23)

/b - waħʃ /	ID-PLACE	OCP-LAB	DEP-μ
a. bwaħʃ		*!	
b. mwaħʃ		*!	
☞c. wwaħʃ			*
d.ʔuw.waħʃ			**!
e.ʔub.waħʃ		*!	**
f. dwaħʃ	*!		

This ranking produces the first optional output for /b-waħʃ/ → [w-waħʃ], the faithful output in this case. Candidates (a, b, e) are penalized and ruled out by OCP-LAB. Candidate (e) also violates DEP-μ twice because two moras are epenthesized, one for the vowel and the other for the coda of the first syllable. Candidate (f) changes the place features of the determiner, so it is ruled out by ID-PLACE. The winning output (c) violates DEP-μ only. The output in (d), which is the other attested form, as we will see in the following tableau, violates DEP-μ twice by the epenthetic vowel and the geminate. ID-PLACE and OCP-LAB are not crucially ranked, but both must outrank DEP-μ. This grammar, however, cannot give us the other attested output (d) given these constraints only because both outputs only violate DEP-μ. To produce the other optional output, the geminate-and-epenthesis [ʔuw.waħʃ], we need a new constraint that is violated by the epenthesis-only output. We will use the *μμμ constraint used in the OCP m-dialect. This constraint penalizes superheavy syllables. It must outrank DEP-μ in this case. See (24).

(24)

/b - waħʃ /	ID-PLACE	OCP-LAB	*μμμ	DEP-μ
a. bwaħʃ		*!		
b. wwaħʃ			*!	*
☞c. ʔuw.waħʃ				**

The question is: can we get the other optional output; the epenthetic [ʔuw-waħʃ], using the PO model without *μμμ? No, because the epenthetic option and the geminating option only violate DEP-μ, so one of them must win all the time, and this should be the one that violates it the less; the epenthetic output cannot win. In this case, *μμμ is a necessary constraint for the PO model to work here. This constraint is violated by the geminating form but not the epenthetic form. *μμμ disfavors initial geminates falling in already heavy syllables. So, to get the optionality here, this constraint must be in partial order with DEP-μ. It must be outranked by OCP-LAB and ID-PLACE. See (25).

(25)

/b - waħʃ /	ID-PLACE	OCP-LAB	*μμμ	DEP-μ
☞a. wwaħʃ			*(!)	*
☞b. ʔuw.waħʃ				*(!)*

In cases of mono-syllabic words of the syllable structure CVG as in (18) above, initial geminates are banned. So, the only attested output for such forms is the geminate-and-epenthesis. We could get the lone

attested output in (26) by ranking *μμμ over DEP-μ invariably but doing so impairs the analysis in (25), where the ranking between these two constraints must be variable. So, instead we need a new constraint like *GVG. This constraint bans syllables containing two geminates. This constraint is *GVG and it is high-ranking.

(26)

/b - fann/	*GVG	ID-PLACE	OCP-LAB	*μμμ	DEP-μ
a. ffann	*!			*	*
☞b. ʔif.fann					**
c. bfann			*!		

So, candidate (a) is ruled out by this new constraint regardless of the variable ranking between the two low-ranking constraints. These syllables are, however, attested optionally in words with multiple syllables such as [fannaan] 'singer/artist'. In the definite form, this word is attested either as [ffannaan] or [ʔifannaan]. *GVG penalizes two geminates completely within one syllable (as opposed to geminates split between syllables). So, the difference between [ffan.naan] 'the artist' and [ffann] 'the art' is that the two geminates are not wholly contained within a single syllable in [ffan.naan]. See (27) for one example of such a case.

(27)

/b - fannaan/	*GVG	ID-PLACE	OCP-LAB	*μμμ	DEP-μ
☞ a. ffan.naan				*(!)	*
☞ b.ʔif.fan.naan					*(!)*
c. bfan.naan			*!		

Another difference between this dialect and the *OCP m-dialect* regards initial geminates in superheavy syllables. Word-initial geminates, which are the product of labial assimilation of the determiner, in superheavy syllables are attested in the *b-dialect* as we saw in (19), e.g., [ffaatuurah] ~ [ʔiffaaturah] 'the bill'. The *OCP m-dialect*, however, bans initial geminates in such syllables and restricts their occurrence to heavy (as opposed to superheavy) syllables only, e.g., [mfaaturah] ~ [ʔiffaaturah] but not *[ffaatuurah]. A superheavy syllable here is one that contains a geminate onset and a long vowel, GVV, or one that contains a geminate onset and a coda cluster, i.e., GVCC, and these are attested only in monosyllabic words as we have seen earlier in (17). So, in the *b-dialect*, the definite determiner either surfaces as an initial geminate in superheavy syllables, which I call the geminate-only variant (28c), or as a geminate accompanied by epenthesis, which I call the geminate-and-epenthesis variant (28d). The geminate-only case is one key difference between the *b-dialect* and the *OCP m-dialect,* and it is illustrated in (28). Compare it to (10-11).

(28)

/b - faatuurah /	*GVG	ID-PLACE	OCP-LAB	*μμμ	DEP-μ
a. bfaa.tuu.rah			*!		
b. mfaa.tuu.rah			*!		
☞ c. ffaa.tuu.rah				*(!)	*
☞ d. ʔif.faa.tuu.rah					*(!)*
e. ʔib.faa.tuu.rah			*!		**

The definite determiner in this word creates a superheavy syllable with the initial geminate in (c). This candidate violates *μμμ and DEP-μ, it violates DEP-μ because the initial geminate creates a mora that is not in the input. The other optional output satisfies *μμμ but violates DEP-μ twice because of epenthesis and the intervocalic geminate. The partial order between these two constraints produces both optional outputs. The candidates in (28a, b, e) are ruled out by OCP-LAB. This analysis is built on the assumption that initial geminates are also moraic in this dialect. The assumption is that initial geminates are not the most optimal onsets, and that epenthesis creates more optimal outputs.

The definite determiner in this dialect becomes nasal when preceding nasal-initial words as seen in (20). I deal with this as nasal assimilation. So, the determiner [b-] is [-nasal] and it assimilates to the following [+nasal] consonant. This nasal assimilation is triggered by AGR-NASAL constraint that requires adjacent consonants to agree in nasality. See (29).

(29)

/b - naħl/	AGR-NAS	*GVG	ID-PLACE	OCP-LAB	*μμμ	DEP-μ
a. bnaħl	*!					
☞b. mnaħl						
c. ʔib.naħl	*!					**
d. ʔim.naħl						*!*
e. nnaħl			*!		*	*

The underlying form of the root starts with a nasal consonant, so AGR-NAS compels the assimilation of the definite determiner into that nasal in (29b). This means the sonority of the determiner changes from obstruent to nasal, thus violating ID-SONORITY which I assume is low-ranking in this grammar. I will not bring it to the tableaus unless it is necessary. Candidates (a) and (c) are ruled by AGR-NAS, and the latter is penalized by DEP-μ. Candidate (d) is ruled out by DEP-μ. The definite determiner can satisfy AGR-NAS by fully assimilating into the following alveolar nasal, but this place assimilation is ruled out by ID-PLACE and is also penalized by DEP-μ in this tableau. It also violates *μμμ because of the geminate. So, there is no way for it to win under this ranking. In (30) is another example of assimilation which creates onset geminate and thus violates *μμμ.

(30)

/b - milħ/	AGR-NAS	*GVG	ID-PLACE	OCP-LAB	*μμμ	DEP-μ
a. bmilħ	*!			*		
b. ʔib.milħ	*!			*		**
☞c. mmilħ					*(!)	
☞d. ʔim.milħ						*(!)*

So, the two optional outputs are produced by the partial ranking just like any other optionality case we went over earlier. We see two possible outputs for this example because nasalization of the determiner before [m] also overlaps with the labial assimilation cases in superheavy syllables we saw in (17) and (19). In other words, since the initial geminate in (30) creates a superheavy syllable then *μμμ triggers the epenthetic variant in (30d). This is different from nasalization in (29) which does not create an initial geminate, thus lack of optionality.

In this dialect, as we see in the examples in (21), the definite determiner also becomes nasal before pharyngeals and glottals, sometimes referred to as laryngeals in some works on Arabic. I deal with this as nasal assimilation. Nasalization before this group of sounds is intriguing. No acoustic or experimental studies have been done on the pharyngeals of this dialect up to the time of writing this article. However, there are few experimental and acoustic studies on the nature of pharyngeals in some Arabic dialects such as Iraqi and Libyan, their effect on adjacent vowels, and the presence or absence of a nasal airflow during their production such as Khattab et al (2016), Ghazeli (1977), Lardi (1983), Butcher and Ahmed (1987), to name a few. Some of these studies, such as Khattab et al, found that the production of pharyngeals in some Arabic dialects was accompanied by a lowering of the velum and a nasal airflow that happened occasionally. What they found was that vowels get some nasal effect, perceptually, when surrounded by pharyngeals. Other studies, such as Ghazeli (1977), did not find any traces of nasalization during pharyngeal production. Yet, when nasalization was found, it was only in the presence of other nasals or vowels surrounding the pharyngeals. Only [ʕ, ħ] were addressed in these studies. Based on studies that claim a nasal airflow accompanying the production of pharyngeals in the dialects presented in these works, I make the hypothesis that it is the nasal airflow that accompanies pharyngeals in this dialect that triggers the assimilation but leave this gap for future research on the acoustics of these sounds in Tihami Yemeni Arabic. For sake of the analysis in this paper, I treat [ʕ, ħ, ʔ, h] as [+nasal]. So, the constraint that triggers the nasalization in (29), i.e., AGR-NASAL, also triggers the nasalization of the determiner before these sounds. The assumption is that the nasal airflow that accompanies the production of these sounds in this dialect spreads to the preceding labial determiner. See (31) for the evaluation of one example.

(31)

/b - ʕasal /	AGR-NAS	*GVG	ID-PLACE	OCP-LAB	*μμμ	DEP-μ
a. bʕa.sal	*!					
☞b. mʕa.sal						
c. ʔib.ʕa.sal	*!					**
d. ʔim.ʕa.sal						*!*

Candidate (a) is ruled out by AGR-NAS because the determiner does not assimilate with the following pharyngeal. The candidate in (b) satisfies AGR-NAS but at the expense of violating the low-ranking ID-SONORITY. Candidate (c) is ruled out early by violating AGR-NAS. It also violates DEP-μ. Candidate (d) satisfies AGR-NAS but loses to the winning output by violating DEP-μ.

It is time to test this ranking on an example from (15) where the definite determiner surfaces faithfully as [b-]. See (32).

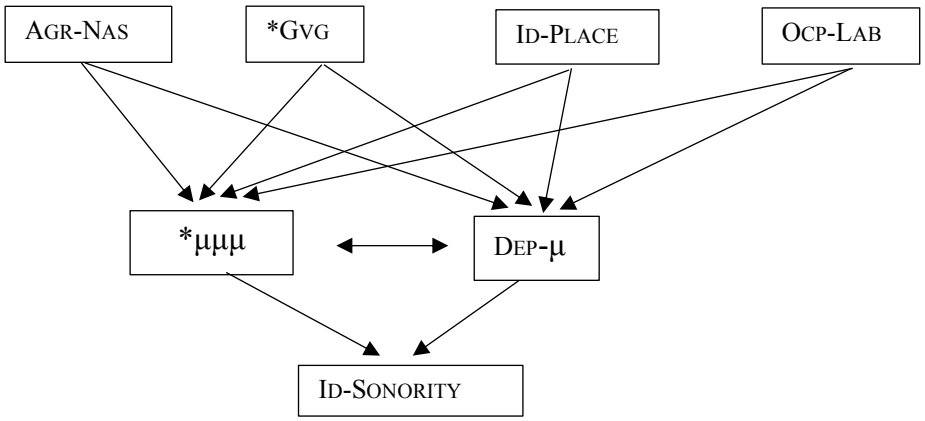
(32)

/b - θoor/	AGR-NAS	*GVG	ID-PLACE	OCP-LAB	*μμμ	DEP-μ
☞a. bθoor						
b. θθoor			*!		*	
c. ʔib.θoor						*!*
d. mθoor	*!					

The faithful winning output does not violate any of the constraints. Candidate (b) is ruled out by ID-PLACE because the determiner's place feature changes from labial to coronal. This candidate also violates *μμμ. Candidate (c) resolves the onset cluster in the winning candidate by epenthesis through violating DEP-μ. Two DEP-μ violations are incurred by this epenthesis, because the epenthetic vowel carries a mora, and the definite determiner falls in the coda of the epenthetic syllable which also contributes a mora. Candidate (d) is ruled out by AGR-NAS which compels the determiner to have the same nasal features of the following consonant. This ranking, then, captures all the facts in the *b-dialect*.

So, the grammar that accounts for the optionality and alternation of the definite determiner in the b-dialect is different from that of the m-dialect in some ways to be addressed in the discussion section. To conclude this section, see (33) for the Hasse diagrams of the b-grammar and the comparison between both.

(33) b-dialect Hasse diagram



This diagram shows that the top four constraints are not crucially ranked in relation to each other. They, however, must outrank *μμμ and DEP-μ. These two constraints are the ones that give optionality in this dialect. They must be in partial order which is indicated by the arrow pointing on both sides. OCP-LAB is high-ranking in this dialect because we saw previously in this section that the determiner cannot surface faithfully before labials. It must fully assimilate into the following labial. AGR-NAS is also high-ranking because the determiner must become nasal before nasals and pharyngeals. *GVG is high-ranking to ban syllables with two geminates, the case of mono-syllabic words in this dialect. In most cases where the definite determiner does not surface faithfully it, violates ID-SONORITY by either becoming a glide before glides or nasal before nasals, so this constraint is low-ranking and must be outranked by all other constraints as the Hasse diagram shows. This constraint is not necessary for this grammar, but I bring it for the sake of comparing its status to that in the *OCP m-dialect* where it is one of the two constraints that produce the optionality.

4 Discussion and Conclusion

This paper is about the definite determiners in Yemeni Tihami Arabic. The determiner is /m-/ for the *OCP m-dialect* and /b-/ for the b-dialect. In the *OCP m-dialect*, the determiner surfaces faithfully except before labials. When followed by labial consonants, the determiner fully assimilates into that labial optionally so as to create word-initial geminates. This assimilation is triggered by OCP-LAB, which bans adjacent labial consonants on the boundaries of the prosodic word and whose partial ranking with ID-SONORITY, which favors the faithful outputs, gives the optionality. However, assimilation of the determiner before labial consonants is banned when these consonants are in syllables that are already heavy. That is because the onset geminates created through assimilation will make these syllables superheavy, thus violating a high-ranking markedness constraint in the language; *μμμ.

The *b-dialect* is different in many ways from the *OCP m-dialect*. OCP-LAB is high-ranking in the b-dialect. In other words, the definite determiner cannot surface faithfully before labials. Second, superheavy syllables created by initial geminates through assimilation are optionally allowed, except if two geminates fall within the same syllable. So, as opposed to the *OCP m-dialect*, *μμμ is not high-ranking. In fact, it is one of the two constraints that give optionality in this dialect. The other constraint is DEP-μ which is violated by the other optional output, the geminate-and-epenthesis output. So, the partial ranking between these two constraints produces optionality in the b-dialect. The definite determiner in this dialect becomes nasal before nasal consonants and before the class of pharyngeals, this includes [ʔ, h]. This nasal assimilation is proposed to be triggered by AGR-NAS.

This paper is significant in many ways. First, it introduces the phonology of two definite determiners in Arabic that have not been discussed in the literature on Arabic definite determiners and develops an analysis that treats them as emerging from different permutations of a small number of constraints. It also displays a rare case of optionality with the definite determiner in Arabic and approaches it using the PO optionality model (Anttila, 1997, 2007). The paper also contributes to the limited literature on initial geminates in the world's languages, particularly to the literature on initial geminates in Arabic. The different labial geminate onsets introduced in this work are, to the best knowledge of the author, the first case of initial geminates in Yemeni Arabic, and moraic initial geminates in Arabic generally. The nasalization in the case of the pharyngeals and glottals in the *b-dialect* could be used as evidence that they make a subclass within the guttural class of sounds to the exclusion of uvulars which do not trigger nasalization in this dialect.

References

Anttila, Arto (2007). Variation and optionality. de Lacy, Paul (ed.), The Cambridge Handbook of Phonology, Cambridge University Press, Cambridge, 519–536.

Behnstedt, Peter (2017). Dialect atlas of North Yemen and adjacent areas. Brill.

Clements, George (1990). The role of the sonority cycle in core syllabification. Papers in laboratory phonology, 1, 283-333.

Davis, Stuart (1999). On the representation of initial geminates. Phonology, 16(1), 93-104.

Hayes, Bruce (1995). *Metrical stress theory: Principles and case studies*. University of Chicago Press.

Hayes, Bruce (1989). Compensatory lengthening in moraic phonology. Linguistic Inquiry, 20(2), 253-306.

Hayes, Bruce, Bruce Tesar, and Kie Zuraw (2013) "OTSoft 2.5," software package, http://www.linguistics.ucla.edu/people/hayes/otsoft/.

Kenstowicz, Michael (1994). *Phonology in generative grammar (Vol. 7)*. Cambridge, MA: Blackwell.

Muller, Jennifer (2001). The phonology and phonetics of word-initial geminates. Ph.D. dissertation, Ohio State University

Topintzi, Nina (2008). On the existence of moraic onsets. Natural Language and Linguistic Theory 26. 147–184

Parker, Stephen (2011). Sonority. In: van Oostendorp, M., Ewen, C., Hume, E., Rice, K. (Eds.), Companion to Phonology. Wiley–Blackwell

Ryding, Karin (2005). *A reference grammar of modern standard Arabic*. Cambridge university press.

Topintzi, Nina (2006). Moraic onsets. Ph.D. dissertation, University College, London. Topintzi,

Topintzi, Nina, & Davis, Stuart (2017). On the weight of edge geminates. The phonetics and phonology of geminate consonants, 2, 260-282.

Watson, Janet (2007). *The phonology and morphology of Arabic*. OUP Oxford.

Watson, Janet (2011). Word stress in Arabic. In Oostendorp, Marc, Colin Ewen, Elizabeth Hume, & Keren Rice (eds.) The Blackwell Companion to Phonology. Volume 5 Oxford: Blackwell.

Smolensky, Paul (1995). On the structure of the constraint component Con of UG (ROA-86).

Leben, Willaim (1973). Suprasegmental phonology [doctoral dissertation]. *Cambridge, MA: Massachusetts Institute of Technology*.