Linguistic Studies in the Arabian Gulf

edited by
Simone Bettega & Fabio Gasparini

«QuadRi»
Quaderni di RiCOGNIZIONI

Volume patrocinato dall'Università degli Studi di Torino

UNIVERSITÀ DEGLI STUDI DI TORINO

Simone Bettega, Fabio Gasparini (edited by), *Linguistic Studies in the Arabian Gulf*, Dipartimento di Lingue e Letterature straniere e Culture moderne – Università di Torino, Torino 2017 – ISBN 978-88-7590-113-4

In copertina: *Veduta dal Jabal Samhan* (Oman). Foto di Simone Bettega

Progetto grafico e impaginazione: Arun Maltese (www.bibliobear.com)

«QuadRi»
Quaderni di *RiCOGNIZIONI*
VII
2017

I «QUADERNI DI *RICOGNIZIONI*»

«QuadRi» – Quaderni di RiCOGNIZIONI è la collana curata dal Comitato scientifico e dalla Redazione di *RiCOGNIZIONI. Rivista di lingue, letterature e culture moderne*, edita online dal Dipartimento di Lingue e Letterature straniere e Culture moderne dell'Università di Torino. La rivista e i suoi *Quaderni* nascono con l'intento di promuovere ri-cognizioni, sia trattando da prospettive diverse autori, movimenti, argomenti ampiamente dibattuti della cultura mondiale, sia ospitando interventi su questioni linguistiche e letterarie non ancora sufficientemente indagate. I *Quaderni di RiCOGNIZIONI* sono destinati ad accogliere in forma di volume i risultati di progetti di ricerca e gli atti di convegni e incontri di studio.
ISSN: 2420-7969

COMITATO DI DIREZIONE

COMITATO DI REDAZIONE

SEGRETERIA DI REDAZIONE

COMITATO SCIENTIFICO

EDITORE

Dipartimento di Lingue e Letterature Straniere e Culture Moderne
Palazzo delle Facoltà Umanistiche
Via Sant'Ottavio, 20, Torino
Sito Web: http://www.dipartimentolingue.unito.it/

CONTATTI

RiCOGNIZIONI. Rivista di lingue, letterature e culture moderne
Sito Web: http://www.ojs.unito.it/index.php/ricognizioni/index
E-Mail: ricognizioni.lingue@unito.it
Issn: 2384-8987

Linguistic Studies in the Arabian Gulf

edited by
Simone Bettega & Fabio Gasparini

Università
di Torino

Dipartimento di
Lingue e Letterature straniere e
Culture moderne

All the contributions in the present volume have been subjected
to a process of double-blind review which attests their validity

SoMMARIO

Linguistic Studies in the Arabian Gulf

edited by Simone Bettega & Fabio Gasparini

INTRODUCTION*

Simone Bettega & *Fabio* Gasparini

It might strike the reader as odd that, to this day, the eastern shore of the Arabian Peninsula remains one of the linguistically least well-known areas of the Arabophone World. Indeed, such scantiness of documentation appears all the more curious if one considers the extreme richness and diversity of the region's linguistic landscape. This variety has no doubt arisen as a result of the Gulf's peculiar history: over the centuries, its waters and coasts have represented one of the world's most important and most diverse commercial hubs, a global crossroads for goods, peoples, cultures and languages. For hundreds of years the Ottomans, Persians, Portuguese and British have fought for control over the region, always confronted by the local powers. Maritime trade routes connected Arabia with Africa to the South and Asia to the East, and the Gulf ports were bustling with activity, swarming with merchants and travelers from three continents, in a dazzling kaleidoscope of languages. In the interior, the mountains and valleys of the southern Peninsula were home to centuries-old settlements, while nomadic peoples roamed immense distances across deserts and plains, their constant migrations representing one of the most powerful forces of linguistic change in the area. Some of the indigenous inhabitants of the region spoke varieties of Arabic, characterized by peculiar features rarely encountered, if ever, outside the Peninsula; others spoke different languages, loosely related with Arabic but ultimately not mutually intelligible. In the course of history, Arabic has taken root in Southern Arabia at the expense of other, pre-existing languages. Some of these have been extinct for centuries, and exist today only in the form of (sometimes enigmatic) carvings and incisions on rocks and monuments. Others are still spoken today, surviving at different levels of endangerment, their history and classification still a matter of debate among scholars.

The history of the Arabian Gulf, and its linguistic history more specifically, are without any doubt fascinating matters. But one would be wrong to assume that its present is any less interesting than its past. Certainly, Arabia is no longer the land of nomads and caravans, the fabled land of frankincense and a mandatory stop along the spice trade route. But it has lost nothing of its commercial importance: if anything, it has become one of the seats of the world's economic power. The discovery of vast

* This issue of "QuadRi – Quaderni di Ricognizioni" appears in a slightly altered graphic design, to allow the correct display of those symbols necessary for the transcription of Spoken Arabic and the Modern South Arabian Languages.

deposits of fossil fuels, after World War II, and the wealth derived by their exploitation, has brought about a change so drastic at the economic, social, and political level, that it hardly knows any parallels in the contemporary world. And, it goes without saying, social change brings along linguistic change. The streets of the thriving coastal towns are packed with immigrants from all over the globe, and the Gulf is again – as it has always been – a kaleidoscope of people and languages.

The complex and fascinating picture we have presented up to this moment is, we believe, a worthy object of study and research. The Arabian Gulf, intended as an abstract object of (socio)linguistic inquiry, confronts us with questions old and new: both are appealing and enticing, both difficult to answer. We hope this volume will be a welcome contribution in this sense, helping as it does in finding such answers, or, at the very least, paving the ground for addressing these questions from a more lucid perspective.

The nature of the contributions we present here is varied, as is their intended time-depth. Considering the variety of the matter at hand, this should come as no surprise. Some of the articles collected in this volume have a uniquely synchronic perspective: this is the case for Sabrina Bendjaballah's, Simone Bettega's, Emma De Murtas' and Roberta Morano's contributions. While the first investigates certain phonological and morphological phenomena characteristic of the Mehri language of Oman, the latter three chapters focus on various syntactic, semantic and lexical aspects of the Arabic dialects of Oman and the Gulf coast more broadly. Taken together, they provide solid proof of how the study of little-documented, often neglected "marginal" varieties can contribute importantly to the definition of central problems in general linguistic theory. This is true, as well, of the work of Fabio Gasparini, whose phonetic analysis of of Baṭḥari is permeated by the additional dimension of urgency concerning the documentation of a heavily endangered language, inexorably destined to disappear in the coming years. The contributions of Julien Dufour and of Janet Watson and Abdullah Musallam al-Mahri are also concerned with some of the Modern South Arabian Languages. While the former can be said to adopt an openly diachronic perspective – providing us with some fascinating insights on the developments of these languages through time – the latter can be thought of as a diachronic study only in relative terms: what Watson and al-Mahri focus on is the indissociable relation which exists between a language, its speakers, and the environment they live in (a theme also adumbrated in De Murtas' article). The question which emerges from this relation is a truly fundamental one: what happens to a language when its environment is drastically altered? What are the linguistic outcomes of strong nonlinguistic stimuli? Sometimes, this process results in language loss, as appears to be the case with Mehri, Śḥerēt (Watson and al-Mahri) and Baṭḥari (Gasparini). Sometimes, however, new forms come into existence as the result of movement, contact, and multilingualism: this is best exemplified in Andrei Avram's contribution, which investigates the origin of several traits that characterize the variety commonly referred to as Gulf Pidgin Arabic. Avram closely inspects the possible sources of these traits, and in so doing unearths parts of the Gulf's rich linguistic history. Such richness is also alluded to in Morano's article, and it is central in the work of Dénes Gazsi, where we see how the forces of history are embodied and made manifest into the linguistic fact. Gazsi's contribution reminds us that language and identity are inextricably entangled, and that discourse is the board onto which the game of self-definition is played. In a linguistic

environment as rich and multifarious as that of the Gulf, such game can grow exponentially complex: but it is in breaking down that complexity, and giving it meaning, that the work of linguists resides. In this sense, it seems to us, the collective efforts of all those who contributed to this volume have felicitously achieved their goals.

GUTTURALS AND GLIDES, AND THEIR EFFECTS ON THE MEHRI VERB*

Sabrina Bendjaballah

ABSTRACT • In the Mehri language of Oman, as in all other Semitic languages, gutturals and glides trigger perturbations on the verb stem, *e.g.* Ga-Stem √rkz *rəkú:z* "straighten" but √sbħ *sú:bəħ* "swim", H-Stem √nsm *hənsú:m* "breathe" but √bny *həbnó:* "build" *etc*. While those perturbations have long been known to exist, their exact nature is still not entirely clear. The aim of this article is to provide an adequate synchronic description of the perturbations triggered by the glides in two specific verb forms, the H- and $š_1$-Stems. The empirical basis includes the standard literature on the Mehri language of Oman as well as original fieldwork data.

KEYWORDS • Mehri, phonology, verb morphology, glides

1. Introduction

In the Mehri language of Oman, as in all other Semitic languages, gutturals and glides trigger perturbations on the verb stem. For example, in the basic stem (Ga-Stem), the regular perfective 3ms form has the shape $R_1əR_2ú:R_3$ (1a). By contrast, in the case of a guttural-final root, the stressed vowel does not show up between R_2 and R_3 but between R_1 and R_2. In the case of a guttural-medial root, the stem-vowel is not *ú:* but *é:* (1b). The perfective 3ms of glide-final and glide-medial Ga-Stems is given in (1c): in glide-final Ga-Stems, the final glide is not realized and the stressed vowel is *ó:* while Ga-Stems of glide-medial roots have a regular shape.

* Acknowledgments: I am grateful to my colleagues from the OmanSaM project (http://omansam.huma-num.fr/) as well as to Aaron Rubin for insightful discussions.

(1) Ga-Stem perfective 3ms[1]

a. sound	CCC	√rkz	rəkú:z	"straighten"
b. gutturals	CC**G**	√blh	bú:ləh	"keep so. off his work"
	C**G**C	√ghm	gəhé:m	"go in the morning"
c. glides	CC**Y**	√bdy	bədó:	"lie"
	C**Y**C	√syr	səyú:r	"go"

The effect of glides and gutturals is not uniform across verb stems. For instance, the alternations triggered by the gutturals in the perfective of the H-Stem differ from those observed in the perfective of the Ga-Stem: contrast (1b) with (2b). In guttural-final H-Stems, the vowel changes its quality: *é:* instead of *ú:*. By contrast, if the guttural is the second root-consonant, nothing specific happens.

(2) H-Stem perfective 3ms

a. sound	CCC	√nsm	hənsú:m	"breathe"
b. gutturals	CC**G**	√nfh	hənfé:h	"recover from a faint"
	C**G**C	√bhl	həbhú:l	"cook"

In some forms, *e.g.* the perfective 3ms of the Ga-Stem, gutturals and glides trigger different effects (1b *vs* 1c). However, in others, *e.g.* the perfective 3ms of the T_2-Stem, they behave alike: the stressed vowel of both guttural-final and glide-final verbs is located between the t-infix and R_2 (instead of between R_2 and R_3), and it is *ó:* instead of *ú:* (3bi, 3ci). Guttural-medial and glide-medial verbs have a regular shape (3bii, 3cii).

(3) T_2-Stem perfective 3ms

a. sound	CCC	√gml	əgtəmú:l	"do a favour to (b-) so.; be well off"
b. gutturals	i. CC**G**	√χlh	əχtó:ləh	"give birth prematurely"
	ii. C**G**C	√nhs	ənthú:s	"sigh"
c. glides	i. CC**Y**	√bdy	əbtó:di	"begin"
	ii. C**Y**C	√hyb	əhtəyú:b	"despair"

The alternations triggered by the gutturals and the glides are complex in the sense that they depend on the verb stem and on the position of the segment in the root. In this article I concentrate on a specific, well delimited configuration: the effect of the radical glides in the H- and $š_1$-Stems. My aim is to provide an adequate synchronic

[1] Conventions: G = guttural, Y = glide, √ = consonantal root, pf = perfective, ipf = imperfective, sbj = subjunctive, cd = conditional, 3ms = third person masculine singular. The data taken from the Mehri Lexicon (ML) are not marked, the data taken from the Mehri Texts (MT) are marked with subscript MT , the data taken from my own fieldwork are marked with subscript FW.

description of the alternations triggered by the radical glides in these stems. I will concentrate on the following questions:

(a) Are these alternations phonologically or morphologically conditioned?

(b) Do these alternations interact with the presence of other natural classes of consonants in the root? If yes, how?

2. Background information

Before we proceed, we need to specify the groups of consonants that constitute natural classes in Mehri. The consonant inventory of the Mehri language of Oman is given in (4).[2]

(4)

			lab.	dent.	alv.	lat.	pal.	vel.	uvul.	phar.	lar.
obstruents	plosives	-voice			t			k			ʔ
		+voice	b		d			g			
		ejective			t'			k'			
	fricatives	-voice	f	θ	s	ɬ	š		χ	ħ	h
		+voice		ð	z				ʁ	(ʕ)	
		ejective		θ'	s'	ɬ'	š'				
sonorants	nasals	(+voice)	m		n						
	liquids				r	l					
	glides						y	w			

[2] The voiced pharyngeal fricative [ʕ] is realized only in very few forms. However, it is clearly part of the phonological inventory of Mehri. For this reason, it appears in parenthesis in (4), but as a segment on its own right in (5). [ʔ] regularly surfaces at the phonetic level, its phonological status, however, is far from being clear. For this reason, it appears as a segment on its own right in (4), but in parenthesis in (5).

Based on phonological behaviour, five natural classes of consonants can be recognized: (5a-e), see Bendjaballah & Ségéral (2014: 165-169) for a survey and references. In this article I will argue that sonorants (5f) must be considered a natural class, too.

(5) a.	gutturals	χ ħ h ʁ ʕ (ʔ)
b.	glides	y w
c.	voiceless non-ejective consonants	f θ s t ɬ š k χ ħ h
d.	consonants with a lowering effect	θ’ s’ t’ ɬ’ š’ k’ χ ħ ʁ ʕ
e.	consonants that assimilate -t-	θ s t ɬ š θ’ s’ t’ ɬ’ š’ ð z d
f.	sonorants	m n w y l r

The second important piece of information concerns the interaction of stress, vowel quality, vowel length and syllabic structure. In Mehri, a stressed vowel alternates in quality and in length depending on the syllabic structure (*cf. e.g. ML*: xiv): *í:* and *ú:* surface as ə́ in closed syllable; *é:* and *ó:* as *á*.

(6)

syllable		example				
open	closed					
í:	ə́	θí:bər (*pf 3ms*)	θə́brək (*pf 1s*)	Gb-Stem	√θbr	“be broken”
ú:		yərú:kəz (*ipf 3ms*)	yərə́kzəm (*ipf 3mp*)	Ga-Stem	√rkz	“straighten”
é:	á	šəné:səm (*pf 3ms*)	šənásmək (*pf 1s*)	$š_2$-Stem	√nsm	“sigh”
ó:		aró:kəb (*pf 3ms*)	arákbək (*pf 1s*)	a-Stem	√rkb	“put (a pot) on fire’

3. Glides in the H-Stem and š1-Stem

In the Mehri verb system, there is no one-to-one correspondence between a given verb stem and a specific meaning. The main meaning of the H-Stem is a causative meaning, however this is far from being systematic (see *e.g.* Rubin 2010: 101). According to Johnstone, the $š_1$-Stem has a “causative-reflexive meaning”, “however the great majority of Š1-Stems can only be categorized as lexical” (Rubin 2010: 105). In this article, I concentrate on the phonology and morphology – not on the syntax and semantics – of the verb stems. The aim of this section is to describe the perturbations triggered by radical glides in the H- and $š_1$-Stems.

3.1. Distributional facts and general information

As far as their morpho-phonological shape is concerned, the Mehri H- and $š_1$-Stems behave in a parallel way (see *e.g.* Rubin 2010: 102). (7) shows their regular pattern.

(7)

		pf 3ms	*ipf 3ms*	*sbj 3ms*	*cd 3ms*	
		h/šəCCú:C	yəh/šəCCú:C	yəh/šáCCəC	yəh/šáCCəCən	
H-	√nsm	hənsú:m	yəhənsú:m	yəhánsəm	yəhánsəmən	"breathe"
$š_1$-	√rkz	šərkú:z	yəšərkú:z	yəšárkəz	yəšárkəzən	"give help"

Table (8) summarizes the distribution of glides in the H- and $š_1$-Stems, based on an exhaustive examination of *ML*. Both in the H-Stem and in the $š_1$-Stem, there is a majority of √CCY, followed by √wCC and √CwC, and finally √CyC. If there are two glides in the root, the major type is √Cwy.

(8)

	√		H-Stem	$š_1$-Stem
a. Y in isolation:	**Y**CC	= wCC	24	14
	C**Y**C	CwC	22	14
		CyC	12	8
	CC**Y**		55	58
b. Y in combination:	**Y**C**Y**	= wCY	4	2
	C**YY**	= Cwy	7	9
	YYC		0	0

Table (8) illustrates several distributional facts that are true of Mehri verbs in general:

(a) absence of y-initial triliteral roots.[3]

[3] The Mehri Lexicon gives a single verb derived from a y-initial root, and this root is a biliteral root: √ys' yə́s's' "be afraid".

(b) neutralization of the contrast w *vs* y in final position.[4]

(c) absence of √YYC: two glides may cooccur as R_1 and R_3, and as R_2 and R_3 but not as R_1 and R_2.

(d) all √CYY are √Cwy.

3.2. Glide-initial and glide-medial roots

(9a) shows the pattern of the H- and $š_1$-Stems derived from √YCC and (9b) that of the H- and $š_1$-Stems derived from √CYC. They are parallel to (7): √YCC and √CYC behave like sound roots.

(9)

	pf 3ms	*ipf 3ms*	*sbj 3ms*	*cd 3ms*	
	h/šəCCú:C	yəh/šəCCú:C	yəh/šáCCəC	yəh/šáCCəCən	
a. YCC					
H-	√wgɬ	həwgú:ɬ	yəhəwgú:ɬ	yəháwgəɬ	yəháwgəɬən	“take out (the animals)”
$š_1$-	√wgɬ	šəwgú:ɬ	yəšəwgú:ɬ	yəšáwgəɬ	yəšáwgəɬən	“go in the early evening”
b. CYC					
H-	√nwf	hənwú:f	yəhənwú:f	yəhánwəf	yəhánwəfən	“beckon so.”
$š_1$-	√ɬ'ym	šəɬ'yú:m	yəšəɬ'yú:m	yəšáɬ'yəm	--	“run short of milk”

Preliminary fieldwork data however show that the situation is more complex. Consider the sample of examples below, all from √YCC. For each root, the first line gives the forms given in ML, the second the one elicited during fieldwork.

[4] The Gb-Stem constitutes an exception to this generalization in that w-final roots and y-final roots do behave differently: e.g. pf 3ms √glw gí:ləw “have fever” vs pf 3ms √ɬny ɬí:ni “see”. Note that ML gives 3 w-final Gb-Stems, only (√glw gí:ləw, √ktw kí:təw “(darkness) fall” and √s'nw s'áynəw “be deaf”).

The overwhelming neutralization of the contrast w vs y in final position led Johnstone to transcribe the final glide of glide-final roots with the symbol “v”. However, based on Semitic comparative data, he sometimes distinguishes between y- and w-final roots. In this article, I will use the symbol Y to note a final glide and give the root as it stands in ML s.v.

		pf 3ms	*ipf 3ms*	*sbj 3ms*	
i.	√wgɬ	šəwgú:ɬ	yəšəwgú:ɬ	yəšáwgəɬ	"go in the early evening"
		šu:gú:ɬ$_{FW}$	yəšgó:ɬ$_{FW}$	yəšáwgəɬ$_{FW}$	
ii.	√wkb	həwkú:b	yəhəwkú:b	yəháwkəb	"put in"
		hu:kú:b$_{FW}$	yəhkó:b$_{FW}$	yəháwkəb$_{FW}$	
iii.	√wkf	šəwkú:f	yəšəwkú:f	yəšákf	"sleep"
		šu:kú:f$_{FW}$	yəškó:f$_{FW}$	yəšákf$_{FW}$	

These forms present two peculiarities:

(a) root-initial w is dropped in the *ipf* (i,ii) or both in the *ipf* and the *sbj* (iii).

(b) if w is dropped, the *ipf* vocalization is ó: and the *sbj* vocalization is ə́.

The parameter that conditions this loss of w and the vowel alternations are not clear to me at the present stage: more data are needed.[5]

3.3. Glide-final roots

Glide-final roots do not pattern with sound roots. Two representative examples of the H-Stem and of the š$_1$-Stem of √CCY appear in (10) and (11) respectively.[6]

(10) CCY, H-Stem

		pf 3ms	*ipf 3ms*	*sbj 3ms*	*cd 3ms*	
A	√bnv	həbnó:	yəhəbáyn	yəhí:bən, yəhé:bən	yəhámnən, yəhámnən	"build"
B	√nħv	hənħó:	yəhənáyħ	yəhánħ yəhánħ$_{MT\ 20:61}$	yəhánħən	"burn"

[5] On the basis of the contrast i, ii vs iii in the sbj, one may conjecture that the identity of R2 and R3 is relevant, and that the diphtong áw is reduced to ə́ because of the presence of two voiceless non ejective consonants to its right, i.e. the "idle glottis" effect described in section 4.1 is at work here. This analysis does not extend to the ipf.

[6] The set of the relevant forms based on an exhaustive count of the Mehri Lexicon is given in the Appendix.

(11) CCY, $š_1$-Stem

		pf 3ms	*ipf 3ms*	*sbj 3ms*	*cd 3ms*	
A	√fk'w	šəfk'ó:	yəšfáyk'	yəšé:fək', yəší:fək'$_{FW}$	yəšə́fk'ən	"cover one self up"
B	√s'fw	šəs'fó:	yəšs'áyf	yəšás'f yəšə́s'f$_{MT, FW}$		"gather news, find out"

It appears that glide-final H- and $š_1$-Stems do not behave uniformly with respect to the subjunctive stem: two patterns must be distinguished, A and B.

Type B has not been explicitly recognized as such until now. Johnstone, in the introduction of the *Mehri Lexicon* (*ML*: xlv) gives type A as representative of the "causative verb final semi-vowel", *i.e.* of H-Stem √CCY. He mentions type B for "causative verb initial w and final v" (*ML*: xlvi), thus suggesting that it involves √YCY roots only, which is not the case (see (10) and (11)). For the $š_1$-Stem, type B is not mentioned at all in the *Mehri Lexicon*. Rubin (2010: 159) does not give examples of type B.

Type B however is not marginal, as evidenced in the table below.[7] It thus needs to be recognized as such.

	type A	type B
H-Stem	21	14
$š_1$-Stem	18	4
Σ	39	18

Focusing on the data in (10) and (11), we may make the following observations. In the perfective and in the imperfective, type A and type B behave alike. In the *pf 3ms*, the stem-vowel is [ó:] (often realized as [óh]) and the root-final glide does not surface as such.[8] In the *ipf 3ms*, the final glide is not realized as such either, and a diphthong surfaces between R_1 and R_2. A comparison of the "sound" stems in (12a) with the glide-final stems suggests the rules in (12b):

[7] This count is based on an exhaustive examination of the H- and š1-Stems for which ML gives the sbj 3ms.

[8] *ML* gives both *ú:* and *ó:* for the *pf 3ms* of Y-final roots. However it is clear from fieldwork as well as from *MT* that the vowel is [ó:].

(12)	√	*pf 3ms*	*ipf 3ms*
a.	CCC	həCCú:C	yə-həCCú:C
b.	CC**Y**	*həCC**ú:Y** → həCCó:	*yə-həCC**ú:Y** → yə-həCáyC

(12b) suggests a ban on the sequence ú:Y. This ban is confirmed by other morpho-phonological processes in the language.

Consider for instance the Ga-Stem of guttural-final roots. The perfective of such roots generally has its stem-vowel between R_1 and R_2, not between R_2 and R_3 (13b). There is a systematic exception to this generalization: the perfective of √CYG systematically behaves differently: it has its stem-vowel located between R_2 and R_3 (13c).[9] Considering now the imperfective of the same forms, we observe that the pattern of √CYG systematically deviates from that of √CCG, too: instead of expected yəCú:YəG, we observe yəCYó:G. The exceptional behaviour of the Ga-Stem of √CYG both in the perfective and in the imperfective may be attributed to a ban on ú:Y: *pf* *Cú:YəG → CəYó:G and *ipf* *yəCú:YəG → yəCYó:G.

(13)	Ga-Stem		*pf 3ms*	*ipf 3ms*	
a.	CCC	√rkz	rəkú:z	yərú:kəz	"straighten"
b.	CCG	√blh	bú:ləh	yəbú:ləh	"keep so. off his work"
c.	CYG	√dyχ	*d**ú:y**əχ dəyó:χ	*yəd**ú:y**əχ yədyó:χ	"be, become dizzy"

Coming back to the H- and $š_1$-Stems of glide-final roots in (10) and (11), we observe that in the subjunctive, two classes of verbs must be distinguished: type A and type B.

As for the identity of the stem-vowel in type A, the data in the *Mehri Lexicon* are contradictory. For instance, for √bnv, *ML*: xlv has *cd* yəhámnən, which suggests /í:/, but *ML s.v.* has *cd* yəhámnən, which suggests /é:/ (see (6)). Such a variation is attested in most forms (*e.g.* √tlv *s.v.*, √wly *s.v.*) and prevents from safely concluding at the present stage. However both *MT* and my own preliminary fieldwork data have í: for type A and á for type B:[10]

[9] Out of the 101 Ga-Stem *pf 3ms* of √CCG listed in the *Mehri Lexicon*, 83 have the stressed vowel between R_1 and R_2, 14 have it between R_2 and R_3, and 4 have both forms as variants. Ga-Stem *pf 3ms* of √CYG all have the stressed vowel between R_2 and R_3.

[10] See appendix. These data need to be systematically extended and checked with a native speaker. However this issue is not directly relevant for my purpose here. Note furthermore that √YCY has á in the *sbj.* (I am grateful to Aaron Rubin for pointing out this fact to me.)

	type A	has	*sbj*	-h/š í: R_1 ə R_2
and	type B	has	*sbj*	-h/š á R_1 R_2

This means that type A and type B both have the same underlying vowel: í: surfaces as á in a closed syllable (6). The only difference between type A and type B is that R_1 and R_2 are separated by schwa in type A while they are not in type B. This is good indication that both types may be derived one from the other.

In order to gain insight into these forms, let us compare them with the forms that we expect on the basis of the rules of the language. In Mehri, a glide in coda position is vocalized [i:] word-internally and [i] word-finally: əy# → [i]# (see Bendjaballah & Ségéral in press). This rule applies for instance in the perfective of the a-Stem: (14a). For the H- and $š_1$-Stems, we would expect the same rule to apply, that is, the subjunctive of √CCY should be -h/šáCCi:

(14) a. a-Stem

	pf 3ms	*example*	
√CCC	aCó:CəC	√rkb	aró:kəb
√CCY	aCó:CəY → aCó:Ci	√msy	amó:si "kiss(b-)so."

b. H/$š_1$-Stem

	sbj 3ms	*example*	
√CCC	-h/šáCCəC	√nsm	-hánsəm
√CCY	-h/šáCCəY → *-h/šáCCi	√bnv	*-hábni, -hí:bən
		√nħv	*-hánħi, -hánħ

Instead of *yəh/šáCCi, type A has yəh/ší:CəC and type B has yəh/šáCC. Both types are characterized by (a) an Umlaut of the stressed vowel, and (b) the fact that the trigger of Umlaut, the final glide, does not surface as expected as i (*i.e.* it is « latent »). Both phenomena are well-attested in Mehri. For example, the *sbj 2fs* of the H-Stem also instantiates the rule /á...i/ → ó: √nsm H-Stem *sbj 2fs* = / tə-hánsəm-i / → təhánsəm (*cf.* Bendjaballah & Ségéral 2015 for analysis).

The first question to answer is about the causality of the distinction between type A and type B: is it phonologically conditioned? If no, we have to assume two different underlying structures for the subjunctive, hence two groups of Y-final H- and $š_1$-Stems. If yes, then we can conclude in favor of a single underlying group.

4. Exploring the phonological conditioning of the divide between type A and type B

The aim of this section is to explore the hypothesis that the divide type A *vs* type B is phonologically conditioned. The question is whether the two consonants involved in the final CəC# (type A) and CC# (type B) differ in any predictable way.

I will proceed as follows: I will examine the identity of the consonants involved in the two subjunctive patterns, and I will systematically compare them with those involved in comparable nominal patterns. We will then be in a position to decide whether the final CC-clusters involved in the CC#-subjunctives are identical with those involved in CC#-nouns, as opposed to those in CəC#-subjunctive and nouns. If yes, this will be a good indication that some phonological – not morphological – conditioning is at work in the divide between type A and type B.

(15) and (16) below give the complete set of the first two root-consonants involved in type A and type B respectively. The exhaustive list of the corresponding verbs along with their glosses are given in the Appendix.

(15) Type A: *sbj* $R_1əR_2$ #

lab		*dent-alv*							*lat*		*vel*				*uvul*		*phar*		*lar*
bn	fk'	ml	ðr	dr	tl	sw	nw	rw	ɬn	lw	gr	kd	k'w	wl	ʁd	χt'	ħg	ʕn	hm
bd				dʕ	tw								k'r		ʁl	χw		ʕf	hw
br				dħ														ʕs'	hd
																		ʕɬ'	

(16) Type B: *sbj* R_1R_2 #

bʕ	ft	t'f	s'f	nħ	rd	ɬt	gf	wf					
bk'					rs			wħ					
					rɬ'			wz					
					rχ								
lab	*dent-alv*					*lat*	*vel*		*uvul*		*phar*		*lar*

Three generalizations involving three natural classes of consonants must be made. The first one involves the class of [-voice, -ejective] consonants, the second one the gutturals, and the third one the sonorants. I address them in the following sections.

4.1. Generalization 1: the "idle glottis" effect

Bendjaballah & Ségéral (2014) show that in Mehri two voiceless non ejective consonants (called "idle glottis" consonants) may never be separated by an unstressed schwa:

(17) *©ə©, where © = [-voice, -ejective]

As a consequence, √©©Y is expected to exclude the A-pattern (sbj *©ə©#) and select for the B-pattern (sbj ©©#). This is indeed the case: there is not a single √©©Y following the A-pattern (see (15)). All √©©Y (2 tokens) select for the B-pattern:

(18) √©©Y systematically selects for type B.

	pf 3ms	*sbj 3ms*
√ftw	həftó:	yəháft
√ɬtw	həɬtú:	yəháɬt

These verbs can thus automatically be filtered out, since the selection of the B-pattern is phonologically conditioned.

4.2. Generalization 2: Gutturals spread schwa to their right

A comparison of (15) and (16) reveals that √GCY selects for the A-pattern: no guttural-initial root is attested in the B-pattern:[11]

[11] All √ʕCY have *sbj* yəh/šé:C (see appendix 1), *e.g.* √ʕny *pf 3ms* ha:nó:, *sbj 3ms* yəhé:n "decide, intend" This form is difficult to interpret. I assume that ʕ behaves on a par with the other gutturals, and expect this form to be derived from the A-pattern: /yəhí:ʕən / > yəhé:n.

(19) √GCY systematically selects for type A.

		pf 3ms	*sbj 3ms*	*cd 3ms*
χ	√χt'v	χt'ú:	yəhé:χət'	
	√χt'v	χt'ú:	yəhí:χət'	yəháχt'ən
	√χt'ʔ	šəχt'ú:	yəšé:χət', yəší:χət'$_{FW}$	yəšáχt'ən
	√χwy	χwú:	yəhé:χəw	
ʁ	√ʁdv	həʁdú:	yəhí:ʁəd	yəháʁdən
	√ʁlv	həʁló:	yəhé:ʁəl	
	√ʁly	šəʁló:	yəšé:ʁəl, yəší:ʁəl$_{FW}$	
ħ	√ħgw	šəħgú:	yəšé:ħəg, yəší:ħəg$_{FW}$	yəšáħgən
h	√hmv	hmú:	yəhí:həm	yəháhmən
	√hdy	šhədú:	yəšé:həd, yəší:həd$_{FW}$	
	√hwy	hwú:	yəhé:həw	

The fact that *sbj* GC# is excluded and *sbj* GəC# is mandatory indicates that, in monomorphemic contexts, gutturals before C# must be followed by schwa. This generalization must be checked against other morphological classes, *e.g.* nouns. We expect CVGC-nouns to be excluded. Since a guttural is expected to be followed by schwa in that context, we expect to find CV:GəC instead.

The *Mehri Lexicon* lists a total of 178 CVCC nouns. 9 of them derive from √CGC. If we compare this number with the number of CV:GəC nouns, we observe that they are proportionally more numerous: 29 for a total of 217 CV:CəC nouns. Finally, the *Mehri Lexicon* gives 6 cases with 2 variants: C_1VGC_3 and $C_1VGəC_3$ (20c).

(20)

		total	$C_2 \neq G$	$C_2 = G$	
a.	$C_1VC_2C_3$	178	169	9	5%
b.	$C_1V:C_2əC_3$	217	188	29	13,3%
c.	C_1VGC_3 ~ $C_1VGəC_3$	6			

The 9 exceptional CVGC nouns appear in (21). Two of them (21.1 and 21.2) may be excluded on independent ground: (17) prevents schwa from breaking the ©©# cluster. We are left with the counter-examples in (21.3-9). Two of them are Arabic words. That is to say, based on Johnstone transcription, there are 5 true counter-examples: (21.5-9).

(21)	1.	√nħf	náħf	“slice of wood, stone, bread”	(√C©©)
	2.	√nħs	náħs	Ar. “ill fortune”	(√C©©)
	3.	√rħm	ráħm	“womb”	(Ar)
	4.	√θ'hr	θ'áhr	“noon, midday”	(Ar)
	5.	√bχs'	báχs'	S “pain”	
	6.	√fʁɬ	fáʁɬ	“plentiful grazing, pasture”	
	7.	√šħk'	šáħk'	“ebb-tide”	
	8.	√shm	sə́hm	“swell (at the sea)”	
	9.	√ðhl	ðáhl	“the person next to me (m)”	

As for the six examples with a variant with schwa (20c), we observe that they behave in an anomalous way: they exhibit a *short* stressed vowel in an open syllable. In that context we would expect a *long* vowel to surface (see section 2). Five of them are Arabic words, and for one of them, šáʁəl in (22.3), the only example in *MT* is in an Arabic phrase (Aaron Rubin p.c.) I thus consider these items not to be ‘true’ counter-examples.

(22)	1.	√dhk'	dáhk', dɛ́həkʼ$_{MT}$	“precipice, slope”	
	2.	√rχs'	bə-ráχs', bə-ráχəsʼ$_{MT}$	“cheaply”	(Ar)
	3.	√šʁl	šáʁl, šáʁəl$_{MT}$	“way of acting; way of doing, preparing st., work”	(Ar)
	4.	√shm	sáhm, sɛ́həm	“arrow”	(Ar)
	5.	√sħt'	sáħt', sɛ́ħətʼ$_{MT}$	“slaughter”	(Ar)
	6.	√sχt'	sáχt', sáχət'	“matches”	(Ar)

These marginal exceptions should be checked with a native speaker in order to firmly establish their correct transcription and their status in Mehri.

Admittedly, the difference between (20a) and (20b) is not large. However, the fact remains that CVGC nouns are rare. I thus conclude that a guttural may not be the first consonant of a monomorphemic final CC-cluster: *GC#.[12] In that configuration, a guttural spreads a schwa to its right: *GC# → GəC#.

4.3. Generalization 3: Sonorants need a vocalic support to their left

Finally, the distribution of type A and type B in function of R_2 reveals a robust generalization: R_2 = [+son] (*i.e.* m n w y l r) systematically selects for type A.

The distribution tables are given in (23) and (24), and the relevant examples in (25).

(23) Type A: *sbj* R_1əR_2 #

R_2 = [+sonorant]					R_2 = [+plosive]			R_2 = [+fricative]		
bn	hm	ml	ðr	sw	kd	χt’	ħg	ʕf	dħ	dʕ
ɬn		tl	dr	nw	bd					
ʕn		wl	gr	rw	ʁd					
		ʁl	br	lw	hd					
			k’r	hw						
				k’w						
				tw						
				χw						

(24) Type B: *sbj* R_1R_2 #

					rd	ft	bk'	t'f	rs	wz	rχ	nħ	rɬ'	bʕ
						ɬt		s'f				wħ		
								gf						
								wf						
R_2 = [+sonorant]					R_2 = [+plosive]			R_2 = [+fricative]						

[12] If G = ©, and final C = ©, see section 4.1.

(25) √CSY systematically selects for type A.

		pf 3ms	*sbj 3ms*	*cd 3ms*
n	√bnv	həbnú:	yəhí:bən	yəhámnən
	√ɬny	ɬnú:	yəhé:ɬən, yəhí:ɬən$_{MT}$	yəháɬnən
		šəɬnú:	yəšé:ɬən, yəší:ɬən$_{FW}$	
m	√hmy	hmú:	yəhí:həm	yəháhmən
l	√ʁly	həʁló:	yəhé:ʁəl	
		šəʁló:	yəšé:ʁəl, yəší:ʁəl$_{FW}$	
	√ml?	mlú:	yəhí:məl	
	√tlv	tlú:	yəhé:təl	yehátlən
	√wly	həwlú:	yəhé:wəl, yəhí:wəl$_{MT}$	yəháwlən
r	√drw	hədró:	yəhé:dər	
		šədró:	yəšé:dər, yəší:dər$_{FW}$	
	√ðry	həðró:	yəhé:ðər	
	√gry	həgró:	yəhé:gər	
	√brʕ	šəbró:	yəšé:bər	
	√k'rv	šək'ró:, šək'rú:	yəší:k'ər, yəší:k'ər$_{FW}$	yəšák'rən
w	√hwy	hwú:	yəhé:həw	
	√k'wy	hək'wú:	yəhí:k'əw	yəhák'wən
	√nwy	hənwó:	yəhé:nəw	
	√rwy	hərwú:	yəhé:rəw	yəhárwən
	√twy	twú:	yəhé:təw	
	√χwy	χwú:	yəhé:χəw	
	√lwy	šəlwú:	šé:ləw$_{imp.sg}$	
	√swy	šəsú:	yəšɛ:səʔ	

Sbj CS# is excluded and *sbj* CəS# is mandatory. This means that sonorants must be preceded by schwa: Accordingly, we expect to find no CVCS-noun, and to find CV:CəS instead.

In the class of CVCC nouns, nouns with a final sonorant are largely underrepresented: 9/178. (See also Dufour 2016). By contrast, they are in the majority in the class of CV:CəC nouns (161/217). Finally *ML* gives 4 nouns with two variants: CVCS and CVCəS.

(26)

		total	$C_3 \neq S$	$C_3 = S$	
a.	$C_1VC_2C_3$	178	169	9	5%
b.	$C_1V{:}C_2əC_3$	217	56	161	74.2%
c.	$C_1VC_2S \sim C_1VC_2əS$	4			

The 9 exceptional CVCS nouns are listed in (27), the 4 CVCS nouns with a CVCəS variant appear in (28). These exceptions overlap with the exceptions listed above in section 4.2 because their behavior with respect to the guttural in C_2 is exceptional, too: (27a) and (28a). We are left with 7 additional exceptions (27b, 28b), 5 of which are clearly Arabic loans.

(27)

a.	CVGS	√θ'hr	θ'áhr	"noon, midday"	(Ar)
		√rħm	ráħm	"womb"	(Ar)
		√shm	sə́hm	"swell (at the sea)"	
		√ðhl	ðáhl	"the person next to me (m)"	
b.	CVCS	√χs'm	χás'm	"enemy"	(Ar)
		√χθm	χáθm	"sickly (livestock)"	
		√ħbr	ħábr	"ink"	(Ar)
		√gbn	gə́bn	"cheese"	(Ar)
		√ʔlm	ʔə́lm	"knowledge"	(Ar)

(28)

a.	CVGS	√shm	sáhm, sɛ́həm	"arrow"	(Ar)
		√šʁl	šáʁl, šáʁəl, šə́ʁəl$_{MT}$	"way of acting ; way of preparing st."	(Ar)
b.	CVCS	√ʔs'l	ʔás'l, ʔás'əl	"origin"	(Ar)
		√ħɬm	ħáɬm, ħáɬəm	"phlegm ; sticky excrescence on trees"	

I thus conclude that *CS# → CəS#. This rule is to be compared to a phenomenon that can be observed at the beginning of the word: #SC → #əSC.[13] Both processes indeed can be unified: a sonorant needs a vocalic support to its left.

To conclude: three phonological generalizations make it possible to predict whether a given H/$š_1$-Stem belongs to type A or to type B:

(29)

a. *©ə©	where © = [- voice, - ejective]	*i.e.*	= f θ s t ɬ š k χ ħ h
b. *GC# → GəC#	where G = [+ guttural]	*i.e.*	= χ ħ h ʁ ʕ
c. *CS# → CəS#	where S = [+sonorant]	*i.e.*	= m n w y l r

5. Discussion and conclusion

The phonological analysis based on the three generalizations in (29) made it possible to filter out 2/3 of the H- and $š_1$-Stems listed in the *Mehri Lexicon*. The remaining 18 CCY roots, for which we have no rule predicting their subjunctive shape, appear in (30) and (31), sorted out according to the identity of R_2 and of R_1 respectively. These forms need to be systematically checked with a native speaker: in some cases indeed, my informant produces a *sbj* form that does not match the one given in the *Mehri Lexicon.* For instance, he chooses the B-pattern for $š_1$-Stem √bdy *pf 3ms* šəbdóh, *sbj 3ms* yəšə́bd "believe so. is lying" where *ML* has the A-pattern. Further fieldwork may reveal additional phonological generalizations and an evolution towards a single class (and a phonological conditioning) for Y-final H- and $š_1$-Stems.

(30)

R_2 =	f	d	z	s	ɬ’	k’	χ	ʕ	ħ
A		kdY, bdY				fk’Y		dʕY	dħY
B	wfY, gfY, s’fY, t’fY	rdY	wzY	rsY	rɬ’Y	bk’Y	rχY	bʕY	nħY, wħY

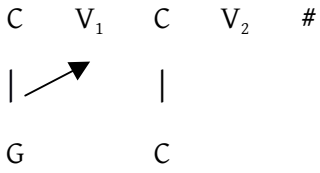
(31)

R_1 =	b	f	d	n	r	s’	t’	k	g	w
A	bdY	fk’Y	dʕY, dħY					kdY		
B	bk’Y, bʕY			nħY	rdY, rsY, rɬ’Y, rχY	s’fY	t’fY		gfY	wfY, wzY, wħY

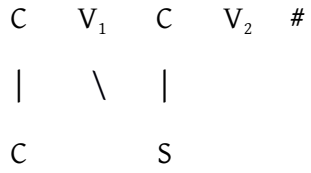
[13] In the Ga-Stem, Johnstone (*ML*) notes the phenomenon for l (*e.g.* √lbd əwbú:d "shoot, strike"), n (*e.g.* √nfz nəfú:z n- "(stone) cut so's foot"), m (√mry mərú: əmrúh "play with a camel's teats till milk comes") and r (*e.g.* √rsm rəsú:m, ɛrsú:m "answer readily"). See also Watson (2012: 43).

At a more general level, the examination of the Mehri data made it possible to establish two generalizations that pertain to the phonological representation of gutturals and sonorants (29b and c). These rules largely correspond to the behaviour of gutturals and glides cross-linguistically: gutturals have an influence on their adjacent vocalic positions, and sonorants need a vocalic support to their left. The interesting aspect is that in Mehri these processes take place word-finally, only: forms like yəšáħgəm $š_1$-Stem *sbj 3ms* √ħgm "(animals) be put in a safe place" and yəháflət H-Stem *sbj 3ms* √flt "escape, flee" are perfectly grammatical. The situation word-finally and word-initially is represented in the CV-framework (Lowenstamm 1996) of Government Phonology (Kaye, Lowenstamm & Vergnaud 1990) in (32a-b) and (32c-d) respectively.

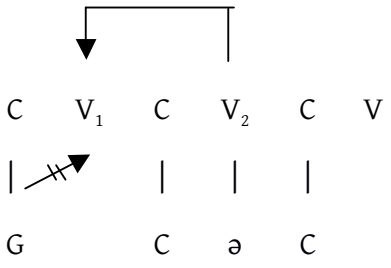
(32) a. *GC# → GəC#



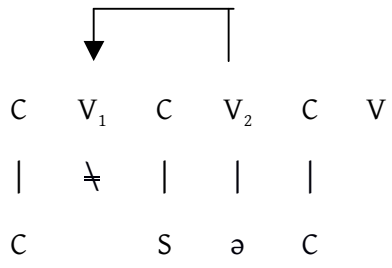
b. *CS# → CəS#



c. GCVC



d. CSVC



The contrast (32a-b) *vs* (32c-d) can readily be derived: in (32c-d), V_2 is phonetically identified, hence it properly governs V_1. As a consequence V_1 may remain phonetically empty. By contrast, in (32a-b), assuming that empty final nuclei may not be governors in Mehri, V_1 is not properly governed. It must be identified, and propagation from the neighbouring consonants takes place. Consider now a salient property of CVC#. Contrarily to what happens word-internally, a stressed vowel in final CVC syllable is lengthened: CVVC#. This indicates that the final empty nucleus is a licensor – *cf*. Bendjaballah & Ségéral (in press). A more fine-grained description of the properties of final CC clusters in Mehri is expected to provide a clearer picture of the ability of final empty nuclei to govern/license.

BIBLIOGRAPHICAL REFERENCES

Bendjaballah, S., Ségéral, P. (2014), *The Phonology of «Idle Glottis» Consonants in the Mehri of Oman (Modern South Arabian),* in "Journal of Semitic Studies" 59, 1: 161-204

Bendjaballah, S., Ségéral, P. (2015), *De l'Umlaut dans le système verbal du mehri d'Oman*, 3è journée d'études sudarabiques modernes, Univ. de Strasbourg, 9 octobre 2015

Bendjaballah, S., Ségéral, P. (in press), *The vocalic system of the Mehri of Oman: stress, length and syllabic structure*, in "Brill's Journal of Afroasiatic Languages and Linguistics" 9.1-2: 160-190

Dufour, J. (2016), *Recherches sur le verbe sudarabique moderne*, Habilitation à Diriger des Recherches, Paris, EPHE

Dufour, J. (in press), *Nouns and adjectives of the shape $C_1\acute{V}C_2(ə)C_3(-)$ in Jibbali (Śḥri) and Mehri,* in "Brill's Journal of Afroasiatic Languages and Linguistics" 9.1-2: 191-217

Kaye, J., Lowenstamm, J., Vergnaud, J-R. (1990), *Constituent structure and government in phonology,* in "Phonology" 7-2: 193-231

Lowenstamm, J. (1996), *CV as the only syllable type*, in J. Durand & B. Laks (eds), *Current Trends in Phonology: Models and Methods*, Manchester, ESRI: 419-442

ML = Johnstone, T.M. (1987), *Mehri Lexicon and English-Mehri Word-List,* with *Index of the English Definitions in the Jibbali Lexicon, compiled by G. Rex Smith*, London, SOAS

MT = Stroomer, H (1999), *Mehri Texts from Oman. Based on the Field Materials of T.M. Johnstone*, Wiesbaden, Harrassowitz

Rubin, A. (2010), *The Mehri Language of Oman*, Leiden / Boston, Brill

Watson, J. (2012), *The Structure of Mehri*, Wiesbaden, Harrassowitz

APPENDIX

In the tables to follow,

- for each verb, the root is the one given in the *Mehri Lexicon*. In some cases, I analyse a verb given in the *Mehri Lexicon s.v.* under a ʔ or ʕ-final root as deriving from a glide-final root (*e.g.* 9 and 20). This results from the analysis of the *ipf*, *sbj* and *cd* forms. If a verb has an ambivalent behaviour (*e.g.* 18), I leave both roots in the first column.
- I adopt the following abbreviations from the *Mehri Lexicon*: os. = oneself, so. = someone, st. = something

1. A-PATTERN

1.1. H-STEM

	√	*pf 3ms*	*ipf 3ms*	*sbj 3ms*	*cd 3ms*	*gloss*
1.	bnv	həbnó:, həbnú:	yəhəbáyn	yəhé:bən, yəhí:bən	yəhámnən, -hɛ́:-, yəhɛ́mnən	build, put up, erect
2.	drw	hədró:	yəhədáyr	yəhé:dər		climb right to the top
3.	ðry	həðró:	yəhəðáyr	yəhé:ðər		(lə-) let the blood of a goat run over a sick person
4.	fk'w	fk'ó:	yəfáyk'	yəhé:fək'	yəháfk'ən	cover (l-) so. (b-) with st.
5.	gry	həgró:	yəhəgáyr	yəhé:gər		put forward, advance
6.	ʁdv	həʁdó:, həʁdú:	yəhəʁáyd	yəhé:ʁəd yəhí:ʁəd	yəhə́ʁdən	forget, lose
7.	ʁly	həʁló:	yəhəʁáyl	yəhé:ʁəl		put (milk) on to boil; sell dear
8.	hmy	hmú:	yəháym	yəhé:həm yəhí:həm	yəhə́hmən	call, name
9.	mlʔ	mlú:	yəhəmáyl	yəhí:məl		fill
10.	ɬny	ɬnú:	yəɬáyn	yəhé:ɬən	yəháɬnən	show; be a medicine-man
11.	tlv	tlú:	yətáyl	yəhé:təl	yehə́tlən	regret, be sorry, repent
12.	wly	həwlú:	yəhwáyl	yəhé:wəl	yəhə́wlən	turn back, go back to, come back, direct os. to
13.	χwy	χwú:	yəχáyw	yəhé:χəw		(us. a woman to a man) send, or cause to be sent, a confidential or secret message (us. in the earliest stages of marriage negotiations); expose o's; be exposed
14.	hwy	hwú:	yəháyw	yəhé:həw		make fall
15.	k'wy	hək'wú:	yəhək'áyw	yəhí:k'əw	yəhə́k'wən	build up (a fire); strengthen
16.	nwy	hənwó:	yəhənáyw	yəhé:nəw		intend, decide (to do st.)
17.	rwy	hərwú:	yəhəráyw	yəhé:rəw	yəhárwən	give to drink
18.	kdv kdʔ	kdú:	yəkáyd -kú:da	yəhé:kəd		make (a camel) trot, race
19.	χt'w	χt'ú:	yəχt'ú:, yəχáyt'	yəhé:χət', yəhí:χət'	yəhə́χt'ən	take clothing for sewing
20.	χt'ʔ	χt'ú: ∞həχt'ó:	yəχáyt'	yəhé:χət'		hit, injure by mistake [χt'ʔ see also χt'w]
21.	twy	twú:	yətwú:	yəhé:təw		cause to eat
22.	ʕny	ha:nó:	yəháyn	yəhɛ́:n	yəhɛ́:nən	decide, intend

1.2. Š1-STEM

	√	*pf*	*ipf*	*sbj*	*cd*	*gloss*
1.	bdy	šəbdó:	yəšbáyd	yəšé:bəd yəšbádyəm$_{3mp}$ yəšəbd$_{FW}$		believe so. is lying
2.	brʕ	šəbró:	yəšbáyr	yəšé:bər		be free, freed from; not do st.
3.	dħw	šədħó:	yəšədáyħ	yəšé:dəħ		come to after a faint, regain consciousness
4.	drw	šədró:	yəšədáyr	yəšé:dər, yəší:dər$_{FW}$		get to the top
5.	fk'w	šəfk'ó:, šəfk'ú:	yəšfáyk'	yəšé:fək', yəší:fək', yəší:fək'$_{FW}$	yəšəfk'ən	cover os. up
6.	ʁly	šəʁló:	yəšəʁáyl	yəšé:ʁəl yəší:ʁəl$_{FW}$		buy dear, at a high price
7.	hdy	šhədú:	yəšháyd	yəšé:həd, yəší:həd$_{FW}$		find, come across (l-) st.
8.	ħgw	šəħgú:	yəšháyg	yəšé:ħəg, yəší:ħəg$_{FW}$	yəšáħgən	stand firm; settle a difficulty
9.	k'rv	šək'ró:, šək'rú:	yəšk'áyr	yəší:k'ər, yəší:k'ər$_{FW}$	yəšək'rən	be hidden, disappear, hide os.
10.	lwy	šəlwú:, šəlwúh		šé:ləw$_{imp.sg}$		make a turn-off
11.	swy	šəsú:, šsúh šsóh-əh$_{3ms+3ms}$	yəšsé:, -é?	yəšɛ́:səʔ		speak to the authorities urging the release of so. imprisoned
12.	ɬny	šəɬnú:	yəšɬáyn	yəšé:ɬən, yəší:ɬən$_{FW}$		be seen
13.	χt'ʔ	šəχt'ú:	yəšχáyt'	yəšé:χət', yəší:χət'$_{FW}$	yəšáχt'ən	be injured by mistake, inadvertently
14.	dʃw	šó:da, šú:da	yəšó:da, yəšú:da	yəšé:da	yəšadan	curse, insult, miscall so (b-)
16.	ʕfw	ša:fó:	yəšáyf	yəšá:f -šɛ́:f		improve in health, get better, recover
17.	ʕs'v	ša:s'ó:	yəšáys'	yəšɛ́:s'		become afraid for so.
18.	ʕɬ'v	ša:ɬ'ó: -óh	yəšáyɬ'	yəšɛ́:ɬ'	yəšɛ́:ɬ'ən	worry about, be upset or worried about (as, *e.g.*, so. coming late)

2. B-PATTERN

2.1. H-STEM

	√	*pf*	*ipf*	*sbj*	*cd*	*gloss*
1.	bʃw	həbó:	yəhábi	yəhábi		caress an animal's teats (with the fingers or gentle words) to coax it into giving milk
2.	bk'y	həbk'ó:	yəhəbáyk'	yəhábk'		leave, put aside
3.	ftw	həftó:	yəhəfáyt	yəháft		give (h-) so. an idea or plan; concentrate; decide on one of several choices; decide definitively on st.
4.	gfy	həgfó:	yəhəgáyf	yəhágf	yəhágfən	throw a shadow
5.	nħv	hənħú:	yəhənáyħ	yəhánħ	yəhánħən	burn
6.	rdw	hərdú:	yəhəráyd	yəhárd		leave, save (food)
7.	rsw	hərsú:	yəhəráys	yəhárs	yəhársən	anchor; stay in one place, in port, lie at anchor
8.	rɬ'y	hərɬ'ú:	yəhəráyɬ'	yəhárɬ'		be reconciled with, bring back o's wife after a quarrel; get restoration of conjugal rights
9.	rχv	hərχú:	yəhəré:χ	yəhárχ		release, slacken (l-) st.
10.	ɬtw	həɬtú:	yəhəɬáyt	yəháɬt		transhume in winter
11.	t'fv	hət'fú:	yəhət'áyf	yəhát'f		extinguish, put out (a light, a fire)
12.	wfy	həwfú:	yəhwáyf	yəháwf		settle, pay in full
13.	wħy	həwħú:	yəhwáyħ	yəháwħ		come to help; call, appeal for help to so.
14.	wzy	həwzú:	yəháyz yəhəwáyz	yəháwz	yəháwzən	put aside

2.2. Š1-STEM

	√	*pf*	*ipf*	*sbj*	*cd*	*gloss*
1.	nħv	šənħú:	yəšənáyħ	yəšánħ		get burnt
2.	s'fw	šəs'fú:	yəšs'áyf	yəšás'f, yəšəs'f_{FW}		gather news, find out
3.	wfy	šəwfú:	yəšwáyf	yəšáwf	yəšáwfən	be paid in full, be revenged for (b-) so.
4.	wzy	šəwzú:	yəšwáyz	yəšáwz	yəšáwzən	bend back over

SABRINA BENDJABALLAH • Directrice de Recherche at CNRS, Director of UMR 6310-LLING, *Laboratoire de Linguistique de Nantes*, CNRS & Université de Nantes; phonology and morphology of Afroasiatic languages, in particular Berber, Somali (Cushitic) and Mehri (Semitic); recent publications: Bendjaballah & Ségéral, 2017. On the verb forms derived from four h-initial roots in the Mehri language of Oman. *Journal of Semitic Studies* LXII/1:199-215, Barillot, Bendjaballah & Lampitelli, to appear. Verbal classes in Somali, Allomorphy has no classificatory function, *Journal of Linguistics*, Bendjaballah, Ben Si Saïd & Haiden, 2015. Le syntagme prépositionnel à la périphérie gauche en Taqbaylit. *Corpus* 14: 263-280.

E-MAIL • sabrina.bendjaballah@univ-nantes.fr

LA MORPHOLOGIE DES FORMES VERBALES SIMPLES EN SUDARABIQUE MODERNE

Hypothèse diachronique

Julien DUFOUR

ABSTRACT • This article explores the phonologically-triggered morphological suppletism exhibited by the basic-stem verbal forms of triliteral roots in the Modern South Arabian languages, concentrating on the default patterns and the patterns selected by roots with a guttural consonant in C_2 or C_3 position. It aims at understanding the diachrony of the system in a comparative perspective. It will be claimed that proto-Modern South Arabian developed word stress according to two constraints: (1) stress favours vowels closer to the last consonant of the word; (2) stress location is sensitive to vowel quality according to a sonority scale *a > *ă > *i, *u (where *ă stands for '*a before a guttural'). This historically phonological process gave rise to what is now synchronically a morphological system of allomorphy.

KEYWORDS • Modern South Arabian, verbal morphology, historical phonology, quality sensitive stress, guttural consonants, Jibbali, Mehri, Soqotri.

1. Introduction

La classification des langues sémitiques la plus communément admise aujourd'hui est sans doute celle présentée dans Huehnergard & Rubin (2011). Acceptant comme primordiale la division est-ouest de la famille, et tout en reconnaissant l'importance de phénomènes aréaux, elle considère le sudarabique dit « moderne » comme un rameau autonome à l'intérieur du sémitique de l'ouest, à côté d'un sémitique central et d'un sémitique d'Éthiopie. Il est vrai que cette classification interne du sémitique de l'ouest repose presque uniquement sur la morphologie de l'inaccompli, et qu'il suffirait que ce pilier s'avère friable pour qu'elle soit remise en cause (cf., par exemple, Lonnet 2017). Cependant un auteur comme Kogan (2015b), qui, tout en tenant compte de la morphologie, centre son étude sur le lexique, arrive lui aussi, mais par une autre voie, à la conclusion que le sudarabique moderne est un rameau autonome: au sein d'un sémitique de l'ouest dont l'unité phylogénétique lui paraît difficile à prouver, il incline à reconnaître une proximité particulière entre sémitique central et sémitique d'Éthiopie face au sudarabique moderne.

Le corollaire de telles conclusions est que, d'un point de vue comparatiste, la leçon du sudarabique moderne pèse au moins ⅓ au sein d'une branche ouest du sémitique qui serait à égalité avec une branche est, et au plus ¼, voire ⅓, à l'intérieur d'une famille où l'unité d'une branche ouest ne serait pas reconnue. C'est dire l'importance de ce groupe de langues en apparence marginal pour l'histoire linguistique du Moyen-Orient. Or jusqu'à présent les conditions ne sont pas réunies pour un emploi rigoureux du sudarabique moderne dans le comparatisme sémitique. On ne saurait se contenter, comme on le voit faire quelquefois, de citer comme exemple *ad hoc*, selon les besoins d'un raisonnement conçu ailleurs, une ou deux formes isolées tirées d'une langue unique, généralement le mehri, sans considération ni du système morphophonologique où ces formes s'inscrivent, ni des autres langues du groupe. Ces dernières doivent avant toute chose être comparées selon des méthodes aujourd'hui bien établies et qu'on ne devrait pas pouvoir négliger.

Peut-être répugnons-nous obscurément à honorer des langues non écrites de la même rigueur qu'on applique aux langues du livre. On ne doit pas céder à cet instinct, même dans une région du monde qui fut le premier berceau des écritures. Il est vrai néanmoins que ces langues presque uniquement orales, imparfaitement décrites et à la morphophonologie très complexe ne facilitent pas la tâche du comparatiste. C'est à ce dernier que les pages qui suivent veulent proposer un outil qui, je crois, permet de jeter une lumière fructueuse sur la diachronie du sudarabique moderne. Cet outil a été élaboré dans le cadre d'une recherche plus large que ce qu'il est possible de présenter ici[1], et certains points importants ne pourront pas être abordés. L'exposé portera sur les formes « simples » (autrement dit non dérivées) du verbe. En amont se trouve un travail sur le vocalisme du jibbali et sur sa morphologie verbale, dont les résultats serviront de base au raisonnement mais ne pourront pas être démontrés en détail[2]. En aval, la rigueur impose d'appliquer la même analyse, pour la tester, aux formes verbales dérivées; il se trouve que ces dernières confirment à mon avis la validité de l'approche proposée, mais la place manquera pour en parler ici.

2. Considérations préliminaires

2.1. Conventions d'écriture

Les consonnes radicales et les positions vocaliques correspondantes (réalisées ou non) sont numérotées comme suit: $V_0C_1V_1C_2V_2C_3V_3$.

Le signe γ représente une consonne gutturale. L'écriture ´γ´ décrit une racine ayant une gutturale en position C_2, ´´γ décrit une racine à gutturale finale. De même w´´ décrit une racine à /w/ initial, etc.

[1] Dufour (2016).

[2] Travail effectué dans le cadre du projet ANR OmanSaM. Il en va de même pour la morphophonologie du mehri d'Oman, explorée par Sabrina Bendjaballah et Philippe Ségéral au cours du même projet, qui a permis de préciser sur certains points les données de Johnstone.

Le signe Ð désigne l'allomorphe par défaut d'une catégorie morphologique donnée, c'est-à-dire celui employé en l'absence d'une des perturbations conditionnant les allomorphies décrites en 2.3.

Langues: ar. cl.: arabe classique; jb.: jibbali[3]; hbN.: hobyot[4]; mh.: mehri; mhO. mehri d'Oman[5]; mhŠ.: mehri de la Šarqiyyah yéménite[6]; sq.: soqotri[7].

Temps verbaux: ACC.: accompli; INACC.: inaccompli; SUBJ.: subjonctif. Il ne sera pas traité du conditionnel. Une notation comme INACC.2fs est à comprendre comme « inaccompli, 2e personne du féminin singulier », etc.

On a visé la cohérence dans la transcription des différentes langues. Le système suivi est relativement proche de celui des travaux de Johnstone, avec quelques adaptations. Les orthographes de Sima (2009) et Nakano (2013) ont été converties dans ce système. Pour le soqotri, la transcription adoptée est inspirée des travaux de Naumkin *& al.* avec quelques modifications.

2.2. L'opposition Ga / Gb

En sudarabique moderne comme dans les autres langues sémitiques, les lexèmes verbaux se répartissent entre des classes morphologiques dont certaines sont communément dites *formes simples* et les autres *formes dérivées*. La morphologie des premières ne met en jeu (hors marques flexionnelles) que les consonnes de la racine et une mélodie vocalique, tandis que les deuxièmes comportent du matériel morphologique supplémentaire (préfixe, infixe, gémination consonantique, longueur vocalique). Il existe en sudarabique moderne deux *formes simples*, qui s'opposent d'une part l'une à l'autre[8] et d'autre part aux *formes dérivées*. On les appellera ici Ga (par ex. mhO. *bərū́k / yəbū́rək / yəbrḗk* 'baraquer') et Gb (par ex. mhO. *θī́bər / yəθbṓr / yəθbṓr* 'se casser'). Il a été reconnu depuis longtemps, sur des bases aussi bien morphologiques que sémantiques, qu'elles correspondaient respectivement aux types ouest-sémitiques représentés, par exemple, par l'arabe classique $C_1aC_2aC_3a$ / $yaC_1C_2uC_3$ (ou $yaC_1C_2iC_3$) et $C_1aC_2iC_3a$ / $yaC_1C_2aC_3$. C'est de ces formes simples Ga et Gb qu'il sera question dans le présent article, à l'exclusion des formes dérivées. La place manquera également pour traiter des personnes autres que la 3e du masculin singulier.

[3] Les formes pour le jibbali proviennent d'un travail d'élicitation mené avec ʿĀmir Azād ʿAdlī al-Kaṯīrī, dialecte de Gúfɛ, Dhofar oriental.

[4] Tel que décrit par Nakano (2013).

[5] Tel que décrit par Johnstone et confirmé par le travail de terrain mené dans le cadre du projet ANR OmanSaM.

[6] Tel qu'attesté dans Sima (2009).

[7] Les formes citées sont tirées tantôt de Naumkin, Bulakh *& al.* (2014a, 2014b) ou Naumkin, Kogan *& al.* (2014), tantôt de mes enquêtes de terrain.

[8] Cette opposition est toutefois vacillante pour les racines à /w/ initial, et pratiquement neutralisée pour les racines bilitères et pour les racines trilitères à glide médian.

2.3. Le supplétisme morphologique phonologiquement conditionné dans le verbe

Une caractéristique commune et remarquable des langues sudarabiques modernes est que leur morphologie verbale est régie par un système d'allomorphies phonologiquement conditionnées par la nature et le nombre des consonnes de la racine. Pour une catégorie morphologique donnée, les racines, en fonction de leurs caractéristiques phonologiques, sélectionnent des schèmes divers, irréductibles les uns aux autres par des processus phonologiques valides en synchronie dans les langues considérées, et qui se trouvent les uns à l'égard des autres en situation de supplétisme. Ainsi, pour la forme dérivée souvent dite « causative », l'inaccompli 3ms en jibbali est par défaut du type *yəcC$_1$eC$_2$ɔ́C$_3$* (*yəffelɔ́t* 'il s'échappe'); mais si la troisième consonne radicale est une gutturale, le schème est *yəcC$_1$ɔ́C$_2$əC$_3$* (*yəffɔ́gəˁ* 'il effraye'); et si la racine est bilitère, le schème est *yəcC$_1$C$_2$éC$_2$* (*yəkkbéb* 'il fait descendre', √kb). Si on laisse de côté la question du *nombre* des consonnes radicales, les consonnes dont la *nature* est susceptible de conditionner de telles allomorphies sont les gutturales (/ˀ/, /h/, /ħ/, /ˁ/, /x/, /ġ/[9]) et les glides (/w/, /y/). Bien que ces phénomènes soient, en synchronie, d'ordre morphologique et inexplicables par la seule phonologie, il est probable qu'il sont historiquement issus de processus phonologiques, qu'on doit s'attacher à élucider.

Pour qu'une explication diachronique du verbe sudarabique moderne soit valide, elle doit rendre compte de l'ensemble du système, et non pas seulement d'un ou de quelques allomorphes. Faute de place cependant, ne seront considérées dans les pages qui suivent que les allomorphes servant aux racines trilitères sans perturbations (qui seront appelés allomorphes par défaut, en abrégé: Ð), ainsi que ceux pour les racines trilitères à gutturale finale (´´γ) et à gutturale médiane (´γ´).

2.4. Phonologie des gutturales en jibbali

À côté de ces *allomorphies* phonologiquement conditionnées qui concernent en propre le système verbal, la morphologie sudarabique moderne dans son ensemble présente également des *allophonies* répondant au jeu normal de la phonologie de chaque langue. Il n'en sera pas question ici. Une règle phonologique du jibbali mérite cependant considération: une voyelle atone en syllabe ouverte devant une gutturale devient régulièrement /ə/ (voyelle ultra-brève, voire susceptible de chuter en contexte VC_CV).

Schème	*Réalisation canonique*	*Avec gutturale*
yəcC$_1$eC$_2$ɔ́C$_3$	yəðelɔ́f	yəgəˁɔ́r
C$_1$eC$_2$íC$_3$	ḳedím	səˁíd
məC$_1$C$_2$əC$_3$ɔ́t	mənṣərɔ́t	məsfəˁɔ́t

yəðelɔ́f 'faire sauter' INACC.3ms, *yəgəˁɔ́r* 'faire tomber' INACC.3ms;
ḳedím 'ancien', *səˁíd* 'heureux';
mənṣərɔ́t 'bande de peau ôtée sur le devant du crâne', *məsfəˁɔ́t* 'gifle'.

[9] Le phonème /ġ/ est le plus souvent réalisé éjectif ([x'], [χ']) en jibbali et mehri du Dhofar.

Ce phénomène s'accompagne en outre d'une transparence régressive des gutturales: dans les mots comme ceux ci-dessus, la réalisation du /ə/ copie normalement le timbre de la voyelle située à droite de la gutturale:

Schème	*Exemple avec gutturale*	*Réalisation phonétique*
yəcC_1eC_2óC_3	yəgəˁór	y^{i}g^{ɔˁ}ór, y^{i}g^{ˁ}ór
C_1eC_2íC$_3$	səˁíd	s^{iˁ}íd
məC_1C_2əC_3ót	məsfəˁót	məsf^{ɔˁ}ót

2.5. L'accent de mot

Toutes les langues sudarabiques parlées sur le continent possèdent un accent de mot. Cela est vrai y compris pour le jibbali, même si les notations de Johnstone (en particulier dans le *Jibbāli Lexicon*) comportent souvent plusieurs accents graphiques par mot: c'est généralement, dans ce cas, le plus à droite des accents de Johnstone qui marque la véritable voyelle tonique; parmi les voyelles prétoniques, celles non dotées par lui d'un accent sont les diverses colorations contextuelles d'un unique phonème /ə/. Le plus souvent, la place de l'accent en mehri et en jibbali est concordante pour les formes cognates. En soqotri en revanche, la place de l'accent paraît n'avoir aucun rapport avec celle des langues continentales.

En mehri (Bendjaballah & Ségéral 2014) comme en jibbali, il ne peut pas y avoir de voyelle atone entre deux consonnes sourdes. Ainsi face à jibbali *ḳədərót* 'elle a pu' (ACC.3fs) on a *ˁəksót* (et non ***ˁəkəsót*) 'elle a photographié', *ħfəθót* (et non ***ħəfəθót*) 'elle a gardé', *skfót* (et non ***səkəfót*) 'elle s'est assise'.

En soqotri cette règle n'est en apparence pas valide. Cependant chaque schème morphologique comporte une « voyelle stable », c'est-à-dire une voyelle qui apparaît dans les formes effectives même lorsque les consonnes de la racine sont telles qu'elle se trouve entre deux consonnes sourdes, alors que les autres voyelles du schème sont « instables », c'est-à-dire qu'elles n'apparaissent généralement pas lorsque la racine est telle qu'elles auraient dû se trouver entre deux consonnes sourdes. Ainsi face à *belégoh* 'elle a mis' (ACC.3ms) on a *ktɛ́boh* (et non ***ketɛ́boh*) 'elle a écrit' et *néśfoh* (à côté de *neśɛ́foh*) 'elle a bu une gorgée', et on peut donc écrire le schème commun à ces items comme *C_1eC_2éC_3oh*, avec voyelle stable (soulignée) en V_3. La voyelle stable des schèmes du soqotri correspond généralement à la voyelle accentuée des schèmes correspondants dans les langues continentales. Elle est la preuve que le soqotri a connu dans sa préhistoire un accent de mot comparable à celui des autres langues du groupe dans son positionnement et dans ses effets. Son identification permet de connaître la place de l'accent en proto-soqotri. Elle permet également de poser l'existence de l'accent de mot pour le proto-sudarabique moderne, les cas de divergence entre les langues devenant des exceptions à expliquer.

3. Hypothèse diachronique à partir des faits du jibbali

3.1. Un vocalisme différentiel

Pour la clarté de l'exposé, le raisonnement sera d'abord présenté uniquement à partir des faits du jibbali. On élargira ensuite l'examen aux autres langues. De même, on s'attachera d'abord à comprendre la place de l'accent dans les différentes formes, avant de considérer le timbre des voyelles.

	Ga ACC.*3ms*	*Gb* ACC.*3ms*	*Ga* SUBJ.*3ms*	*Gb* SUBJ.*3ms*	*Ga* INACC.*3ms*	*Gb* INACC.*3ms*
Jibbali (schèmes Ð)	ḳədɔ́r	féðər	yɔ́ḳdər	yəfðɔ́r	yəḳɔ́dər	yəfeðɔ́r

ḳədɔ́r 'pouvoir', *féðər* 'trembler de peur'.

L'accompli et le subjonctif du Ga et du Gb sont les formes pour lesquelles les étymons à supposer sont les moins douteux. Ce sont en effet celles où les leçons des différentes langues sémitiques concordent le mieux. À l'accompli, le type Ga a toutes les chances de remonter à *$C_1aC_2aC_3a$, et le type Gb à *$C_1aC_2iC_3a$. Au subjonctif, cognat de ce qu'on nomme dans d'autres langues *jussif* ou *apocopé*, la comparaison suggère pour le Gb une forme de type *$yVC_1C_2aC_3$; pour le Ga, deux types de vocalisation sont attestés en sémitique de l'ouest, reflétés par ar. cl. *yaktub* et *yaḍrib*, mais qui ne sont pas distingués en sudarabique moderne. On écrira par convention *yVC_1C_2u/iC_3.

Le principal point d'incertitude dans ces proto-formes est la vocalisation des préfixes flexionnels aux conjugaisons préfixales. Certaines langues ouest-sémitiques attestent, directement ou indirectement, une vocalisation /a/ pour les Ga et /i/ pour les Gb (type *yaktub ~ *yilbaš), selon la loi dite de Barth-Ginsberg. En l'occurrence, ce n'est le cas ni du guèze ni de l'arabe dans sa forme classique, qui seront cités ici pour comparaison.

On peut ainsi poser les proto-formes suivantes pour l'accompli et le subjonctif:

	Ga ACC.*3ms*	*Gb* ACC.*3ms*	*Ga* SUBJ.*3ms*	*Gb* SUBJ.*3ms*
Proto-formes	*$C_1aC_2aC_3a$	*$C_1aC_2iC_3a$	*yVC_1C_2u/iC_3	*$yVC_1C_2aC_3$
Arabe classique	kataba	labisa	yaktub	yalbas
Guèze	ḳätälä	läbsä	yəḳtəl	yəlbäs
Jibbali (schèmes Ð)	ḳədɔ́r	féðər	yɔ́ḳdər	yəfðɔ́r

Ar. *kataba* 'écrire'; ar., guèze *labisa*, *läbsä* 'revêtir'; guèze *ḳätälä* 'tuer'.

S'agissant du jibbali, et à supposer que les proto-formes soient les bonnes pour cette langue, on peut faire plusieurs remarques. D'une part le *a final des formes d'accompli a visiblement chuté à un moment ou un autre de l'évolution de ces formes. Ailleurs, le *a semble être devenu /ɔ/, sauf dans *féðər* où l'on a /e/. Le *i et le *u/i, en revanche,

sont devenus /ə/. Si /ɔ/ est le produit de *a, alors *yɔ́ḳdər* suppose *yaḳdu/ir avec un *a dans le préfixe; pour *yəfðór*, c'est plus difficile à dire puisque le préfixe est inaccentué et donc susceptible d'avoir un vocalisme réduit.

Considérons maintenant la place de l'accent. Dans *ḳədór* il est final, dans *féðər* pénultième. La forme *läbsä* du guèze pourrait fournir une explication à *féðər*: le *i en syllabe ouverte de *faðira aurait été syncopé à un stade de l'évolution, ce qui n'aurait laissé comme voyelle accentuable que le *a en V_1 (puisque visiblement le *a final est resté atone ou a chuté avant le développement de l'accent moderne). Cela implique de supposer que, dans un deuxième temps, après la chute du *a final, le groupe consonantique C_2C_3 a été disjoint par insertion d'une voyelle: *faðira > *fVðra > *fVðr > *féðər*. La chute des voyelles historiquement hautes en contexte VC_CV est attestée par exemple en guèze et dans certains dialectes arabes; de même, la disjonction des groupes consonantiques finaux est attestée dans plusieurs langues sémitiques.

Cette explication, toutefois, ne rendrait pas compte de l'accentuation en V_0 de *yɔ́ḳdər* < *yáḳdu/ir, où le *u/i peut difficilement avoir été syncopé. D'autre part, *yəfðór* face à *yɔ́ḳdər* exclut l'hypothèse que le subjonctif aurait en tant que tel une accentuation particulière, initiale par exemple. On peut en revanche remarquer que, si l'on fait abstraction du *a final de *kataba qui a chuté, l'accent du jibbali moderne tombe, dans les formes ci-dessus, sur le premier *a en venant de la droite. D'où l'hypothèse que le positionnement de l'accent serait sensible à la nature des voyelles, *a étant plus accentuable que les voyelles hautes. Cette hypothèse n'implique aucune syncope. La « faiblesse » des voyelles hautes ne consisterait pas en ce qu'elles seraient inaptes à constituer le noyau d'une syllabe ouverte, mais en ce qu'elles seraient inaptes à constituer le noyau d'une unité accentuable.

Testons maintenant cette hypothèse sur l'inaccompli (schèmes Ð):

	Ga ACC.*3ms*	*Gb* ACC.*3ms*	*Ga* SUBJ.*3ms*	*Gb* SUBJ.*3ms*	*Ga* INACC.*3ms*	*Gb* INACC.*3ms*
Proto-formes	*$C_1aC_2aC_3a$	*$C_1aC_2iC_3a$	*yaC_1C_2u/iC_3	*$yVC_1C_2aC_3$	*$yVC_1\alpha C_2u/iC_3$	*$yVC_1\alpha C_2aC_3$
Arabe classique	kataba	labisa	yaktub	yalbas	yaktubu	yalbasu
Guèze	ḳätälä	läbsä	yəḳtəl	yəlbäs	yəḳättəl	yəläbbəs
Jibbali (schèmes Ð)	ḳədór	féðər	yɔ́ḳdər	yəfðór	yəḳɔ́dər	yəfeðór

En V_2, les voyelles de l'inaccompli sont exactement les mêmes que dans le subjonctif correspondant en jibbali comme en arabe – mais pas en guèze, où Ga et Gb confondent leur morphologie à l'inaccompli, le Gb adoptant la voyelle qui était caractéristique du Ga au subjonctif. Si nos règles sont valides, *yəfeðór* suppose un *a en V_2 et *yəḳɔ́dər* un *u/i.

Dans *yəḳɔ́dər*, l'accent se trouve sur une voyelle /ɔ/ en V_1 qui n'a pas d'équivalent dans la forme arabe correspondante, et qui peut rappeler le /ä/ qu'a le guèze en cette position. Dans *yəfeðór* il y a également une voyelle dans la même position, mais elle est atone et a le timbre /e/. Dans *yəḳɔ́dər* l'accent frappe la voyelle en V_1, dont le produit /ɔ/ évoque les produits de *á ailleurs. On peut faire la supposition que la voyelle en V_1

a une même origine dans *yəḳ́dər* et dans *yəfeðór*, et que son produit n'y est différent que parce qu'elle est atone dans celui-ci et tonique dans celui-là (la différence d'accentuation étant due à la différence de nature de la voyelle en V_2 entre les deux formes). Notons pour le moment *α cette voyelle en V_1 dans les proto-formes, afin de signaler sa ressemblance avec *a sans préjuger de son origine. Cette voyelle paraît propre à être accentuée, puisqu'on a *yəḳ́dər* et non pas ***yVḰḳVdər*. Il est vrai qu'on pourrait songer à d'autres explications: une règle de limitation de remontée de l'accent, ou une vocalisation du préfixe flexionnel différente de celle du subjonctif; ces hypothèses paraissent toutefois *a priori* plus coûteuses que de considérer *α comme accentuable.

3.2. Les Ga et Gb ´ɣ´ et ´´ɣ: l'effet gutturale

Intéressons-nous maintenant aux formes dédiées en jibbali aux racines ayant une gutturale en deuxième ou en troisième position. Pour les racines ´ɣ´, le type étiqueté ici Ga est très minoritaire dans le lexique (une douzaine de lexèmes), et visiblement en voie de disparition.

	Ga ACC.3ms	*Ga* SUBJ.3ms	*Ga* INACC.3ms	*Gb* ACC.3ms	*Gb* SUBJ.3ms	*Gb* INACC.3ms
	*$C_1aC_2aC_3a$	*yaC_1C_2u/iC_3	*$yVC_1\alpha C_2u/iC_3$	*$C_1aC_2iC_3a$	*$yVC_1C_2aC_3$	*$yVC_1\alpha C_2aC_3$
Ð	ḳ͘odór	yóḳ͘dər	yəḳ͘ódər	féðər	yəfðór	yəfeðór
´ɣ´	gəˁár	yógˁər	yḗgˁər	ṣ͘əhér	yəṣ͘hór	yəṣ͘əhór
ɣ´´	défəˁ	yódfəˁ	yədófəˁ	férəħ	yəfráħ	yəfórəħ

gəˁár 'tomber', *ṣ͘əhér* 'apparaître', *défəˁ* 'pousser', *férəħ* 'se réjouir'.

3.2.1. Précisions sur le tableau ci-dessus

Sont en gris dans le tableau les formes des verbes dont les consonnes radicales ne sont pas de nature à induire une perturbation (et qui ont les schèmes par défaut) ainsi que celles qui, pour les verbes à gutturales, suivent les schèmes par défaut. Dans toutes ces formes, la gutturale, quand il y en a une, a à sa gauche soit une autre consonne, soit un /ə/ atone.

Sont en blanc dans le tableau les formes pour lesquelles les racines à gutturales sélectionnent un allomorphe de schème différent du schème par défaut.

Le cas de Gb INACC.3ms ´ɣ´ *yəṣ͘əhór* est particulier: la gutturale a bien à sa gauche un /ə/ mais le schème auquel on peut comparer cette forme (le schème de *yəfeðór*) a dans cette position un /e/ atone. Cependant (cf. 2.4. *supra*) le processus /e/→/ə/ en syllabe atone devant une gutturale en attaque peut être considéré comme régulier dans la phonologie synchronique du jibbali. Il n'est donc pas nécessaire de poser ici une allomorphie de schème.

Ga INACC.3ms *yḗgˁər* requiert une explication. Les voyelles longues du jibbali, bien que très fréquentes dans la chaîne parlée, sont à peu près toujours secondaires et issues soit d'une contraction par chute d'une consonne intervocalique, soit d'un allongement

accompagnant une resyllabification (par ex. Ga ACC.3fs **dələfə́t* → *dəlfə́t* 'elle a sauté', ou **dəfəˁə́t* → **dəfəˁə́t* → *dəfˁə́t* 'elle a poussé'). On proposera donc ici de voir dans *yə́gˁər* la réalisation de **yégəˁər*.

3.2.2. Remarques sur la place de l'accent

On peut faire les remarques suivantes sur les formes à allomorphie de schème (formes en blanc dans le tableau):

- Ces formes à allomorphies sont toutes des formes où, si la racine avait sélectionné le schème sain, la voyelle accentuée aurait été juste à gauche de la gutturale.
- Dans tous les cas sauf un (Gb SUBJ.3ms ´´ɣ *yəfráħ*), le schème réellement sélectionné a l'accent ailleurs qu'à gauche de la gutturale.

L'existence d'allomorphes propres aux racines à gutturales a ainsi comme résultat d'éviter que l'accent tombe juste à gauche d'une gutturale, sauf dans Gb SUBJ.3ms ´´ɣ *yəfráħ*. Ce fait peut être mis en rapport avec les deux éléments suivants, déjà rencontrés:

- En synchronie, un processus phonologique fait passer en jibbali une voyelle atone à /ə/ devant une gutturale.
- En diachronie, nous avons supposé que l'accent frappait la voyelle basse *a de préférence à *i et *u, ces dernières aboutissant à /ə/ atone.

Il est alors tentant de considérer que le jibbali a connu dans sa préhistoire un effet des gutturales en partie similaire à celui qu'elles ont aujourd'hui. Toute voyelle située à la gauche d'une gutturale aurait été « faible » comme *i et *u, c'est-à-dire que:

- Comme *i et *u elle aurait été préférentiellement inaccentuable.
- Comme *i et *u elle aboutit généralement à /ə/ atone.

3.2.3. Évaluation de l'hypothèse

On peut à la lumière de cette hypothèse reprendre une à une les formes qui diffèrent des schèmes sain[10]. La « faiblesse » d'une voyelle est notée par la brève: ŭ.

- Ga ACC.3ms ´´ɣ *dafaˁa > *dafăˁa > *dáfăˁa, parallèle à Gb ACC.3ms *faðira > *faðĭra > fáðĭra, les deux formes aboutissant à un résultat similaire: *défəˁ*, *féðər*.

[10] Les formes notées ici avec un astérisque et en romain, et qui sont appelées « proto-formes », ne sont que des raccourcis schématiques visant à rendre facilement intelligible à l'œil un processus supposé avoir eu lieu. Une écriture comme *dáfăˁa n'implique ni qu'il existait un verbe Ga sur la racine √dfˁ à l'époque où le processus de développement d'un accent de mot s'est déroulé, ni que les voyelles des formes de ce types avaient alors le timbre [a], ni même que le *a final était encore présent à cette date. Il ne s'agit donc pas d'un étymon au sens propre, et le mot *dáfăˁa n'est pas censé avoir existé. De telles formes sont notées en romain. Les formes en italiques précédées d'un astérisque sont réservées aux cas où l'on peut considérer que le processus illustré est valable en synchronie dans la phonologie de la langue (par exemple **yégəˁər* > *yə́gˁər*).

- Ga INACC.3ms ´γ´ *yagαˁu/ir > * yagăˁŭ/ĭr > *yágăˁŭ/ĭr > *yégəˁər → *yḗgˁər*.
- Ga ACC.3ms ´γ´ *gəˁár*: cette forme a exactement même schème que la suivante (*ṣəhér*), [a] n'étant qu'un allophone de /ɛ/ au contact d'une consonne abaissante (/ħ/, /ˁ/, /x/, /ġ/). Il paraît délicat de la faire dériver de *gaˁara, dont on attendrait **gəˁór → **gəˁór* (cf. *ḳadara > *ḳədór*). Elle est donc problématique, mais nous y reviendrons dans le cadre d'une comparaison des différentes langues sudarabiques modernes.
- Gb ACC.3ms ´γ´ *ṣahira > *ṣ̌ăhĭra*. Cette forme, selon notre hypothèse, comporte deux voyelles « faibles ». On doit se borner à constater que c'est celle de droite qui est accentuée: *ṣəhér*. On peut donc observer ici ce qu'il advient d'un *i lorsqu'il est accentué au terme de l'évolution: *í > /é/.
- Gb INACC.3ms ´´γ *yVfαraħ > *yVfαrăħ > *yVfárăħ > *yəfórəħ*. Cette forme est parallèle à Ga INACC.3ms Ð *yVḳαdu/ir > *yVḳαdŭ/ĭr > *yVḳádu/ir > *yəḳódər*. L'inaccompli d'un Gb ´´γ est donc identique à celui d'un Ga. La raison en est que les deux formes ont été accentuées en V_1, alors même que cette accentuation identique a des causes différentes: pour *yəfórəħ*, c'est la présence d'une gutturale en C_3, pour *yəḳódər*, c'est la présence d'un *u/i en V_2. Cela confirme la supposition faite en 3.1. *supra* que la voyelle apparaissant en V_1 à l'inaccompli des Ga et des Gb est historiquement la même dans les deux cas, et qu'elle n'a un aboutissement différent dans les schèmes par défaut (respectivement /ɔ/ dans *yəḳódər* et /e/ dans *yəfeðór*) qu'à cause de la différence de placement de l'accent. En effet, dès qu'une gutturale en C_3 provoque la remontée de l'accent en V_1 dans un Gb, alors cette voyelle a même aboutissement que dans un Ga. Il y a donc bien un élément *α commun aux Ga et aux Gb pour la formation de l'inaccompli.
- Gb SUBJ.3ms ´´γ *yVfraħ > *yVfrăħ*, avec comme résultat *yəfráħ*. On aurait pu s'attendre à ce que l'accent frappe le préfixe comme dans Ga SUBJ.3ms Ð *yóḳdər*, avec comme résultat **yVfrəħ*. Ce n'est pas le cas. La première explication qui vient à l'esprit est que la voyelle du préfixe flexionnel n'était pas *a mais *i et que, dans *yĭfrăħ comme dans Gb ACC.3ms ´γ´ *ṣ̌ăhĭra > *ṣəhér*, c'est la voyelle faible la plus à droite qui a été accentuée. Nous verrons que d'autres explications sont possibles. Quoi qu'il en soit, le produit de *á « faible » accentué est /é/, exactement comme celui de *í, ce qui confirme l'affinité entre *a « faible » et les voyelles hautes, qui concordent non seulement par leur propriétés accentuelles et leur aboutissement final en position atone /ə/, mais aussi par leur aboutissement en position tonique /é/.

3.2.4. Note sur les timbres vocaliques

Nous avons rencontré deux aboutissements de *á tonique: /ɔ/ et /é/. /ɔ/ est attesté dans Ga ACC.3ms Ð *ḳədór*, Ga SUBJ.3ms Ð *yóḳdər*, Gb INACC.3ms Ð *yəfeðór* et Gb SUBJ.3ms Ð *yəfðór*; /é/ est attesté dans Gb ACC.3ms Ð *féðər*, Ga ACC.3ms ´´γ *défəˁ* et Ga INACC.3ms ´γ´ *yḗgˁər* < *yégəˁər*. Malgré le petit nombre de formes considérées, il semble que *á donne /ɔ/ en syllabe fermée et /é/ en syllabe ouverte. Le *ά tonique, en revanche, évolue vers /ɔ/.

Il est temps maintenant d'élargir la comparaison aux autres langues SAM.

4. Les allomorphes Ð, ˊɣˊ et ˊˊɣ des formes verbales simples en SAM

Les tableaux ci-dessous donnent les formes pour le jibbali, le soqotri, le mehri d'Oman, le mehri de la Šarqiyyah yéménite (Sima 2009)[11] et le hobyot (Nakano 2013). La voyelle stable des schèmes du soqotri est soulignée dans les items servant à les représenter.

Ð	*Ga* ACC.3ms	*Ga* SUBJ.3ms	*Ga* INACC.3ms	*Gb* ACC.3ms	*Gb* SUBJ.3ms	*Gb* INACC.3ms
	$*C_1aC_2aC_3a$	$*yaC_1C_2u/iC_3$	$*yVC_1\alpha C_2u/iC_3$	$*C_1aC_2iC_3a$	$*yVC_1C_2aC_3$	$*yVC_1\alpha C_2aC_3$
Jibbali	ḳ͘ɔdɔ́r	yɔ́ḳ͘dər	yəḳ͘ɔ́dər	féðər	yəfðɔ́r	yəfeðɔ́r
Soqotri	bézo̲g	ʎibzé̲g[12]	yebɔ́zəg	bé̲zɛg	ʎibzɔ́g	yebézo̲g
Mehri O.	bərṹk	yəbrɛ́̃k	yəbṹrək	θíbər	yəθbɔ́̃r	yəθbɔ́̃r
Mehri Š.	[bərɔ́̃k]	[yəbrɛ́̃k]	[yəbɔ́̃rək]	[θíbər]	[yəθbɔ́̃r]	[yəθbɔ́̃r]
Hobyot	bərɔ́̃k	yəbrɛ́̃k	ybɔ́̃rək	fírəḳ͘	yəfrɔ́̃ḳ͘	yfarɔ́̃ḳ͘

Sq. *bézog* 'couper', *bézɛg* 'être coupé'; hbN. *bərɔ́̃k* 's'agenouiller', *fírəḳ͘* 'craindre'.

ˊɣˊ	*Ga* ACC.3ms	*Ga* SUBJ.3ms	*Ga* INACC.3ms	*Gb* ACC.3ms	*Gb* SUBJ.3ms	*Gb* INACC.3ms
	$*C_1a\gamma_2aC_3a$	$*yaC_1\gamma_2u/iC_3$	$*yaC_1\alpha\gamma_2u/iC_3$	$*C_1a\gamma_2iC_3a$	$*yVC_1\gamma_2aC_3$	$*yVC_1\alpha\gamma_2aC_3$
Jibbali	(gəˢár)	yɔ́gˢər	yɛ́̃gˢər	ṣəhɛ́r	yəṣhɔ́r	yəṣəhɔ́r
Soqotri	nɔ́ħo̲g)	ʎinḥá̲g	yená̲ħag	łáˢa̲f	ʎiłˢɔ̲́f	yełɔ́ˢo̲f
Mehri O.	(rəḥɛ́̃ṣ)	yərḥɛ́̃ṣ	yərṹ̃həṣ	zəḥɛ́̃f	yəzhɔ́̃f	yəzhɔ́̃f
Mehri Š.	?	?	[yənɔ́̃ħəg]	[ḳəhɛ́̃b]	[yəḳhɔ́̃b]	[yəḳhɔ́̃b]
Hobyot	(nəḥɛ́̃g)	yənḥɛ́̃g	ynɛ́̃həg	zəḥɛ́̃f	yəzhɔ́̃f	yəzhɔ́̃f

Sq. *nɔ́ħog* 'jouer', *láˢaf* 's'habituer'; mhO., hbN. *rəḥɛ́̃ṣ* 'laver', *zəḥɛ́̃f* 'ramper'.

ˊˊɣ	*Ga* ACC.3ms	*Ga* SUBJ.3ms	*Ga* INACC.3ms	*Gb* ACC.3ms	*Gb* SUBJ.3ms	*Gb* INACC.3ms
	$*C_1aC_2a\gamma_3a$	$*yaC_1C_2u/i\gamma_3$	$*yVC_1\alpha C_2u/i\gamma_3$	$*C_1aC_2i\gamma_3a$	$*yVC_1C_2a\gamma_3$	$*yVC_1\alpha C_2a\gamma_3$
Jibbali	défəˢ	yɔ́dfəˢ	yədɔ́fəˢ	férəħ	yəfráħ	yəfɔ́rəħ
Soqotri	ká̲naħ	ʎikná̲ħ	yeká̲naħ	śé̲baˢ	ʎiśbá̲ˢ	yeśá̲baˢ
Mehri O.	sṹ̲bəħ	yəsbɛ́̃ħ	yəsṹ̲bəħ	fírəħ	yəfrēħ	yəfṹ̲rəħ
Mehri Š.	[sṹ̲bəħ]	[yəsbɛ́̃ħ]	[yəsɔ́̃bəħ]	[fírəħ]	[yəfráħ]	[yəfɔ́̃rəħ]
Hobyot	nṹ̲kəˢ	yənkɛ́̃ˢ	ynɔ́̃kəˢ	ríbəħ	yərbɛ́̃ħ	yrɔ́̃bəħ

Sq. *kánaħ* 'refaire', *śébaˢ* 'avoir assez mangé'; mhO. *fírəħ* 'se réjouir', *sṹ̲bəħ* 'nager'; hbN. *nṹ̲kəˢ* 'arriver', *ríbəħ* 'gagner'.

[11] Les textes de Simah (2009) attestent bien la morphologie présentée dans le tableau, mais pas toujours à tous les temps pour un même verbe, et pas toujours à la 3ms. Les formes du tableau sont donc en partie refaites. Leur orthographe a par ailleurs été rapprochée de celle employée pour le mehri d'Oman. Pour toutes ces raisons, elles sont ici données entre crochets.

[12] Les formes soqotries de SUBJ.3ms comportent une particule *l-* préfixée, la réalisation de la séquence *l-ye-* étant ici transcrite *ʎi-*.

Dans les formes mehries ci-dessus, [ē] et [ā] sont des allophones d'un même phonème, l'abaissement [ē]→[ā] ayant des modalités légèrement différentes dans les deux dialectes considérés. Il en va de même en hobyot.

Semblablement, en jibbali, [ɛ] et [a] sont deux allophones d'un même phonème.

En soqotri [a] est souvent un allophone de /ɛ/, mais il est aussi parfois un allophone de /o/, comme nous le verrons.

Pour les verbes ´ɣ´, le type Ga est encore plus marginal en mehri qu'en jibbali. À l'accompli, il a même schème que le type Gb ´ɣ´ dans toutes les langues continentales. Seul le soqotri distingue les deux types à ce temps. Dans cette langue, le type *nóḥo̲g* doit être considéré comme un Ga (*naḥaga > *năḥaga > *nəḥóg → *nóḥo̲g*) et le type *láʿaf* comme un Gb (*laʿifa > *lăʿifa > *ləʿéf → *láʿa̲f*)[13]. Il est probable que le soqotri est ici archaïque, et que dans les langues continentales le type Ga ´ɣ´, devenu résiduel, a emprunté sa morphologie d'accompli au Gb ´ɣ´. Les formes d'accompli Ga ´ɣ´ des langues continentales sont donc mises entre parenthèses dans le tableau ci-dessus, parce qu'elles sont secondaires et ne dérivent pas de *$C_1aC_2aC_3a$.

Sont en grisé dans les tableaux ci-dessus les types où l'accentuation du jibbali diffère de celle attestée par les autres langues; en blanc celle où elle concorde. (Pour le soqotri, la comparaison se fait avec la voyelle stable du schème considéré.) La plupart des formes concordent, sauf:

- Le subjonctif du Ga, où, dans toutes les langues, le schème par défaut sert aussi bien aux racines saines qu'aux racines ´ɣ´ et ɣ´´, mais où le jibbali a l'accent en V_0 (*yóC₁C₂əC₃*) tandis que les autres langues l'ont en V_2 (*yVC₁C₂ɛ̄/ɛ́C₃*).
- L'inaccompli Ga ´ɣ´, où le jibbali a l'accent en V_0 (*yɛ̄́C₁ɣ₂əC₃*) alors que les autres langues l'ont en V_1 (*yVC₁V́ɣ₂VC₃*).
- On peut y ajouter l'inaccompli par défaut du Gb, où, bien que toutes les langues aient l'accent en V_2, le mehri est le seul à avoir zéro en V_1 (*yəθbṓr*) alors que toutes les autres langues ont une voyelle atone dans cette position (*yVC₁VC₂V́C₃*).

Nous commencerons par examiner les cas où la place de l'accent concorde. C'est alors la correspondance entre les timbres vocaliques des différentes langues qui nous intéressera. Puis nous traiterons des trois catégories où la concordance est imparfaite.

4.1. Correspondances des voyelles accentuées

4.1.1. {jb. ó, sq. o̲, mhO. ṓ, mhŠ. ṓ, hbN. ṓ}

C'est la correspondance que l'on trouve en V_2 dans les schèmes du subjonctif et de l'inaccompli des type Gb Ɒ et ´ɣ´ (*yVC₁C₂aC₃, *yVC₁ɣ₂aC₃ et *yVC₁αC₂aC₃, *yVC₁αɣ₂aC₃).

[13] Cf. Naumkin, Bulakh *& al.* (2014 : 47-21), qui soulignent néanmoins que beaucoup de verbes ´ɣ´ ont un accompli de type Ga et des conjugaisons préfixales de type Gb, ou inversement. Ce fait est symbolisé par une double barre verticale à droite de l'accompli du soqotri dans le tableau. Noter en outre que je suppose ici en soqotri une transparence régressive des gutturales similaire à celle du jibbali.

4.1.2. {jb. ó, sq. o̲, mhO. ű́, mhŠ. ő́, hbN. ő́}

On trouve cette correspondance dans deux cas:
- En V_2 à l'accompli du Ga Ð (*$C_1aC_2\underline{a}C_3a$). On doit y ajouter l'accompli du Ga ´γ´ *nóħo̲g* en soqotri (*$C_1a\gamma_2\underline{a}C_3a$), qui n'a pas de vrai cognat dans les langues continentales.
- En V_1 à l'inaccompli du Ga Ð, où il s'agit, selon l'analyse proposée, d'un *α accentué (*$yVC_1\alpha C_2u/iC_3$).

4.1.3. {jb. ó, sq. a̲, mhO. ű́, mhŠ. ő́, hbN. ő́}

On trouve cette correspondance en V_1 à l'inaccompli aussi bien du Ga ´´γ que du Gb ´´γ. Elle ne diffère de la précédente que par le /a̲/ du soqotri *yəkána̲ħ*, *yəśá̲ba*ˁ. Il s'agit encore une fois d'un *α accentué (*$yVC_1\underline{\alpha}C_2u/i\gamma_3$ et *$yVC_1\underline{\alpha}C_2a\gamma_3$). Ces formes du soqotri pourraient laisser penser que les racines ´´γ sélectionnent un schème profondément différent de celui des racines saines. Elles doivent cependant être remises dans le contexte de la morphologie de l'inaccompli dans cette langue[14]. (Dans le tableau ci-dessous, φ note une consonne emphatique, c'est-à-dire éjective.)

Type de racine	*Ga INACC.3ms*	*Ga INACC.3mp*		*Ga ACC.3ms*	
Sans consonne abaissante	yego̲zem	yego̲zem	*jurer*	gézo̲m	
φ´´	yeḳáder	yeḳóder	*cuire*	ḳédo̲r	
´φ´	yebáṣ̌ar	yebóṣ̌er	*couper*	bő́ṣ̌a̲r	
´´φ	yefánaḳ	yefóneḳ	*attendre*	fénaḳ	
l´´	yełóget	yełóget	*frapper*	łégo̲t	
´l´	yegółem	yegółem	*cueillir*	géło̲m	
´´l	yedábał	yedóbeł	*réunir*	dőbo̲ł	
γ´´	yeħófer	yeħófer	*creuser*	ħfó̲r	
´γ´	yesá̲ham	yesé̲hem	*pisser*	nóħo̲g	*jouer*
´´γ	yesáka^{ʕ}	yesóke^{ʕ}	*traverser une vallée*		
	Gb INACC.3ms	*Gb INACC.3mp*			
´´γ	yeśába^{ʕ}	yeśóbe^{ʕ}	*avoir assez mangé*		

Sont en grisé dans le tableau les formes comportant une perturbation du vocalisme par rapport aux formes considérées ici comme canoniques, c'est-à-dire celles des verbes

[14] Le système présenté dans le tableau ci-dessus est tiré des analyses de Naumkin, Bulakh *& al.* 2014a et des données de Naumkin, Bulakh *& al.* 2014b et Naumkin, Kogan *& al.* 2014. Il est confirmé par mes données là où elles sont pertinentes.

dont les racines ne comportent pas de consonne abaissante (et dont un exemple est donné en première ligne du tableau).

Les consonnes éjectives et /l/ perturbent le vocalisme de 3ms, mais pas celui de 3mp. Ces deux tiroirs sont ainsi identiques pour les racines saines et ne diffèrent que pour les racines induisant des perturbations. Considérons la 3ms.

Une éjective en C_1 modifie le timbre de V_1; en C_2 elle modifie le timbre de V_1 et de V_2 (voyelles adjacentes); mais lorsqu'elle est en C_3 elle modifie non seulement V_2 mais aussi V_1 (*yefánaḳ*). Il y a donc une action à distance de C_3 sur V_1.

Un /l/ en C_1 ou C_2 n'induit aucune perturbation. En C_3 en revanche, il modifie le timbre à la fois de V_1 et de V_2. Là aussi il y a une action à distance de C_3 sur V_1.

Une gutturale en C_1 n'a aucun effet perturbateur. En C_2, elle perturbe à la fois la forme du singulier et celle du pluriel, et cette situation mérite un traitement particulier (voir *infra* 4.2.3.). Les formes à gutturale en C_3 sont celles qui nous occupaient, et il s'avère que leur morphologie est identique à celle des formes ayant une éjective ou un /l/ en C_3.

Or dans les autres langues SAM, les éjectives et /l/ ne conditionnent pas d'allomorphies de schème et n'ont, quand elles en ont, que des effets purement phonologiques sur la réalisation du schème; en position C_3 dans les schèmes d'inaccompli des formes simples, elles ne provoquent aucune perturbation. Le vocalisme -CaCaC des inaccomplis ´´γ du soqotri s'inscrit donc dans un système propre à cette langue, où ce sont non pas les gutturales en tant que telles, mais les consonnes abaissantes qui provoquent une perturbation du vocalisme[15]. Cette perturbation doit sans doute être considérée comme allophonique et, autant qu'on puisse en juger, n'implique pas d'alternance dans la place de la voyelle stable, donc pas d'alternance dans la place de l'accent proto-soqotri. On doit simplement considérer que le produit de *α tonique à l'INACC.3ms est de manière générale en soqotri une voyelle normalement réalisée [o] mais qui devient [a] en contexte abaissant[16].

[15] Toutes les gutturales ne sont pas abaissantes, et /h/ par exemple ne l'est pas. Pour prouver définitivement que l'allophonie n'est pas liée aux gutturales, il faudrait pouvoir observer un verbe de ce schème sans consonne abaissante en C_1 ou C_2 et avec un /h/ en C_3. Je n'en ai pas trouvé.

[16] Il faut reconnaître néanmoins que les processus exacts mis en jeu par cette allophonie n'ont pas encore été compris, comme de manière générale le statut phonologique des voyelles de timbre [a], [ɑ], et [ɒ] en soqotri. Il est en tout cas visible ici que les modalités de l'abaissement par les éjectives ne sont pas les mêmes que celles de l'abaissement par /l/ et par les gutturales abaissantes, et que l'abaissement du /o/ n'a pas non plus le même conditionnement que celui de la voyelle qui apparaît ici en V_2 dans les INACC.3ms considérés (et dont on peut montrer qu'elle est phonologiquement un /ɛ/).
Des abaissements /o/ → /a/ au contact d'une éjective sont attestés dans de nombreuses catégories morphologiques du soqotri. On le voit par exemple dans la colonne de droite du tableau ci-dessus pour la voyelle en V_2 de Ga ACC.3ms *C_1aC_2aC_3a.

Il faut donc regrouper la correspondance {jb. ɔ́, sq. a̲, mhO. ū́, mhŠ. ṓ, hb. ṓ} avec {jb. ɔ́, sq. o̲, mhO. ū́, mhŠ. ṓ, hb. ṓ} en une correspondance unique {jb. ɔ́, sq. o̲/a̲, mhO. ū́, mhŠ. ṓ, hb. ṓ}, valable pour tous les inaccomplis des formes simples où un *α est accentué.

4.1.4. {jb. ɛ́, sq. ɛ̲, mhO. ḗ, mhŠ. ḗ, hbN. ḗ}

On trouve cette correspondance au Gb ACC.3ms ´γ´ (*$C_1a\gamma_2iC_3a$) et au Gb SUBJ.3ms ´´γ (*$yVC_1C_2a\gamma_3$), ainsi qu'en V_2 au subjonctif du Ga Ɖ (*yVC_1C_2u/iC_3), bien que le jibbali ait l'accent en V_0 dans *yɔ́ḳdər* (cf. 4.2.2. *infra*). Dans toutes les langues, la voyelle connaît des allophones de timbre [a] conditionnés par la présence d'une consonne abaissante (cas statistiquement le plus fréquent). Le fait, remarqué plus haut pour le jibbali, que le produit de *í tonique dans *$C_1a\gamma_2iC_3a$ était identique à celui de *á tonique devant gutturale dans *$yVC_1C_2a\gamma_3$ est donc valable pour tout le SAM[17].

4.1.5. {jb. é, sq. a̲, mhO. ū́, mhŠ. ū́, hbN. ū́}

On trouve cette correspondance dans le Ga ACC.3ms ´´γ (*$C_1aC_2a\gamma_3a$). Notons que si le mehri d'Oman a la même voyelle tonique dans Ga INACC.3ms Ɖ *yəkū́təb* que dans Ga ACC.3ms ´´γ *sū́bəħ*, ce n'est le cas ni du mehri de la Šarqiyyah ni des autres langues SAM. Notons également la différence entre cette correspondance et la suivante, différence qui n'était pas apparue précédemment dans l'étude du seul jibbali.

4.1.6. {jb. é, sq. e̲, mhO. ī́, mhŠ. ī́, hbN. ī́}

On trouve cette correspondance dans le Gb ACC.3ms Ɖ *$C_1aC_2iC_3a$. Cette forme n'est identique à la précédente qu'en jibbali. Soulignons que cette identité vaut pour la 3ms mais qu'une différence réapparaît à la 3fs, où les deux catégories sont distinguées par le vocalisme de la base (/i/ pour le Gb), d'une manière qui trouve un parallèle en Soqotri:

	Jibbali		*Soqotri*	
	ACC.3ms	*ACC.3fs*	*ACC.3ms*	*ACC.3fs*
Ga ´´γ	défəˁ	dōfˁɔ́t	ká̲naħ	kenóħo̲h
Gb Ɖ	féðər	fiðirɔ́t	bé̲zɛg	bizɛ́go̲h

[17] Il peut être intéressant de noter que tel est également l'aboutissement de *i atone dans soqotri Gb ACC.3ms Ɖ *bé̲zɛg* < *baziga. Ce /ɛ/ surface comme /a/ à proximité d'une consonne abaissante (gutturale, éjective, /l/). Sans consonne abaissante, il est réalisé [ɛ] par certains locuteurs, alors que chez d'autres il est impossible à distinguer d'un /e/, selon un phénomène touchant la plupart des /ɛ/ post-stables de la langue.

Le *ĭ de *fáðĭra et le *ă de *dáfă^{ʕ}a, bien qu'ils se comportent l'un et l'autre comme inaccentuables, ont ainsi conditionné des évolutions divergentes pour ces deux types. Il paraît clair que le *i en V_2 de *faðira a palatalisé le *a en V_1. Il est difficile de dater cette palatalisation relativement aux phénomènes accentuels étudiés jusqu'ici, comme d'évaluer quel a pu être le timbre de la voyelle produite par cette palatalisation. On peut remarquer toutefois d'une part que ce *a palatalisé a finalement été accentué de préférence au *i en V_2 (à la différence, par exemple, de ce qui s'est passé dans *ṣ̌ăhira > jb. *ṣ̌əhér*), et que son aboutissement final est tout à fait différent de celui de *i. Il est donc inopportun de supposer un stade *fiðira.

Cette palatalisation a peut-être eu lieu également dans le type *ṣ̌ăhĭra > jb. *ṣ̌əhér*, et cela peut aider à expliquer pourquoi c'est finalement le *i en V_2 qui a été accentué: le *a en V_1 aurait ainsi été non seulement situé à gauche d'une gutturale mais également élevé par le *i de la syllabe suivante, voire totalement assimilé à lui si l'on suppose une transparence régressive des gutturales comme en soqotri ou jibbali modernes.

4.2. Examen des formes les plus problématiques

4.2.1. Le Gb INACC.*3ms **yVfɑðar*

Le mehri est la seule langue à n'avoir pas de voyelle en V_1 dans le schème par défaut correspondant à cette forme: mhO. *yəfðṓr*, vs. hbN. *yfaðṓr*, jb. *yəfeðór*, sq. *yebézo̲g*. Il est donc également la seule langue où l'inaccompli est identique au subjonctif dans les schèmes par défaut du Gb. Pour les racines ´´γ en revanche[18], le mehri se comporte exactement comme les autres langues du groupe, avec une voyelle (tonique) en V_1 à l'inaccompli et pas de voyelle au subjonctif: mhO. INACC.3ms *yəfū́rəħ* vs. SUBJ. *yəfré̲ħ*.

En somme, le mehri a une voyelle en V_1 à l'inaccompli des Gb là où les autres langues ont une voyelle tonique, et pas de voyelle là où les autres langues (y compris le hobyot) ont une voyelle atone. Le plus simple est de considérer que le mehri a connu lui aussi cette voyelle atone dans sa préhistoire mais qu'il l'a perdue, comme ce semble être le cas pour la plupart des voyelles atones en contexte VC_CV dans cette langue. C'est la supposition que font Johnstone (1975) et Rubin (2015: 324-325) au vu des seuls schèmes par défaut. L'examen des autres allomorphes la confirme.

4.2.2. Le Ga SUBJ.*3ms **yaḳdŭ/ĭr*

Toutes les langues ont, pour ce tiroir morphologique, un schème unique que la racine soit saine, ´γ´ ou ´´γ. Le schème par défaut est donc sélectionné par les trois types de racine.

[18] Ainsi que dans les racines à glide final.

En soqotri, mehri et hobyot, on a dans Ga SUBJ.3ms Ð *yaC$_1$C$_2$u/iC$_3$ > *yəC$_1$C$_2$éC$_3$*, *yəC$_1$C$_2$ḗC$_3$* une voyelle haute accentuée en V$_2$ qui aboutit à /ɛ́/, */ḗ/* (comme dans Gb ACC.3ms ´ɣ´ *śahira > jb. *śəhér*, mhO. *śəhḗr*). Mais la forme du jibbali *yóḳdər*, avec /ɔ́/ accentué en V$_0$ s'oppose à celle de toutes les autres langues.

Pour prendre l'exemple du mehri d'Oman comme typique des langues autres que le jibbali, on voit donc l'accent remonter en V$_1$ dans Ga INACC.3ms Ð *yVbαru/ik > *yəbū́rək*, Gb ACC.3ms Ð *θabira > *θī́bər*, Gb INACC.3ms ´´ɣ *yVfαraħ > *yəfū́rəħ*, mais pas jusqu'en V$_0$ dans Ga SUBJ.3ms Ð *yabru/ik > *yəbrḗk*. Quelle que soit la raison de ce fait, il a des parallèles ailleurs dans la morphologie. Ainsi le mot arabe *maġrib* 'coucher du soleil, prière du soir' a été emprunté par les langues du Dhofar, sans doute à époque islamique, et l'on a mhO. *məġrā́b* mais jb. *múġrəb* <*móġrəb*[19], parallèle à mhO. *yəḳdḗr* face à jb. *yóḳdər*.

Il y a plus: l'accentuation de mhO. *yəbrḗk* est attestée pour le subjonctif du Ga même en jibbali. En effet, ce dernier n'a l'accent en V$_0$ que pour les personnes du singulier. Au pluriel, l'accent est en V$_2$: 1p *nəḳdér* (et 3mp *yəḳdór*). On pourrait certes chercher quel conditionnement phonologique a fait évoluer *yaḳdu/ir et *taḳdu/ir vers *yóḳdər*, *tóḳdər* et *naḳdu/ir vers *nəḳdér*. Mais il est sans doute plus simple de considérer que les deux accentuations ont existé comme variantes, et qu'elles ont été ensuite réparties morphologiquement, l'une caractérisant le singulier et l'autre le pluriel. Le déclencheur d'une telle répartition pourrait bien être à chercher dans les personnes du masculin pluriel, où le suffixe *-ū de *yaḳdu/irū, *taḳdu/irū a affecté les timbres vocaliques et a pu faire pencher la balance vers une accentuation en V$_2$: *yəḳdór*, *təḳdór* et par analogie *nəḳdér* éliminant la variante **nóḳdər*. Il apparaît en tout cas que la remontée d'accent en V$_0$ est moins automatique que la remontée en V$_1$, et que dans beaucoup de cas elle ne se produit pas, même en jibbali.

Nous avions rencontré le cas du Gb SUBJ.3ms ´´ɣ (jb. *yəfráħ*, mhO. *yəfrḗ̆ħ* < *yVfraħ), où l'effet de la gutturale en C$_3$, bien qu'il ait rendu le produit de *á semblable à celui de *í ou *ú/í, n'a pas été suffisant pour conditionner une remontée de l'accent en V$_0$, alors même que cette remontée a lieu dans Ga SUBJ.3ms Ð jb. *yóḳdər* (< *yaḳdu/ir). Nous avions évoqué la possibilité, pour expliquer cette différence d'accentuation, de supposer une vocalisation différente du préfixe flexionnel (*yifraħ vs. *yaḳdir). Nous constatons maintenant que la remontée d'accent en V$_0$ est de toute façon un processus limité. Or *yVfrăħ et *yaḳdŭ/ĭr ne sont pas exactement superposables. Certes, la voyelle en V$_2$ y est dans les deux cas « faible » au sens où nous l'avons défini. Cependant, rien ne dit qu'un *ă « affaibli » par position ait eu exactement le même poids qu'un *ĭ ou qu'un *ŭ sur ce qui semble être une échelle de sonorité, bien qu'ils se comportent dans beaucoup de cas de manière similaire d'un point de vue accentuel, et qu'ils évoluent tous vers /ɛ́/ etc. quand ils sont toniques. Si l'accent ne remonte pas en V$_0$ dans *yVfraħ, ce n'est donc pas forcément à cause de la voyelle du préfixe, cela peut être également dû à la nature de la voyelle en V$_2$.

Il peut être utile à ce stade de faire un excursus dans les formes dérivées, en considérant la morphologie de la forme dite « causative »:

[19] Le processus /ɔ/ → /u/ est régulier au contact d'une nasale en jibbali.

		ACC.3ms	*SUBJ.3ms*	*Racine*
		*$\check{h}aC_1C_2aC_3a$	*$yV\check{h}aC_1C_2iC_3$	
Ð	*Jb.*	efflét	yɛ́flət	flt
	MhO.	hənsűm	yəhánsəm	nsm
´´γ	*Jb.*	effgáˁ	yɛ́fgəˁ	fgˁ
	MhO.	həndɛ́x	yəhándəx	ndx

Ces formes soulèvent plusieurs difficultés (timbre du /é/ de *efflét*, nature de la consonne du préfixe transcrite ici *ȟ) qu'on n'abordera pas ici. Cependant, la concordance de plusieurs langues sémitiques est ici assez bonne pour ce qui est de la vocalisation, et on peut tenir celle des proto-formes pour relativement assurée. Au subjonctif *yVȟansĭm > mhO. *yəhánsəm*, l'accent remonte jusqu'en V_0. À l'accompli en revanche, *ȟandăxa > mhO. *həndɛ́x*, l'accent frappe V_2 (où le *a est affecté exactement comme dans Gb SUBJ.3ms ´´γ *yVfrăħ > mhO. *yəfrɛ́ħ*). On a donc l'impression qu'un *ă en V_2 n'est pas suffisant pour faire remonter l'accent en V_0, alors qu'un *i l'est.

La même explication peut s'appliquer à jb. Gb SUBJ.3ms ´´γ *yəfráħ* vs. Ga SUBJ.3ms *yóḳdər* < *yaḳdu/ir, et il n'est pas nécessaire de supposer *yifraħ avec un *i dans le préfixe flexionnel. Quand bien même on poserait *yafraħ avec *a, l'exemple de *ȟandaxa > *həndɛ́x* ferait prédire *yafraħ > *yəfrɛ́ħ*[20]. Il est donc inutile d'expliquer la différence d'accentuation entre *yəfrɛ́ħ* et *yóḳdər* par le timbre originel de V_0, une telle explication ne rendant de toute façon pas compte de la différence d'accentuation entre *həndɛ́x* et *yəhándəx*.

*4.2.3. Le Ga INACC.3ms ´γ´ *yagăˁŭ/ĭr*

La morphologie attestée est la suivante:

	INACC.3ms	*INACC.3mp*
	*$yaC_1\alpha\gamma_2u/iC_3$	*$yaC_1\alpha\gamma_2u/iC_3\bar{u}$ (+ əm) [21]
Jb.	yɛ́gˁər	yəgˁór
Sq.	yesa̲ħam	yese̲ħem
MhO.	yərűħəṣ, yənűħəg	yərəħṣəm
MhŠ.		ynáħgäm
HbN.	ynáħəg, yráħəṣ, ygáˁər	ynáħgəm

[20] On pourrait objecter que *ȟandaxa comporte un *a final absent de *yafraħ et que l'accentuation en V_2 de *həndɛ́x* est peut-être due à sa présence. Le Ga ACC.3ms ´´γ *sabaħa > mhO. *sűbəħ* (et non **sabɛ́ħ*), accentué exactement comme Gb INACC.3ms *yVfɑraħ > mhO. *yəfűrəħ*, suggère qu'il n'en est rien.

[21] Le marquage du masculin pluriel par un suffixe *-əm* est une innovation que partagent, non sans une certaine variabilité dans son emploi, les dialectes mehri et le hobyot (cf. Lonnet 2005 ; Rubin 2015 : 321).

D'après notre hypothèse, la proto-forme *yagăʿŭ/ĭr contient deux voyelles « faibles », en V_1 et V_2, et un *a en V_0, ce qui rend *a priori* son accentuation problématique. Toutes les langues sauf le jibbali attestent un accent en V_1 au singulier comme au pluriel.

- En mehri[22], la voyelle accentuée a le même timbre que dans le schème par défaut du Ga INACC. (mhO. 3ms *yəkū́təb*, 3mp *yəkátbəm*; mhŠ. 3ms *yəkṓtəb*, 3mp *yəkátbəm*).
- Ce n'est pas le cas en hobyot (3ms *ykṓtəb*, 3mp *ykótbəm*), où le timbre /ā́/ en V_1 de *ynā́ħəg* est remarquable.
- La voyelle stable du soqotri *yesá̱ħam* atteste d'une accentuation identique à celle du mehri et du hobyot.
- Le jibbali semble encore une fois avoir moins de contraintes sur l'accentuation de V_0 que les autres langues, et a au singulier *yḗgʿər*, réalisation sans doute de **yégəʿər*. Au pluriel, il a *yəgʿór*, réalisation sans doute de **yəgəʿór*, avec une alternance accentuelle singulier ~ pluriel semblable à celle du subjonctif par défaut *yóḳdər* ~ *yəḳdór*.

On doit remarquer que l'alternance accentuelle au pluriel de ce type en jibbali ne se fait pas au profit d'un accent en V_1 (**yəgV́ʿər), alors même que cette accentuation est attestée par les autres langues. L'accentuation *yəgʿór* semble ainsi confirmer que c'est l'affection des timbres vocaliques par le*-ū final qui explique l'accentuation différentielle entre singulier et pluriel. En effet, en jibbali, de manière générale, l'affection vocalique de 2mp/3mp touche beaucoup plus souvent V_2 (par ex. Gb INACC.3ms *yəfeðór* ~ 3mp *yəfeðér*) que V_1 (cf. Ga INACC.3ms=3mp *yəḳódər*). Ce serait alors, dans *yagaʿu/irū, l'affection de V_2 par le suffixe *-ū qui l'aurait rendue accentuable et aurait empêché l'accentuation en V_0, mais également l'accentuation en V_1. Car, bien que la chose paraisse étonnante, force est de constater que l'effet d'un *-ū sur un *u/i en V_2 est de le faire « ressembler à un *a », puisque l'aboutissement en est alors /ó/, ici comme dans Ga SUBJ.3mp *yəḳdór* < *yaḳdirū (cf. également Gb SUBJ.3mp *yəfróħ* < *yVfraħū, où c'est un *ă qui est affecté).

À la 1p, le jibbali a *nəgʿár*, qui est difficilement explicable par pure évolution phonologique. Car en l'absence d'affection de V_2 par le suffixe *-ū on s'attendrait *a priori* à ce que l'hésitation accentuelle se soit faite entre *nagắʿu/ir > **nəgV́ʿər et *nágăʿu/ir > **nḗgʿər. Mais **nəgV́ʿər aurait été totalement isolé dans le paradigme, face à 3ms *yḗgʿər* et 3mp *yəgʿór*; **nḗgʿər* aurait eu pour lui de concorder avec le singulier, mais dans tous les types jibbalis à alternance accentuelle (et même dans quelques autres), la 1p s'est alignée sur les 2mp/3mp. C'est donc probablement la pression paradigmatique qui a prévalu et qui a produit *nəgʿár* à partir de 3mp *yəgʿór*, sur le modèle de Ga SUBJ. {1p *nəḳdér* ~ 3mp *yəḳdór*} ~ 3ms *yóḳdər*.

[22] Pour le mehri d'Oman : du moins chez les informateurs consultés.

5. Conclusions

5.1. Remarques sur les correspondances entre phonèmes vocaliques

On peut résumer ainsi les évolutions rencontrées pour les voyelles toniques:

1	**a en syllabe fermée*	{jb. *ɔ́*, sq. *o̲*, mhO. *ő*, mhŠ. *ő*, hbN. *ő*}	*yVC$_1$C$_2$aC$_3$ *yVC$_1$αC$_2$aC$_3$ jb. *yaC$_1$C$_2$u/iC$_3$
2	**a en syllabe ouverte*	{jb. *é*, sq. *a̲*, mhO. *ű́*, mhŠ. *ű́*, hbN. *ű́*}	*C$_1$aC$_2$aγ$_3$a jb. *yaC$_1$αC$_2$u/iC$_3$ (*yə́gˁər*)
3	**α, ...*	{jb. *ɔ́*, sq. *o̲/a̲*, mhO. *ű́*, mhŠ. *ő*, hbN. *ő́*	*yVC$_1$αC$_2$u/iC$_3$ *yVC$_1$αC$_2$aγ$_3$ *C$_1$aC$_2$aC$_3$a
4	**a en contexte *_Ci*	{jb. *é*, sq. *e̲*, mhO. *ű́*, mhŠ. *ű́*, hbN. *ű́*}	– *C$_1$aC$_2$iC$_3$a
5	**i, *u/i, *ă*	{jb. *ɛ́*, sq. *ɛ̲*, mhO. *ə́*, mhŠ. *ə́*, hbN. *ə́*}	*C$_1$aγ$_2$iC$_3$a *yVC$_1$C$_2$aγ$_3$ sq., mh., hbN. *yVC$_1$C$_2$u/iC$_3$ jb. *nVC$_1$C$_2$u/iC$_3$ (1p *nəḳdér*)

Ces correspondances entres phonèmes vocaliques des différentes langues avaient déjà été relevées par Hahn (2012), qui posait pour chacune d'entre elles un proto-phonème différent. Le tableau ci-dessus s'efforce d'aller plus loin, en remontant à des voyelles supposables pour le protosémitique. Les évolutions et surtout les conditionnements indiqués sont contestables à ce stade, dans la mesure où ils reposent sur un nombre d'exemple à chaque fois très limité. Ils ne pourront être confirmés que par un examen des formes dérivées.

Celles-ci fourniront d'autres attestations des mêmes cas de figure, par exemple:

- (1) T1 INACC.3ms Ð *yVśtaðar > jb. *yəśteðór* 'il se raidit';
- (2) T1 INACC.3ms ´´γ *yVḳtaṭaˁ > jb. *yəḳtéṭəˁ* 'il se coupe', mhO. *yəḳtűməħ* 'il est déçu';
- (4) T1 SUBJ.3ms Ð *yVśtaðir > jb. *yəśtéðər*, mhO. *yərtī́kəz* 'il se tient droit';
- (5) Š1 ACC.3ms ´´γ *stanfăˁa > jb. *šənfáˁ* 'il a profité';

La correspondance (3) requiert un traitement particulier. Elle serait identique à (1) si le mehri d'Oman n'y avait *ű́*, c'est-à-dire la même voyelle qu'en (2). Les problèmes posés par cette correspondance (3) lorsqu'elle apparaît en V_2 (par exemple ici au Ga ACC.3ms *C$_1$aC$_2$aC$_3$a) sont trop complexes pour être abordés ici. Il faut dire un mot

cependant de cette correspondance là où on la rencontre en V_1, c'est-à-dire là où elle est l'aboutissement de l'élément baptisé ici *α.

*5.2. L'élément *α et la formation de l'inaccompli*

Cet élément *α apparaît en V_1 comme {jb. *e*, sq. *ɛ*, mhO. *Ø,* mhŠ. *Ø,* hbN. *a*} quand il est atone, et comme {jb. *ó*, sq. *o̲/a̲*, mhO. *ū,* mhŠ. *ō,* hbN. *ō*} quand il est accentué. Il sert à la formation de l'inaccompli des formes simples Ga et Gb, mais aussi des formes dérivées suivantes:

- Ȟ1 (ou « causatif », jb. 3ms Ð *yəffelót* 'il s'enfuit', ´´γ *yəffógəˁ* 'il effraie');
- Š1 (ou « causatif-réfléchi », jb. 3ms Ð *yəšəḳeṣór* 'il est à court de', ´´γ *yəšenúfəˁ* ← **yəšenófəˁ*[23] 'il profite');
- ᵠȞ1 (quadrilitère, jb. 3ms Ð *yəġɛðefór* 'il jette', ´´´γ *yə̄rórḳəħ* ← **yəmerórḳəħ* 'il aplanit');
- ᵠN1 (quadrilitère, jb. 3ms Ð *yəngērdóś* ← **yəngeredóś* 'il tombe', ´´´γ *yənbelóṭəħ* 'il se cache').

Nous rencontrons ici la délicate question de la formation de l'inaccompli en sudarabique moderne et, partant, en sémitique. Pour le sudarabique moderne, la plupart des auteurs[24] ont vu dans le mehri *yəkū́təb* la même formation que le duratif akkadien *iparras* et que l'imparfait guèze *yənäggər* (en supposant soit que le sudarabique moderne a perdu la gémination de C_2, soit qu'elle est au contraire secondaire en akkadien et en guèze). Cette hypothèse est généralement faite à partir de formes isolées, sans que soit proposée une phonologie historique du sudarabique moderne rendant compte de l'ensemble de la morphologie (ne serait-ce que des formes simples). En outre, le problème que pose l'absence de concordance entre akkadien, guèze et sudarabique moderne pour ce qui est des timbres vocaliques en V_2 (selon les temps verbaux et selon les classes morphologiques) n'est généralement pas soulevé. Enfin, le jeu des allomorphies n'étant pas élucidé, l'existence de l'élément *α en mehri ailleurs qu'au Ga n'est le plus souvent pas repérée.

David Cohen (1972: 49-52; 1973; 1974; 1984: 68-76) est pratiquement le seul à proposer une explication de la morphologie des formes simples du sudarabique moderne en tant que système. Il le fait au moyen d'une hypothèse très différente, à partir d'une règle évolutive fondée sur des processus de resyllabification attestés dans de nombreux dialectes arabes. Il souligne que l'on peut faire dériver une forme comme mehri INACC.3ms *yəkū́təb* d'un étymon similaire à ar. cl. *yaktubu* sans qu'il soit nécessaire d'invoquer le guèze *yəḳättəl* ni l'akkadien *iparras*. Le cœur de son hypothèse vient de la constatation que de nombreux dialectes arabes supposent en diachronie et souvent même attestent en synchronie un traitement différentiel des voyelles brèves, les voyelles hautes /i/ et /u/ chutant en syllabe ouverte tandis que la voyelle basse /a/ se maintient. Dans ces

[23] L'élévation /ɔ/→/u/ au contact d'une nasale est phonologiquement régulière en jibbali.
[24] Pour un exposé détaillé de l'histoire de cette controverse, voir Lonnet 2017.

dialectes, une forme comme Ga INACC.3mp *yiktubū* est impossible et l'on a à la place la forme resyllabifiée *yikutbū*, tandis que INACC.3ms *yiktub* est stable, de même que sont stables Gb INACC.3ms *yilbas* et INACC.3mp *yilbasū*. Cohen suppose qu'un suffixe *-u a pu avoir des effets similaires dans la préhistoire du sudarabique moderne. Ce suffixe *-u ne serait pas cette fois le morphème opposant le masculin pluriel au singulier (comme dans l'inaccompli des dialectes arabes ci-dessus), mais le suffixe vocalique employé par au moins certaines langues ouest-sémitiques pour produire un inaccompli (par ex. ar. cl. INACC.3ms *yaktub-u*) à partir de l'ancienne conjugaison en yVktVb, cette dernière restant en usage (par ex. ar. cl. *yaktub*) avec diverses valeurs – et recevant diverses dénominations: jussif, apocopé, subjonctif. Cette supposition permet d'expliquer la morphologie des formes simples en jibbali de la manière suivante (Cohen 1984: 73):

	I	*II*	*III*	*IV*
[Ga ACC.3ms]	*kataba	*kataba	*katab	ketób
[Ga INACC.3ms]	*yiktubu	*yikutbu	*yikutb	ikóteb
[Ga SUBJ.3ms]	*yiktub	*yiktub	*yiktub	iktéb
[Gb ACC.3ms]	*rikiba	*rikba	*rikb	ríkeb
[Gb INACC.3ms]	*yirkabu	*yirkabu	*yirkab	irkób
[Gb SUBJ.3ms]	*yirkab	*yirkab		

Cette hypothèse suppose une chute (ou une resyllabification) des voyelles hautes en syllabe ouverte (II), suivie de la chute des voyelles finales (III). Elle suppose, dans un stade ultérieur, une disjonction des groupes consonantiques finaux (IV) comparable à la « ségolisation » de l'hébreu et d'autres langues sémitiques (y compris de nombreux dialectes arabes). Elle admet implicitement une règle accentuelle fixant l'accent sur la syllabe finale au stade III (ou sur la voyelle en V_2 au stade I avec recul de l'accent en cas de chute de la voyelle ou resyllabification). Elle est conçue pour rendre compte d'un système final où il n'y a pas d'opposition entre l'inaccompli et le subjonctif dans les verbes de type Gb (*irkób*).

Dans cette hypothèse, ce que j'ai nommé « élément *α » ne serait que le produit d'une resyllabification. C'est au départ cette hypothèse qui m'avait servi de fil d'Ariane. Deux points cependant m'empêchent aujourd'hui d'y adhérer. D'une part, on sait maintenant qu'il n'est pas vrai que le subjonctif et l'inaccompli des Gb soient identiques. Cela n'est vrai qu'en mehri, et même là seulement pour les allomorphes par défaut[25]. Partout ailleurs, l'élément *α est bien visible à l'inaccompli des Gb (jb. Ð *yəfeðór*), mais

[25] Pour les formes jibbalies du tableau ci-dessus, des resyllabifications superficielles dues à la liquide /r/ jointes à des imprécisions dans les transcriptions ont obscurci la morphologie du Gb : le subjonctif est bien *yərkób*, mais l'inaccompli est *yērkób* ← **yerekób*.

aussi à l'inaccompli des formes Ĭ1, Š1, ^{Q}Ĭ1 et ^{Q}N1. Or toutes ces formes ont un *a en V_2, qui n'est pas susceptible d'avoir été syncopé; il ne peut donc pas y avoir eu de resyllabification. D'autre part, on peut montrer dans les noms (Dufour 2017), mais aussi dans certaines formes verbales (Ga ACC.1s *ḳadarku > jb. *ḳədórk* et non **ḳədórək*) qu'il n'y a pas eu de ségolisation en sudarabique moderne (sauf dans le cas particulier où la consonne finale est une liquide). L'hypothèse Cohen devient donc très coûteuse à maintenir s'agissant de l'inaccompli, et sa vertu explicative devient assez faible puisqu'elle ne rendrait compte finalement que de l'inaccompli du Ga.

Il est donc plus économique de supposer la présence originelle d'un morphème en V_1 pour expliquer l'élément *α. Mais c'est là que les problèmes commencent. Car si mhO. *yəkū́təb* remonte à *yVkatu/ib, sans gémination de C_2, alors pourquoi le produit de *α accentué est-il, dans toutes les autres langues, différent de celui de *a accentué en syllabe ouverte (cf. 5.1. *supra*) ? Peut-être alors doit-on plutôt poser *yVkattu/ib avec gémination. Mais d'une part, cela implique au passage une deuxième supposition qui, pour plausible qu'elle soit, demanderait démonstration: c'est que la gémination en question a historiquement disparu en sudarabique moderne. Et d'autre part, pour que *yVkattu/ib ne soit pas une supposition *ad hoc*, il faudrait trouver aux évolutions phonétiques que cela implique une confirmation ailleurs dans la morphologie. On pourrait argüer de ce que le produit de *α accentué est partout ailleurs qu'en mehri d'Oman identique à celui de *a en syllabe fermée et non ouverte. Il est vrai que cela plaide en faveur d'une syllabe historiquement fermée, et on devrait alors simplement supposer que le mehri d'Oman fait exception, et que *áC_iC_iV y a évolué comme *áCV. Il nous faudrait alors chercher un autre exemple de *áC_iC_iV pour vérifier la régularité de l'évolution supposée. Les autres formes où l'on peut – avec plus ou moins de certitude – supposer une ancienne gémination de C_2 sont la forme « intensive » (mhO. SUBJ.3ms *yarṓkəb* < *yVĭ̆harakkib ?) et la deuxième forme à *-t-* infixé (mhO. SUBJ.3ms Ð *yəftkū́r*, ´´γ *yəmtṓdəħ* < *yVmtaddaħ). Or en mehri d'Oman la voyelle accentuée en V_1 dans ces formes est *ṓ*, et non pas *ū́* comme le produit de *á.

Il y a donc des arguments aussi bien contre *yVkatu/ib que contre *yVkattu/ib, et le cas est indécidable dans l'état actuel de l'exploration de la phonologie historique du sudarabique moderne[26]. L'opposition *ṓ*~*ū́* du mehri d'Oman, qui prend souvent à rebours

[26] On pourrait, en faveur d'une gémination de C_2, tirer argument du fait que V_1 n'est nulle part palatalisée dans ces inaccomplis (conçus comme issus de *yVkattib) alors qu'elle l'est dans mhO. Gb ACC.3ms *fī́ðər* < *faðira (vs. Ga ACC.3ms ´´γ *sū́bəħ* < *sabaħa sans palatalisation) : la différence de traitement de la voyelle serait due à la gémination de la consonne subséquente. Cependant, cet argument ne vaut que si l'on suppose un *i en V_2 (*yVkattib). Or dans la mesure où la voyelle en V_2 à l'inaccompli des formes simples (Ga et Gb) paraît en SAM correspondre régulièrement à celle du subjonctif, la comparaison sémitique suggère qu'elle a pu être aussi bien *u que *i, et peut-être même plus souvent *u que *i (cf. Kogan 2015a et la doxographie qu'il établit). L'absence de palatalisation est peut-être simplement due à la généralisation du type en *u aux dépens de celui en *i à l'inaccompli comme au subjonctif des Ga, voie empruntée également par l'hébreu et l'araméen dans leurs conjugaisons préfixales. Les formes de duel du soqotri (accentuées en

la leçon des autres langues du groupe, est visiblement au cœur du problème. La seule certitude est que l'élément *α est bien un morphème, et qu'il doit donc être considéré dans le cadre des procédés morphogénétiques des langues sémitiques consistant en l'insertion de matériel morphologique en V_1. On peut donc le *comparer* à l'akkadien *iparras* ou au guèze *yənäggər*, mais l'*identifier* à celui-ci, à celui-là ou aux deux à la fois est un pas supplémentaire qui mérite démonstration et que la leçon du sudarabique moderne n'impose pas.

En outre, il reste à comprendre pourquoi l'élément *α est absent dès que la racine est « trop courte » et oblige à une réduplication, sélectionnant dans ce cas un allomorphe de schème spécial. C'est le cas avec les racines bilitères pour les formes:

- G (Ga et Gb étant confondus, jb. 3ms *yədlél* 'il indique'),
- Ȟ1 (jb. *yəkkbéb* 'il fait descendre') et
- Š1 (jb. *yəšəmdéd* 'il attrape'),

ainsi qu'avec les racines bilitères et trilitères pour les formes:

- ^{Q}Ȟ1 (jb. *yədəfdéf* 'il caresse', *yəzəħlél* 'il glisse') et
- ^{Q}N1 (jb. *yənḳəśḳéś* 'il craque', *yənḳəfrér* 'il fait la moue').

L'akkadien, lui, a *iškuk / išakkak* comme *iprus / iparras*, et le guèze *yəsdəd / yəsäddəd* comme *yəngər / yənäggər*. L'inaccompli du sudarabique moderne atteste donc d'une histoire particulière, que des différences importantes séparent de l'akkadien comme de l'éthiopien.

5.3. Le développement d'un accent de mot

Ce qui de l'hypothèse Cohen ne peut, à mon sens, pas être maintenu, c'est la syncope et/ou resyllabification. Mais ce qu'il en demeure, c'est le traitement différentiel du vocalisme, *a ayant une proéminence supérieure à celle de *i et *u. Le scénario proposé dans le présent article implique qu'à un stade de son évolution l'ancêtre du sudarabique moderne a développé un phénomène prosodique donnant une proéminence particulière à une des voyelles de chaque mot. Au départ, cette proéminence semble avoir été fonction de la structure du mot et de la nature des voyelles[27]. Au vu de Ga ACC.3ms *ḳadara > jb. *ḳədór* et 3fs *ḳadarat > jb. *ḳədərót*, il apparaît que dans une forme ne comportant que des *a, c'est le *a non final le plus à droite qui a porté ce qui allait devenir un accent. Le Gb ACC.3ms *faðira > jb. *féðər* etc. nous a montré que la formulation « le *a non final le plus à droite » valait même lorsque ce *a était en V_1 et qu'il y avait en

V_3 en proto-SAM) paraissent confirmer cette dernière hypothèse, puisqu'elles ont en V_2 au subjonctif et à l'inaccompli du Ga (3ms *ʎibłégo̲h* / *yebɛłégo̲h*) un /ɛ/ qui contraste avec le /i/ des formes où la comparaison sémitique permet raisonnablement de poser un *i (par ex. Ȟ1 SUBJ.3ms *ʎándiḳoh*, cf. ar. Cl. IV *yufʿilā* ; Ȟ2 SUBJ.3ms *yesɛ́biroh*, cf. ar. Cl. II *yufaʿʿilā*, III *yufāʿilā*) : la voyelle thématique des conjugaisons préfixales du Ga ne se comporte donc pas comme un *i (Dufour 2006 : 334-345).

[27] Cela revient à supposer un *quality-sensitive stress* tel qu'étudié par Kenstowicz (1997).

V_2 une voyelle haute. En revanche, mh. *yəḳdə́r* face à jb. *yóḳdər* révèle que lorsque le *a non final le plus à droite était en V_0, la remontée de l'accent n'a pas été sans restrictions. Enfin, on a vu dans *dafaˁa > jb. *défəˁ* qu'un *a situé devant une gutturale se comportait souvent d'un point de vue prosodique comme les voyelles hautes, bien que dans certains cas il apparaisse davantage susceptible que ces dernières d'être accentué (par ex. jb. SUBJ.3ms Ga *yaḳdu/ir > *yóḳdər* mais Gb ´´γ *yVfraħ > *yəfráħ*).

Il y a ainsi deux échelles de préférences: une échelle de sonorité relative au timbre des voyelles (*a > *ă > *i, *u), et une échelle relative à leur position ($V_3 > V_2 > V_1 > V_0$)[28]. C'est lorsqu'il y a contradiction entre ces deux échelles qu'on voit apparaître des divergences entre les langues: Ga SUBJ.3ms *yaḳdu/ir > jb. *yóḳdər* mais mh. *yəḳdə́r*; Ga INACC.3ms ´γ´ *yagăˁu/ir > jb. *yə́gˁər* mais mhO. *yərū̆ħaṣ*. Le fait que le jibbali connaisse des alternances accentuelles à l'intérieur même du paradigme des formes concernées par ces divergences (3ms *yóḳdər* mais 1p *nəḳdə́r*, 3mp *yəḳdór*; 3ms *yə́gˁər* mais 1p *nəgˁár*, 3mp *yəgˁór*) laisse entendre que le paramétrage des poids relatifs des différentes voyelles selon les différentes positions a pu être complexe, avec au bout du compte des morphologisations diverses. Certaines formes isolées du jibbali sont visiblement les restes d'une situation antérieure à la morphologisation ayant aboutit au système aujourd'hui régulier. Ainsi dans le Ga SUBJ.3ms irrégulier *yəġréb* 'savoir', la gutturale en C_1 paraît avoir bloqué la remontée de l'accent en V_0 (*yăġrŭ/íb > *yəġréb*). Il est probable qu'une étude dialectologique du jibbali serait instructive à cet égard.

Mais ce qu'on doit également souligner, c'est qu'une fois que le développement d'un accent de mot a eu lieu, cet accent morphologisé a par la suite profondément affecté l'évolution du vocalisme. La place a manqué ici pour étudier les timbres des voyelles atones, mais on a vu, par exemple, que *α atone donnait en jibbali /e/, alors que *ά tonique donnait /ɔ́/; de même les voyelles de jb. SUBJ.3ms *yóḳdər* et 1p *nəḳdə́r* (< *yaḳdu/ir, *naḳdu/ir) ne sont aujourd'hui différentes que par suite d'une différence historique d'accentuation. On peut également montrer que dans toutes les langues modernes l'inventaire vocalique est plus riche en position tonique (ou, pour le soqotri, stable) qu'ailleurs. Le vocalisme des langues modernes est ainsi profondément conditionné par l'accent, alors même que celui-ci était initialement, à l'époque de son apparition, fonction du vocalisme hérité du protosémitique.

[28] Le conditionnement proposé ici suppose que l'unité de base servant de repère au placement de l'accent proto-SAM est la position consonantique et non pas la syllabe. Ainsi Ȟ1 ACC.3ms ´´γ *ȟandaxa (CVCØCVCV) n'est pas superposable à Ga ACC.3ms ´´γ *sabaħa (CVCVCV), et l'accentuation en est différente au terme de l'évolution : mhO. *həndə́x* (cf. 4.2.2) vs. *sū́bəħ*. Ce conditionnement peut être considéré comme agissant régulièrement à un niveau purement phonologique. Dans la mesure cependant où la morphologie verbale sémitique fait appel également à des positions vocaliques, repérées par rapport aux consonnes de la racine, un tel conditionnement phonologique a immédiatement des implications morphologiques et, en retour, on doit s'attendre *a priori* à ce que ses aboutissements soient parfois perturbés par des phénomènes évolutifs devant être expliqués à un niveau morphologique.

C'est ce nouvel accent de mot qui a donné au verbe sudarabique moderne son aspect caractéristique. En effet, les anciennes alternances vocaliques ont été converties en alternances accentuelles. Ainsi, de l'alternance *u/i~*a en V_2 exprimant l'opposition Ga~Gb aux conjugaisons préfixales (par ex. INACC.3ms *yVḳαdu/ir ~ *yVfαðar) il reste surtout aujourd'hui une alternance dans la place de l'accent: jb. *yəḳódər ~ yəfeðór*. Les exemples abondent également dans les formes dérivées: T1 SUBJ.3ms *yVśtaðir ~ INACC.3ms *yVśtaðar est devenu jb. *yəśtéðər ~ yəśteðór*.

Mais il y a plus. Les consonnes gutturales[29] ont eu soit sur les timbres vocaliques, soit directement sur la prosodie du mot un effet dont on ne peut plus guère aujourd'hui que constater le résultat. Bouleversant le placement de l'accent, elles ont donné naissance à ce qui s'est morphologisé sous la forme des allomorphies supplétives de schèmes qui parcourent le verbe sudarabique moderne. Le vocalisme ayant, depuis, connu des évolutions conditionnées par la place de l'accent, ces allomorphies ne sont plus explicables par la pure phonologie dans les langues modernes; elles relèvent désormais d'une information morphologique, même si ce sont les propriétés phonologiques de la racine qui déterminent le choix entre les allomorphes[30].

5.4. La vocalisation des préfixes flexionnels dans les formes simples

Dans les pages précédentes, on a autant que possible dans les proto-formes évité de se prononcer sur la nature de la voyelle des préfixes flexionnels au subjonctif et à l'inaccompli. Ce n'est que dans Ga SUBJ.3ms Ð *yaḳdu/ir et Ga INACC.3ms ´γ´ *yagαˁu/ir qu'on a écrit un *a, parce que les formes jibbalies correspondantes (*yóḳdər* et *yɛ́gˁər*) comportent la même voyelle que ce qui est ailleurs le produit de *a dans des positions comparables.

Or la question de la vocalisation de ces préfixes dans l'histoire des langues sémitiques est des plus débattues[31]. L'akkadien a, du moins dans sa morphologie majoritaire, des voyelles qui diffèrent suivant les personnes (*a-*, *ta-*, *i-*, *ni-*) mais qui ne sont pas sensibles au vocalisme de la base verbale. Les langues classées dans le rameau

[29] Ainsi que, avec des effets ressemblants, les glides.

[30] Hayward *& al.* 1988 ont souligné l'importance des gutturales pour le placement de l'accent en jibbali. Il faudrait plusieurs pages pour répondre point par point à ce travail précis, clairvoyant et qui a fait avancer notablement notre compréhension des faits. Ce qui lui manque est sans doute de ne pas énoncer clairement si les règles dégagées sont synchroniques ou diachroniques, et d'avoir cru qu'on pouvait considérer le vocalisme du jibbali moderne préalablement à la détermination de la place de l'accent, ce qui produit des raisonnements qui apparaissent comme circulaires si l'on veut bien reconnaître que les vocalismes toniques et atones obéissent aujourd'hui à des logiques très différentes. En outre, les auteurs ont négligé les types Ga ´γ´ et Gb ´´γ, qui leur auraient sans doute fourni la clé du système.

[31] Pour une présentation récente de la discussion, de son histoire et des principales références, voir Kogan 2015 (146-152).

sémitique central révèlent au moins à l'état de reliques une répartition différente: dans un paradigme donné, tous les préfixes ont la même voyelle, mais celle-ci est fonction du vocalisme de la base (**yaCCi/uC*, **taCCi/uC*, etc. mais **yiCCaC*, **tiCCaC*, etc.). L'éthiopien a /ə/ à toutes les personnes quel que soit le vocalisme de la base. Des scénarios divers ont été proposés pour rendre compte de cette situation. La leçon du sudarabique moderne est donc capitale, et le lecteur est en droit d'attendre qu'on la dégage ici. Hélas, disons tout de suite qu'elle est peu assurée.

La question est abordée par Kogan (2015b: 150-151), qui voit dans jb. Ga SUBJ.3ms *yóḳdər* la marque d'une vocalisation *yaḳdu/ir, et considère que le /ə/ des formes jibbalies et mehries où le préfixe est inaccentué ne sont pas pertinentes, le timbre de la voyelle étant probablement dû à sa position atone (cf. jb. 3mp *yəḳdór* face à 3ms *yóḳdər*). Il paraît clair en effet que, d'une manière générale, la plupart des voyelles atones du mehri sont passées à /ə/ ou zéro; or le mehri n'accentue jamais le préfixe personnel, et ses formes sont donc peu probantes. Les autres langues ont mieux conservé les voyelles atones (cf. l'exemple du *α atone, préservé partout ailleurs qu'en mehri). Cependant la vocalisation des préfixes flexionnels en hobyot est normalement /ə/ pour les Ga comme les Gb, Nakano (2013) notant même zéro en syllabe ouverte: Ga SUBJ.3ms *yəktə̄b* mais INACC.3ms *ykə̄təb*. Le hobyot n'est donc pas d'une plus grande aide que le mehri.

5.4.1. La voyelle des préfixes flexionnels en jibbali

En jibbali, *yóḳdər* implique visiblement *a au subjonctif du Ga. Quant à *yə̄gˁər*, il est intéressant puisqu'il implique *yagαˁu/ir, et montre donc que l'inaccompli des Ga a un *a tout comme le subjonctif. Ailleurs, la voyelle du préfixe est atone et donc peu probante[32]. Certaines formes parlent cependant en faveur de *a dans les Ga. En syllabe ouverte atone en effet, la voyelle du préfixe est souvent réalisée /e/ devant une consonne liquide en C_1: INACC.2ms *telóḳəṭ* plutôt que *təlóḳəṭ* 'tu cueilles'. Or /e/ est souvent ailleurs le produit de *a atone en syllabe ouverte, cf. le type Ga ACC.3ms avec liquide en C_2 *derós* 'étudier' < *darasa, ou la forme dérivée T1 INACC.3ms *yəśteðór* < *yVśtaðar, ou encore des schèmes nominaux comme celui de *bedén* 'corps' (cf. arabe *badan* 'dº'). Un phénomène semblable affecte le préfixe *šə-* des formes dérivées Š1 et Š2: normalement réalisé *š(ə)-*, il devient *še-* devant liquide (Rubin 2014: 182-184); or ce préfixe est probablement issu de *šta- avec un *a.

Pour les Gb du jibbali, c'est toujours le *a en V_2 qui est accentué dans les conjugaisons préfixales (ou, à défaut, le *α en V_1 dans les inaccomplis à gutturale ou glide finaux), et le préfixe est donc toujours atone. On a vu également (4.2.2.) qu'on ne pouvait rien conclure de l'accentuation en V_2 du Gb SUBJ.3ms ´´γ *yəfráħ* concernant la

[32] Sauf dans le subjonctif Ga à glide final, type *yéḳər* (√ḳry) 'qu'il lise', dont la voyelle en V_1 est difficile à expliquer, mais où le préfixe suppose *ya-. Les formes des racines w´´ seront étudiées plus bas.

vocalisation originelle du préfixe: *yəfráħ* autorise à poser *yifraħ, mais peut aussi bien être expliqué comme l'aboutissement régulier de *yafraħ. Dans les inaccomplis Gb cependant, quand C_1 est une liquide, on a *yērkób* 'il monte sur' qui suppose **yerekób*, donc probablement *yarɑkab. Les préfixes flexionnels auraient donc un vocalisme *a dans les Gb exactement comme dans les Ga. On aimerait cependant avoir plus d'arguments pour le montrer.

La leçon du soqotri est difficile à exploiter dans l'état actuel de notre compréhension de la phonologie historique de la langue. À côté de nombreuses réalisations à vocalisme réduit [ə], des variantes en [e] suggèrent une situation proche de celle du jibbali.

5.4.2. Les verbes à première radicale /w/

Kogan (2015b: 151) évoque le cas des SUBJ.3ms w´´ du mehri d'Oman: Ga *yāzḗm* et Gb *yəwṣṓl*. Il souligne avec raison que *yəwṣṓl* n'implique pas un préfixe *yi-. On pourrait ajouter que *yāzḗm* n'implique pas davantage *ya-[33]. Le *ā* long prétonique de cette forme suppose une consonne sous-jacente (en l'occurrence certainement /w/), mais la phonologie des voyelles longues atones du mehri n'est pour l'instant pas suffisamment comprise pour qu'il soit possible d'en dire plus.

Ces formes doivent cependant être remises dans le cadre de la morphologie des autres langues sudarabiques modernes, afin que n'en soient pas tirées des conclusions erronées:

		ACC.3ms	*INACC.3ms*	*SUBJ.3ms*	
Jibbali	Ga	eḳóf, ḳóf	yēḳóf	yéḳəf (*1p* nəḳéf, *2mp* yəḳóf)	*se taire*
	Gb	éṣəl	yēṣól	yəṣól	*arriver*
Soqotri	Ga	édof	yóudof	ʎidéf	*saisir*
	Gb	(ékod)	yóukod	ʎikód	*avoir peur*
Mehri d'Oman	Ga	wəfűd	yəwfőd	yāfḗd	*demander en mariage*
	Gb	wḯṣəl	yəwṣṓl	yəwṣṓl	*arriver*
Hobyot N.	Ga	wsők	yəwsők	yesḗk	*chiquer*
	Gb	wīṣəl	yəwṣṓl	yəṣṓl	*arriver*

Dans toutes les langues, les Ga et les Gb w´´ ont normalement des inaccomplis identiques[34], qui évoquent les inaccomplis Gb des racines saines.

[33] Il en va de même, *pace* Kogan (2015 : 151 note 419), de Ga SUBJ.3ms *yaʔmḗr* (√ʕmr), la coloration [a] d'un /ə/ étant un effet régulier de /ʕ/ sous-jacent en mehri.

[34] Sauf pour une poignée de verbes (principalement mhO. INACC.3ms / SUBJ.3ms *yəwűkəb / yākḗb*

En jibbali, la voyelle longue de *yēḵ́f*, *yēṣ́l* suppose une contraction, et donc **yəweḵ́f*, **yəweṣ́l*, avec chute régulière du /w/ intervocalique. On peut donc poser **yVwɑḵaf*, **yVwɑṣal*. Ces mêmes étymons sont sans doute à la base des formes du soqotri, avec leur diphtongue /ou/, ainsi que de celles du hobyot. Dans cette dernière langue, l'absence de réflexe visible du **α* est explicable. En effet, alors que **α* atone y devient normalement /a/ (**yVfɑðar* > *yfaðṓr*), ce /a/ disparaît régulièrement lorsque C_1 est une liquide (**yVlɑḵaf* > **yəlaḵ́f* → *yəlḵ́f* 'il attrape'), suivant un processus proche de ce que connaît le jibbali (**yerekə́b* → *yērkə́b*). Ici, /w/ se comporte visiblement comme une liquide, et on a **yəwasṓk* → *yəwsṓk*. En mehri, il est normal que **α* atone aboutisse à zéro.

Au subjonctif, le jibbali suppose Ga **yaḵu/if* et Gb **yVṣal*. Le soqotri suggère également des formes sans **w*. La comparaison sémitique indique qu'une telle morphologie est ancienne (cf. ar. cl. Ga *yaqif* √wqf 's'arrêter', malgré Gb *yawrab* √wrb 'périr').

Au subjonctif du mehri, le **w* apparaît comme tel dans Gb *yəwṣṓl*, mais Ga *yāfḗd* lui aussi suggère un glide sous-jacent. Ces formes, comparées à celles archaïques du jibbali et du soqotri, sont *a priori* suspectes d'être des réfections (sur lesquelles il est donc risqué de se fonder pour évaluer la vocalisation originelle des préfixes). On est simplement étonné que la restitution du /w/ de la racine ait donné des résultats différents dans *yəwṣṓl* et dans *yāfḗd*. Le corollaire de cette dissymétrie est que, pour les Gb, le subjonctif est homophone de l'inaccompli.

Au subjonctif Ga du hobyot, il est important de noter que Nakano transcrit régulièrement *yəzɑ̄́m*, *yeḵɑ̄́ṣ*, etc. avec *yə-* ou *ye-* et non pas ***yzɑ̄́m*, etc. Il ne s'agit donc pas ici du simple préfixe /yə-/, lequel est toujours transcrit *y-* en syllabe ouverte (*ykṓtəb*, *yfaðṓr* etc.). Ce *ə/e* contient donc un élément sous-jacent (sans doute le **w* de la racine), et *yesḗk* est l'équivalent de mhO. *yāfḗd*, et non pas des formes soqotries et jibbalies sans **w*.

Il en va de même au subjonctif Gb du hobyot, qui est *yəṣṓl* et non pas ***yṣṓl*, et qui ne remonte donc pas à **yVṣal*. Le hobyot atteste donc très probablement, comme le mehri, la restitution d'un /w/ dans les formes de subjonctif. Mais il est remarquable que, à la différence du mehri, il oppose, pour les Gb, l'inaccompli *yəwṣṓl* au subjonctif *yəṣṓl*. Dans la mesure où ce dernier comporte bel et bien un glide sous-jacent, la seule façon d'expliquer la différence de forme entre les deux temps est de considérer que le subjonctif *yəṣṓl* vient de **yVwṣal*, tandis que l'inaccompli *yəwṣṓl* vient de **yVwɑṣal* avec **α* en V_1 (ce qui corrobore la supposition faite plus haut sur *yəwṣṓl* comme issu de **yəwaṣṓl*).

Le subjonctif Gb *yəṣṓl* du hobyot, parallèle comme il est au subjonctif Ga *yəzɑ̄́m*, est donc l'exact équivalent d'un mehri ***yāṣṓl* qu'on s'étonnait de ne pas trouver face au subjonctif Ga *yāfḗd*. Il est donc très probable que le subjonctif *yəwṣṓl* du mehri est refait sur l'inaccompli et a éliminé un ancien ***yāṣṓl* dont le cognat demeure en hobyot. En

'entrer' et *yəwű́zəm* / *yāzḗm* 'donner'). En outre, certains verbes ont un accompli de type Gb et un subjonctif de type Ga : hbN. *wī́rəθ* / *yəwrṓθ* / *yərḗθ* 'hériter', sq. *érət* / *yóurot* / *ʎirét* 'dº'.

effet, en mehri, dans la morphologie par défaut des Gb, et suite à la chute du produit de *α atone, inaccompli et subjonctif étaient devenus identiques (*fī́ðər / yəfðṓr / yəfðṓr*). La morphologie des Gb w´´ (*wī́ṣəl / yəwṣṓl / **yāṣṓl*) contrevenait dès lors au syncrétisme établi entre inaccompli et subjonctif pour cette classe morphologique, et devenait vulnérable à une réfection analogique. Une telle analogie n'avait en revanche aucune raison de se produire en hobyot où l'inaccompli Gb reste distingué du subjonctif par la présence en V_1 du produit de *α (*fī́ðər / yfaðṓr / yəfðṓr*).

Un autre indice va dans le même sens. Le type Gb w´ɣ en mehri d'Oman (ACC.3ms *wī́ḳa* √wḳˁ 'être') a en effet un subjonctif de forme *yāḳā́* et non pas **yəwḳā́*. *yāḳā́* représente certainement la forme originelle. Ici, la présence de la gutturale finale avait régulièrement conditionné {*yVwαḳaˁ ~ *yVwḳaˁ} > {*yəwṓḳa ~ yāḳā́*}. De ce fait, ailleurs qu'au parfait, la morphologie de ce type ne ressemblait plus à celle du Gb par défaut (elle coïncidait plutôt avec celle des Ga), et l'analogie n'avait pas lieu de s'exercer.

Le type SUBJ. Gb *yəwṣṓl* du mehri est donc un faux subjonctif, et ce qui oppose sa première syllabe à celle de SUBJ. Ga *yāḳā́f* remonte historiquement non pas à une opposition Gb ~ Ga mais à une opposition inaccompli ~ accompli (présence ~ absence de l'élément *α).

Les verbes w´´ ne révèlent donc aucune différence de vocalisation des préfixes flexionnels entre les types Ga et Gb. On doit donc en rester à la conclusion précédemment énoncée: dans les rares cas où le sudarabique moderne révèle la vocalisation originelle des préfixes, il semble qu'on ait partout un *a, bien que la chose soit mieux assurée pour les Ga que pour les Gb.

BIBLIOGRAPHIE

Bar-Asher, E. A. (2008), *The imperative forms of Proto-Semitic and a new perspective on Barth's Law*, in "Journal of the American Oriental Society", 128: 1-23

Bendjaballah, S., Ségéral, P. (2014), *The Phonology of "Idle Glottis" Consonants in the Mehri of Oman (Modern South Arabian)*, in "Journal of Semitic Studies", 59: 161-204

Cohen, D. (1972), *Problèmes de linguistique chamito-sémitique*, in "Revue des études islamiques", 40/1: 43-68

Cohen, D. (1973), *Sémitique comparé*, in "Annuaire de l'École pratique des hautes études", IVe section, 1973-1974, p. 175-191

Cohen, D. (1974), *La forme verbale à marques personelles préfixées en sudarabique moderne*, in "IV Congresso internazionale di studi etiopici (Roma, 10–15 aprile 1972)", Rome, Accademia Nazionale dei Lincei, vol. 2, p. 63-70

Cohen, D. (1984), *La phrase nominale et l'évolution du système verbal en sémitique. Études de syntaxe historique*, Paris / Louvain, Peeters: 68-76

Dufour, J. (2016), *Recherches sur le verbe sudarabique moderne*, mémoire présenté pour l'obtention de l'habilitation à diriger les recherches, sous la direction de Gilles Authier, Paris, École pratique des hautes études

Dufour, J. (2017), *Nouns and adjectives of the shape* $C_1\acute{V}C_2(ə)C_3(-)$ *in Jibbali (Śḥri) and Mehri*, in "Bulletin of Afro-Asiatic Languages and Linguistics", 9/2-1: 191-217

Hahn, M. (2012), *Proto-Modern South Arabian vowels – A first approximation*, travail inédit mis en

ligne sur https://www.academia.edu/12966895/Proto-Modern_South_Arabian_Vowels_A_First_Approximation
Hasselbach, R., (2004), *The markers of person, gender, and number in the prefixes of G-preformative conjugations in Semitic*, in "Journal of the American Oriental Society", 124: 23-35
Hayward, K. M., Hayward, R. J., Bakhit Al-Tabuki, S. (1988), *Vowels in Jibbali verbs*, in "Bulletin of the School of Oriental and African Studies", 51/2: 240-250
Huehnergard, J., Rubin, A. (2011), *Phyla and Waves: Models of Classification of the Semitic Languages*, in S. Weninger, G. Khan, M. P. Streck, J. C. E. Watson (éds), *The Semitic Languages. An International Handbook*, Berlin / Boston, De Gruyter Mouton: 259-278
Johnstone, T. M. (1975), *The Modern South Arabian languages*, in "*Afroasiatic Linguistics*", 1/5: 1-29
Johnstone, T. M. (1981), *Jibbāli Lexicon*, Londres, Oxford University Press
Kenstowicz, M. (1997), *Quality-sensitive stress*, in "Rivista di linguistica" 9/1: 157-187 (numéro daté de 1996, article daté de 1997)
Kogan, L. (2015a), *Observations on Proto-Semitic vocalism*, in "Aula Orientalis", 23: 131-167
Kogan, L. (2015b), *Genealogical Classification of Semitic. The Lexical Isoglosses*, Boston / Berlin, De Gruyter
Lonnet, A. (2005), *Quelques réflexions sur le verbe sudarabique moderne*, in A. Mengozzi (éd.), "Studi Afroasiatici: XI Incontro Italiano di Linguistica Camitosemitica", Milan, Franco Angeli: 187-201
Lonnet, A. (2017), *Modern South Arabian* ikōtəb *is not necessarily* iparras *or* yənaggər, in "Bulletin of Afro-Asiatic Languages and Linguistics", 9/2 (à paraître)
Nakano, A. (2013), (éd. R. Ratcliffe), *Hōbyot (Oman) Vocabulary with Example Texts*, Tokyo, Institute for the Study of Languages and Cultures of Asia and Africa
Naumkin, V., Bulakh, M., Cherkashin, D., Kogan, L., Issa, A., Gumaan, I. (2014a), *Studies in the verbal morphology of Soqotri I/1: Strong triconsonantal roots in the basic stem (the analysis)*, in "Zeitschrift für arabische Linguistik", 59: 25-56
Naumkin, V., Bulakh, M., Cherkashin, D., Kogan, L., Issa, A., Gumaan, I. (2014b), *Studies in the verbal morphology of Soqotri I/2: Strong triconsonantal roots in the basic stem (the lexical data)*, in "Zeitschrift für arabische Linguistik", 60: 25-56
Naumkin, V., Kogan, L., Gumʿan al-Daʿrhi, ʿI, ʿIsa al-Daʿrhi, A., Cherkashin, D., Bulakh, M., Vizirova, E. (2014), *Corpus of Soqotri Oral Literature*, vol. 1, Brill, Leyde
Rubin, A. (2014), *The Jibbali (Shaḥri) Language of Oman*, Brill, Leyde
Rubin, A. (2015), *The classification of Hobyot*, in A. M. Butts (éd.), *Semitic Languages in Contact*, Leyde, Brill: 311-32

JULIEN DUFOUR • is an associate professor of Arabic at the University of Strasbourg, and a member of the OmanSaM research project on Modern South Arabian (MSA) languages. His main research interests are Yemeni Middle-Arabic poetry, Yemeni Arabic dialects, MSA languages and the Occitan dialects of Corrèze. Most of his publications deal with Yemeni Arabic. He has recently defended a habilitation dissertation on Jibbali and the MSA verb (2016) based on published data as well as on extensive fieldwork.

E-MAIL • espagnac@outlook.com

PHONETICS OF EMPHATICS IN BAṬḤARI*

Fabio GASPARINI

ABSTRACT • Baṭḥari is one of the six Modern South Arabian languages spoken in Oman and Yemen and belonging to the Western branch of the Semitic family. Once supposedly spoken in a wider area at the extreme East of Dhofar region, recent fieldwork revealed that only 12 elders from the Baṭaḥira tribe remember the language at various degrees of proficiency. It is very likely that Baṭḥari will disappear in less than a couple of decades with the death of the last speakers and no action other than documentation and description can be made to secure memory of this language. The present paper presents the first results of an ongoing study conducted over a corpus sample of oral texts collected by prof. Miranda Morris and Khalifa Hamood al-Baṭḥari for the "Documentation and Ethnolinguistic Analysis of the Modern South Arabian languages" project. The concurrence of two phonetic processes in the Baṭḥari phonetic system, namely pharyngealization and glottalization, will be explored in detail. These two processes are involved in the realization of the so-called "emphatics", which are known to show both kinds of realization in various contexts in MSAL (Watson & Bellem 2011; Ridouane & al. 2015). The analysis of Morris' sample was followed by a period of fieldwork in Autumn 2016, in order to gather more material and gain a better knowledge on this heavily endangered language.

KEYWORDS • Modern South Arabian Languages, Baṭḥari, Emphatics, Glottalization, Pharyngealization

1. Introduction

Emphatics in Semitic languages are well known to be a phonologically contrastive class. Their realization though is not uniform across the language family. In Arabic they are usually realized as pharyngealized/velarized sounds, while in the Ethio-Semitic

* I want to sincerely thank Giuliano Castagna for valuable and constant feedback and Massimo Radin for helping me with the statistical analysis provided for this paper. Valuable help on the field in organizing the interviews and translating the material from Baṭḥari to Arabic came from Khalifa Hamoud alBaṭḥari, whom I hereby thank – together with the whole Baṭaḥira community - and without whom the present study would have never taken place. I also thank Miranda Morris, whose kindness and willingness to assist me in my attempt at knowing more about Baṭḥari was critical. Any misjudgement or error whatsoever is of course solely mine to blame.

family they tend to show ejective realization (Heselwood 1996: 26-27; Shosted & Rose 2011). The conventional use of a subscript dot in the phonological transcription of Semitic emphatics can be problematic at times, since it indirectly misleads scholars' attention from the real phonetic realization of such segments. It is surely of high interest the areal case of Yemeni and southern Omani languages, in which both processes (pharyngealization and glottalization) are involved at a phonemic level (Watson & Bellem 2011). Still, the description of the entity and distribution of such processes in the languages of Southern Arabia is not complete and deserves greater scholarly attention.

The first part (§2) of this paper will be devoted to a general description of the terms here discussed, namely glottalization and pharyngealization. In §3 a brief description of the language investigated and its language family will be given. In §4 data and methodology will be presented, followed in §5 by research results.

2. The Emphatics

Both glottalization and pharyngealization are involved in the realization of emphatics in MSAL at various degrees, according to dialectal and socio-linguistic factors (Watson & Bellem 2011). This holds even truer for Baṭḥari. If on the one hand the realization of Baṭḥari emphatics may resemble that of Mehreyyet, the Omani variety of Mehri (as described by Watson & Bellem 2011), on the other hand pharyngealization surely plays a heavier role. It is important to point out that presence of glottalization does not exclude pharyngealization and vice versa, since the two processes are not necessarily in conflict from an articulatory point of view.

In the following paragraphs I will briefly explain what pharyngealization and glottalization are and in which way do they occur in the data analyzed so far, taking into account the observations made during fieldwork.

2.1. Glottalization

Glottalization refers to a secondary articulatory process in which narrowing or closure of the glottis takes place. It varies along two different parameters on a continuum: the degree of closure leads from modal voice (no closure) to stiff voice and creaky voice (partial closure) to ejective articulation (full closure). Also the time of release can vary, ranging from a simultaneous segment to an onset or coda or to glottal reinforcement (Ladefoged & Maddieson 1996: 73-81).

Ejective consonants are present in 18% of the world languages (Maddieson 1984). Ejectives are produced by the action of the closed glottis, while there is an occlusion in the oral cavity. The action of the larynx compresses the air in the vocal tract which, once released, produces a greater amplitude in the stop burst (Ladefoged & Maddieson 1996: 78)

The acoustic characteristics of ejectives can vary cross-linguistically, both in terms of Voice Onset Time (henceforth, VOT) and spreading of the glottalization process to the following vowel (Kingston 1985, Wright & al. 2002). Voice lag is an easily detectable

factor in determining ejective realization: usually, the longer the VOT, the higher the supraglottal pressure and therefore the ejective burst (Fallon 2002). The presence of creaky voice (phonation with irregular pulses) at the onset of the following vowel can also be a marker of glottalization, but this feature seems not to be universal. Finally, the high burst amplitude of the release is another hint of ejective realization (Bellem 2007: 31). As for creaky voice, it shows irregular F0 and aperiodic voice (Keating et al. 2015).

Johnstone (1975: 6) claimed that MSALs emphatics show (pre-)glottalized realization. As showed in § 2,2, this claim, though having had the merit of attracting scholarly interest, should be definitely examined in depth.

2.2. Pharyngealization

Pharyngealization is a kind of secondary articulation involving a constriction of the pharynx usually realized through tongue root retraction (Ladefoged & Maddieson 1996: 365). It is a process well attested throughout the whole Semitic family but not enough investigated outside Arabic dialects.

There can be variation on the locus of constriction, and scholars themselves came to different results on the matter. In fact, according to the variety taken into consideration one should more properly talk about laryngealization or uvularization (Ali & Daniloff 1972, Ghazeli 1977). I will use the term "pharyngealization" as a cover term to indicate a general involvement of the pharynx region leading to a "backed" articulation.

As Yeou (2001: 4) says, "pharyngealization can be studied from an acoustic point of view by examining its effects on the formant frequencies of the adiacent vowels". In fact, acoustic analysis shows that a backed articulation of emphatics in Arabic causes strong lowering of F2 and slight raise of F1 (Giannini & Pettorino 1982), particularly at vowel onset where F2 drop is particularly dramatic and might be seen as the most important factor determining a "backed" perception of a given sound (Obrecht 1968).

It is important to point out that pharyngealization is not the only trait that builds what we usually call an "emphatic" (leaving aside the discussion over the exact locus of constriction in the pharyngeal trait itself). Rather, it contributes with other phenomena, such as lip protrusion (as happens in ṣanʕāni Arabic, Watson & Bellem 2011), lowering of the jaw and sulcalization of the tongue dorsum (Bellem 2007: 44-47).

3. The Modern South Arabian Languages

The Modern South Arabian languages (henceforth, MSAL) are a group of six minority languages (Mehri, Soqoṭri, Jibbali/Śeḥri, Ḥarsūsi, Hobyōt and Baṭḥari) belonging to the Western branch of the Semitic family[1] and spoken by an approximative

[1] I hereby follow Rubin's (2012: 263) tree-model.

total amount of 200.000 people from the Eastern part of Yemen, Soqotra included, to the governorate of Dhofar of Oman. The degree of endangerment of MSAL varies from «definite» to «severe» and «critical», according to UNESCO. The degree of endangerment of Baṭḥari is described as "critical" in the UNESCO Atlas of endangered languages (Moseley 2010) due to the low number of speakers left and the sociolinguistic threat of Arabic towards minority languages in the whole area of Dhofar.

Since the situation of severe endangerment that characterizes Baṭḥari is fundamental for a better understanding of our findings, a general report about the Baṭḥari language and its speakers will be given in the next paragraph. A brief account of the studies related to MSAL phonetics and phonology will then follow.

3.1. The Baṭḥari language and its speakers

Still very little is known about Baṭḥari and its people at present time. Already reported to be "a language itself in sharp decline" (Morris 1983: 143), no further reports followed up until the start of Morris & Watson's project, which attested the presence of few speakers left in the Eastern coastal area of Dhofar. Over a period, Morris conducted a meticulous work of gathering, transcription and translation of ethnolinguistic material which is on its way towards publication. As for my own work, I conducted three weeks of fieldwork in Oman in late 2016, then followed by another month of fieldwork between March and April 2017.

The current area of settlement of the Baṭāḥira is the coastal area of al-Ḥallāniyya bay to the East of Hāsik. Oral tradition reports a wider dominion over Dhofar, from *wadi* Ġadūn (around 50 kms to the West of Ṯumrayt, a small city to the North of Ṣalalah) to *wadi* Ġadūn, homonym of the former *wadi* (near Khalil, in al-Wusṭa governorate). The name of the tribe itself is interpreted as "people of the dust" (*bit bəṭḥā*) as a reference to their numerous ancestry. Their land was later reduced by supposed migrations by the Bayt al-Kaṯīr from the North-West, the Mahra and the Qāra from the West and the Janaybā from the East, which took control of most of the Baṭaḥira land pushing them towards the coast. No historical evidence proves these narratives, but the linguistic heterogeneity of these once nomadic tribes may be illustrative of their former migrations. The members of the tribe, labelled as *ḍāʕīf* ("weak"), were at the lowest levels of the local tribal assets and could not carry weapons, managing to survive with great difficulties in a difficult terrain. At the time of Sultan Qaboos' unification of Oman in the 1970es, the members of the tribe abandoned their nomadic traditional way of life, settling down in different locations along the coast. The major part of the speakers lives now in Shuwaymiya, close to an area the Baṭaḥira themselves recognize as their original homeland (that of Warx and Minji), with few others scattered in nearby coastal villages (Sharbithat, Likbi, Ṣawqara). The radical improvement in their traditional way of life – which, according to elders' reports, was afflicted by continuous lack of food and extreme poverty - had a great impact over their language status. The community, who probably had never been purely monolingual (Morris 1983: 130) went through a fast process of language switch towards Arabic, the prestige language of the area, and/or other MSAL. In fact, there are no monolingual speakers, being all of them fluent in Arabic plus Ḥarsūsi and/or Mehri.

			Lab.	interd.	Alv.	Lat.	Pal.	Vel.	Uvul.	Phar.	Lar.
Obstr	Stops	Voiceless			t			k			ʔ
		Voiceled	b		d			g			
		Emphatic			ṭ			ḳ			
	Fric.	Voiceless	f	ṯ	s	ś	š		x[2]	ḥ	h
		Voiced		ð	z				ġ	ʕ	
		Emphatic		ð̣	ṣ	ṣ́	(ṣ̌)[3]				
Sonorants	Nasal		m		n						
	Liquids				r	l					
	Glide						y	w			

Table 1: phonological system of Baṭḥari

At the moment only 12 elder proficient speakers above 60 y. o. are known to me. It is likely that in one or two decades the language will cease to exist together with the life of its last speakers. It is striking to note how fast the process of language loss took place: mid-aged speakers seem to preserve varying degrees of passive competence, while the younger members of the tribe had virtually no access to the language, nor do they have any interest in learning it.

3.2. State of the field

As mentioned before, the only scientific paper dealing with Baṭḥari is the one by Miranda Morris (1983), which dealt with traditional songs and poems. Much has still to be understood and written about this language. On the contrary, the topic of the realization of emphatics has gathered many scholars' attention since Johnstone (1975) because of the presence of ejective realization, which was said to be akin to that found in Ethio-Semitic languages. Scientific interest focused mainly on Śeḥri and Mehri, which are still widespread across Dhofar and whose speakers are surely easier to track down.

In the first reports about MSAL, emphasis is said to be realized in a way similar to Arabic, but less salient (Viennese expedition 1902-1907, Thomas 1939 and Leslau 1947

[2] In various contexts both /ġ/ and /x/ are velarized rather than uvularized.

[3] Morris (1983: 143) lists /ṣ̌/ as part of the phonological system of Baṭḥari, but this phoneme seems to have undergone merging with /ṣ́/, so that ethimological /ṣ̌/ is synchronically undetectable. For example, from the root *ŠBʕ: Mehreyyet [tʃʾo:bəʔ] (Castagna G., p. c.) vs. Baṭḥari [haɮˁaba:ʕ] "finger" (from my own set of data).

based on Thomas' work). As a breaking point with this tradition, Johnstone (1970, 1975) talks about a post-glottalized realization for the whole set of emphatics. Starting from this observation various authors dealt with this topic, redefining the nature and the distribution of pharyngealization and glottalization as ways of realizing emphasis. The common idea behind most of these works is that both processes take part into the realization of emphatics according to the phonetic environment of the emphatic segment and dialectal variation (Lonnet & Simeone-Senelle 1997; Simeone-Senelle 2011). The Russian expedition (Naumkin & Porkhomovskij 1981) individuates for Soqoṭri an ongoing process of transition from a glottalized to a pharyngealized realization, being only emphatic stops realized as fully glottalized items. Lonnet (2009) points out a connection between presence and degree of pharyngealization and diatopic dialectal variation. The detailed works from Watson & Bellem (2010; 2011) show the presence of many pre-pausal phenomena involving pre-glottalization and the co-occurence of pharyngealization and glottalization in Sanaʿani Arabic, Mahryōt and Mehreyyet, two varieties of Mehri respectively spoken in Yemen and Oman. Finally, Dufour (2016: 22) states that "le caractère éjectif des phonèmes emphatiques ne fait aucun doute, en jibbali comme en mehri"[4].

As seen, there is no consensus over the status and realization of MSAL emphatics. A detailed comparative study aware of each language-related dialectal variation would surely be of great interest (and much needed). As for this paper, I will deal solely with Baṭḥari emphatics, a really interesting case study due to its peculiar situation of endangerment.

4. Data and Methodology

Analyzed data comes from a selection of recordings made between 2014 and 2016 by Miranda Morris for the "Documentation and Ethnolinguistic Analysis of the Modern South Arabian languages" project[5] and from a series of recordings made by me during fieldwork, between October and November 2016. The language of communication used during my interviews was Arabic.

[4] Ridouane et al. (2017) shows the outcomes of a study concerning Mehri emphatic fricatives. Unfortunately, this paper was published during the editing of the present volume and it was not possible to include those data in my paper.

[5] The project is funded by Leverhulme Trust. Over 2,000 sound files are now lodged at ELAR (Endangered Languages Archive) at SOAS (School of Oriental and African Languages), London. Around 300 texts have been transcribed and translated by Morris and a selection of them will soon be published.

	Utterance-initial	Intervocalic	Pre-pausal	Total
/ḳ/	20	20	5	45
/ṭ/	20	20	8	48
/ð̣/	10	10	3	23
/ṣ/	10	15	3	28
/ṣ́/	10	10	5	25
Total	70	75	24	169

Table 2: counts of tokens analysed according to their position

Recordings from 4 speakers - 2 women (S1 and S2) and two men (S3 and S4) - were taken into account. In order to obtain easily comparable data it was chosen to take into account tokens in initial, intervocalic and pre-pausal position. Each token comes from natural speech. As shown in Table 2, a total of 169 tokens were included in the acoustical analysis (table 2). Greater attention was given to segments in which the emphatic was followed by the vowel /a/. This choice was made mainly for two reasons, as to say 1) abundance of occurrences of the aforementioned environment, being /a/ present both in stressed and unstressed syllables and 2) clearer backing processes triggered by the emphatics, which allowed an easier individuation of pharyngealization processes. This does not mean that observation of other environments was neglected; a higher number of occurrences was thought to be more appropriate for a better statistical perspective, lacking a proper list of elicitated words. Positional variants for the other vowels need to be studied with greater attention. Only descriptive statistics will be given though, being the analysed sample too small to be proven significant at an inferential level. As the work proceeds I hope to be able to bring further evidence for the preliminary findings presented in this paper.

Acoustic data was segmented and analysed manually using PRAAT software (version 6.0.23). In order to detect presence of pharyngealization, formants of the following vowel were measured at 1/3 and 1/2 of the vowel. As for glottalization, presence, length and number of pre- and post-emphatic glottal lags were considered. VOT was measured from the start of the oral release burst to the first glottal pulse associated with the vowel. Frication length and intensity at midpoint for fricatives was also investigated.

Values reported for plain counterparts of emphatics will be used to provide a means of comparison and are ancillary to the main discussion presented here. These data will be presented soon, along with my PhD dissertation (Gasparini, forthcoming).

4.1. Problems

Before getting into the main topic, I will highlight some methodological problems I went through and which might somehow hinder the results of the study conducted so far:

i. *Contact with other MSAL and Janaibi Arabic*[6]: as for now it is quite a challenge to define such a degree of contact (though bilingualism was widespread way before Sultan Qaboos) and the lack of descriptions concerning the local Arabic varieties does not help.

ii. *Speakers' proficiency*: as a consequence of the previous point, it is not surprising that the data might have been heavily influenced by the linguistic background of each speaker. A comparison with data from Janaibi Arabic would especially be highly valuable and much needed. Only a careful look at Morris' still unpublished set of data from the 1980es could have told us more about the time when Baṭḥari was still healthy. Unfortunately, the audio quality of those recordings is not good enough to allow a precise phonetic analysis.

iii. *Speakers' selection*: it was chosen to conduct a phonetic analysis only over the speakers which had the most intelligible pronuntiation. Due to shaking voice and stuttering part of the speakers were judged not suitable for this study.

iv. *Tokens from spontaneous speech*: there was no chance to conduct elicitation over a list of words with pre-programmed, controlled environments. Fast-speech processes thus might play a certain degree of influence on the chosen samples. The number of tokens for each phoneme also varies according to the incidence of each phoneme in the recordings, which is obviously incidental.

v. *Statistical evidence*: for the reasons outlined above, it is still hard to provide a reliable inferential statistical evidence.

5. Results

Results will be presented in three different sections, in order to make informations more easily available. First, I will deal separately with emphatic and fricative stops in initial and intervocalic position. Utterance-final positions was proven to trigger a whole set of phenomena also involving other phonological classes. It was chosen not to deal in detail with tokens in this position; rather, future work will specifically address this specific topic.

Data analysis showed idiolectal variation to be more important than any other variable, as expected.

[6] By this label I refer to the local undescribed variety of Arabic, named after the Janaiba, a neighbouring Bedouin tribe, as firstly reported by Morris (1983).

5.1. Emphatic stops

	Utterance-initial		Intervocalic		Total
	/ḳ/	/ṭ/	/ḳ/	/ṭ/	
S1	6	5	6	5	22
S2	5	4	9	3	21
S3	5	5	5	3	18
S4	4	6	5	4	19
	20	20	25	15	80

Table 3: Counts of tokens analysed divided for each speaker

Analysis of emphatic stops proves that /ḳ/ strongly shows signs of glottalization both in initial and intervocalic position. A first clue comes from waveform analysis, being the ejective burst clearly visible in all the data from S1, S2 and S3.

Emphatic stops clearly differ from their plain counterparts in terms of VOT in the case of utterance-initial /ḳ/ (SD 10,83 ms) and intervocalic /ṭ/ (SD 0,52 ms) (see table 4). F0 measurements show a slight increase of F0 at vowel onset in the case of utterance-initial stops (F0 at onset – F0 at midpoint = 12,53 Hz, SD 7,9), while for intervocalic stops raise of F0 is less salient (2,79 Hz, SD 1,21).

S4 diverges from the other speakers by showing significantly lower VOT for utterance-initial /ḳ/ (25,9 ms, SD 0,4) (see figure 1 and 2). He also shows creaky voice at the onset of the following vowel for 2,3 ms (SD 0,2) and a weaker release burst, akin to that of plain /k/.

Vowel formants analysis of the following vowel showed signs of pharyngealization in the case of /ṭ/, while for /ḳ/ formants did not prove to be relatable to clear pharyngealization processes (Table 5). For utterance-initial /ṭ/a lowering of F1 of 16,48 Hz (SD 7,92) from onset to midpoint was detected together with raising of F2 of 128,42 (SD 41,06). For intervocalic /ṭ/ F1 udergoes a lowering of 7,85 Hz (SD 2,75), while F2 raises of 93,28 (SD 23,91).

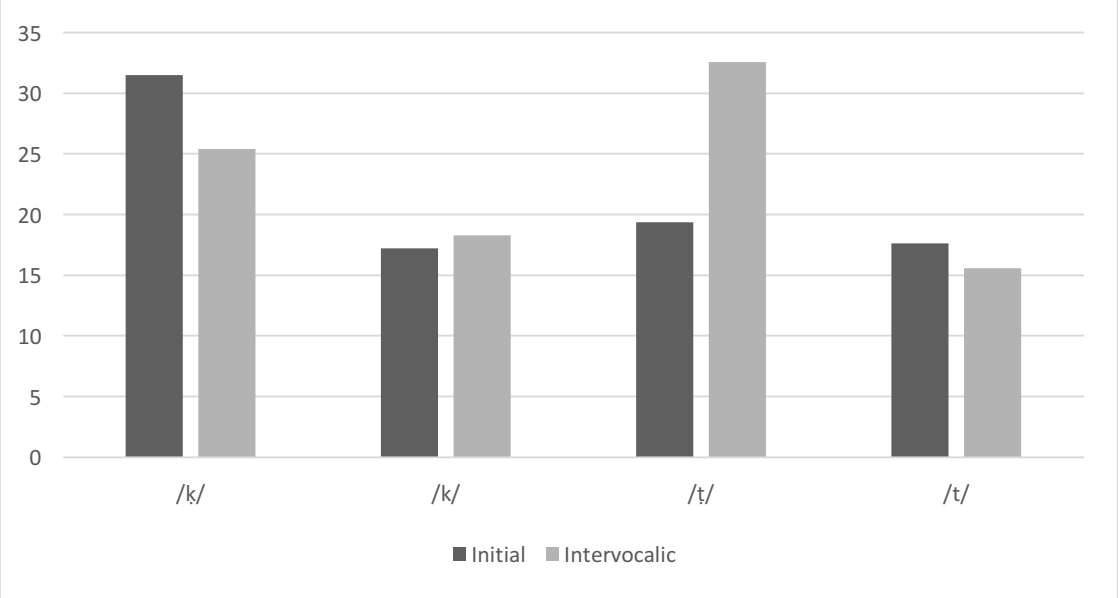


Table 4: VOT values (ms) for emphatic stops

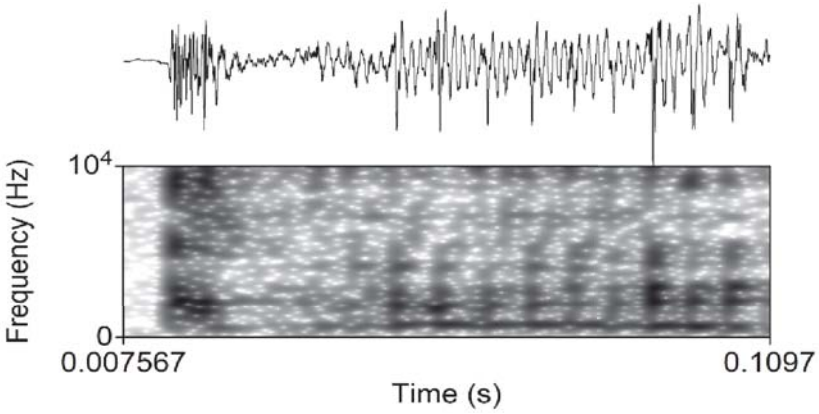


Figure 1: S1 pronouncing /ḳa/: strong burst and long VOT

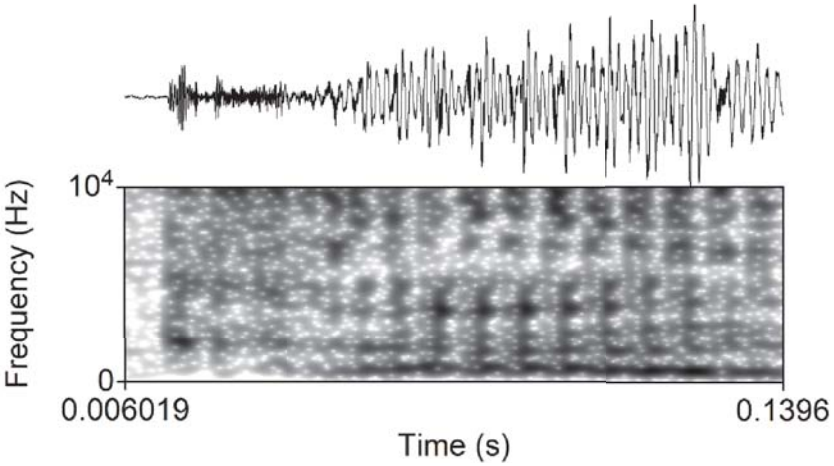


Figure 2: waveform and spectrogram of S4 pronoucing /ḳa/: weak burst, low VOT and aperiodic vowel onset

		F2 - F1 at onset (Hz)	F2 – F1 at midpoint (Hz)
/ḳ/	Utterance-initial	862,72 SD 57,172	879,52 SD 30,176
	Intervocalic	796,954 SD 72,98	803,648 SD 66,72
/ṭ/	Utterance-initial	699,58 SD 65,55	754,65 SD 74,45
	Intervocalic	621,98 SD 70,72	709,6 SD 50,05

Table 5: values for F2 - F1

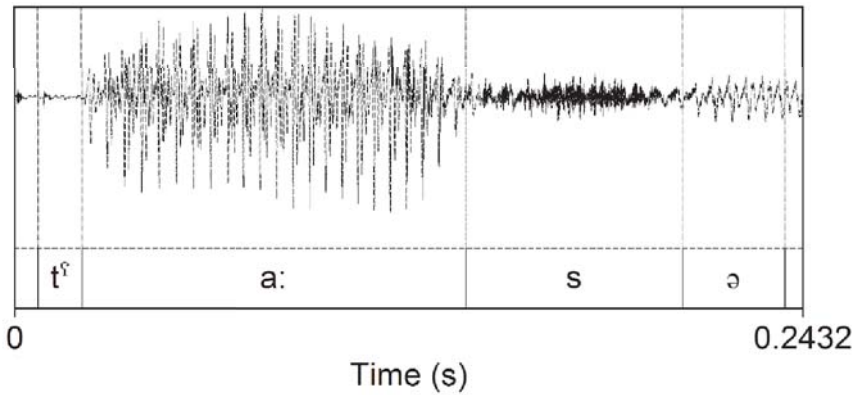


Figure 3: waveform of S03 pronouncing /ṭasəh/ "bowl"

5.2. Emphatic fricatives

	Utterance-initial			Intervocalic			Total
	/ð̣/	/ṣ/	/ṣ́/	/ð̣/	/ṣ/	/ṣ́/	
S1	3	2	3	1	3	3	15
S2	2	2	2	4	4	1	15
S3	2	3	4	3	4	2	18
S4	3	3	1	2	4	4	17
	10	10	10	10	15	10	65

Table 6: counts of tokens analysed divided for each speaker

The study of emphatic fricatives is more problematic than that of emphatic stops. The main problem consists in determining whether and how ejectivity is involved in the realization of emphatics, since there is a natural articulatory incompatibility between the continuing air flow typical of frication and increasing of air pressure in the oral cavity leading to ejective realization (Maddieson 1998). In fact, only 3,7 % of the world's languages show at least one ejective fricative segment (Maddieson 2013).

Since there was no chance to organize any recording session in a proper lab for phonetic analysis (nor I would have ever dared to pursue such a challenging task with my old interviewees living hundreds of kms away from Ṣalalah), only direct observation was possible of how Baṭḥari emphatics are articulated as far as lips and jaw position are concerned. This observation is not rigorous enough as an articulatory analysis, so it cannot be held as a satisfying description and cannot be considered reliable in this context. Futhermore, it would be interesting to compare these articulations with those of Janaibi Arabic, but lack of data on this behalf prevents from pursuing such a task at the present time.

Pre-frication and post-frication silent lags were proven to be systematically absent in the data (figure 4). No speaker ever produced any sort of silent lag while articulating emphatic fricatives, nor affrication processes seem to happen, unlike in Mehreyyet (Ridouane et al. 2015).

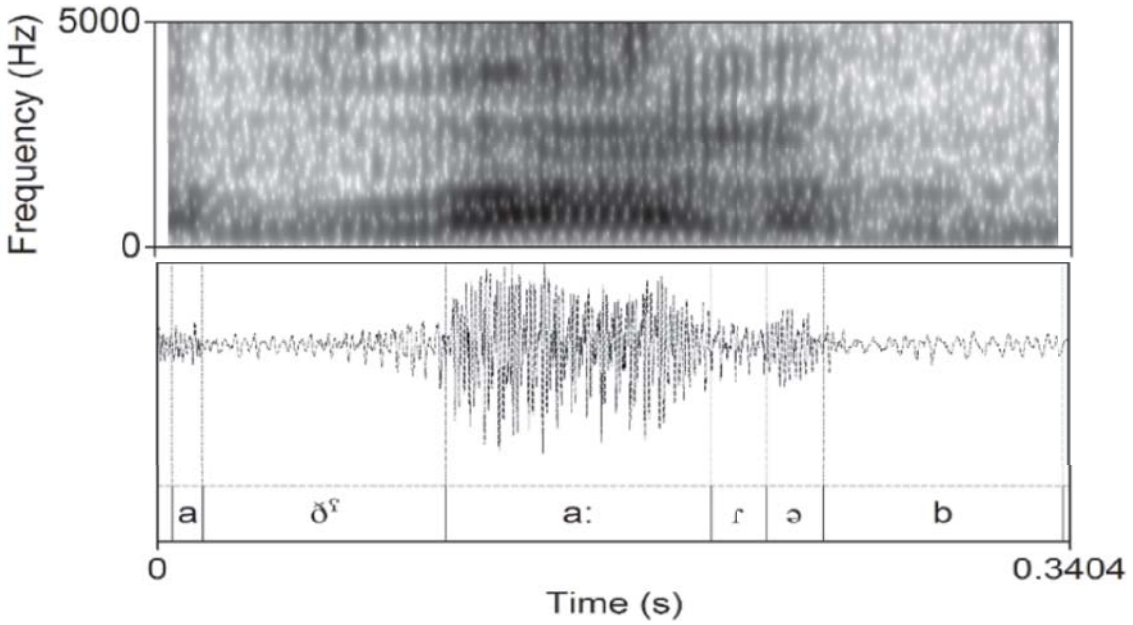


Figure 4: S03 pronouncing /aḍarb/ "stick"

Emphatic /ḍ/ is uniformly articulated as a pharyngealized voiced dental fricative [ðˤ]. Measurement of formants of the following vowel proves this statement (Table 7), with a raising of F1 at vowel onset of 23,48 Hz (SD 8,02) and a strong lowering of F2 of 168,83 Hz (SD 36,27). Frication length and intensity are higher than their plain

counterpart for /ð̣/ and /ṣ/, while they are lower for /ṣ́/ (table 8 and 9). Sibilants, on the other hand, are much more unstable, as Standard Variation values might suggest. Still, pharyngealization seems to be more salient than glottalization in this context (Table 7). Variation happens at idiolectal level and more data should be analysed before making safe assumptions on this matter.

From spectrogram analysis voicing processes for both sibilants were detected through the presence of a voice bar during frication time (7/25 occurrences for /ṣ/ and 12/20 occurrences for /ṣ́/). It is unclear whether voicing is conditioned only by idiolectal variation or there are phonotactical constraints for this to happen, given the low number of occurrences that I could examine. There appears to be a correlation between voicedness and the position of the token inside the word – namely, voiced realization is favoured in intervocalic position, while in word-initial position voiceless realization shows up in few cases (only 1/10 for /ṣ/ and 2/10 for /ṣ́/). Moreover, I do not have physiological data on vocal fold vibrations so I can only make an educated guess on this matter.

		F2 - F1 at onset (Hz)	F2 – F1 at midpoint (Hz)	Frication length (ms)	Frication intensity (dB)
/ð̣/	Utterance-initial	548,98 SD 32,45	775,735 SD 45,34	94,1 SD 11,3	60,59 SD 1,56
	Intervocalic	572,319 SD 23,94	673,005 SD 21,72	48,6 SD 15,7	63,39 SD 1,32
/ṣ/	Utterance-initial	589,121 SD 67,79	609,02 SD 81,66	91,5 SD 15,39	57,10 SD 0,98
	Intervocalic	577,48 SD 78,32	579,52 SD 56,37	75,13 SD 11,02	56,88 SD 2,80
/ṣ́/	Utterance-initial	583,159 SD 85,61	627,056 SD 74,45	60,49 SD 8,87	57,08 SD 1,62
	Intervocalic	601,531 SD 50,83	627,75 SD 73,11	78,45 SD 9,64	55,25 SD 3,205

Table 7: values measured for emphatic fricatives

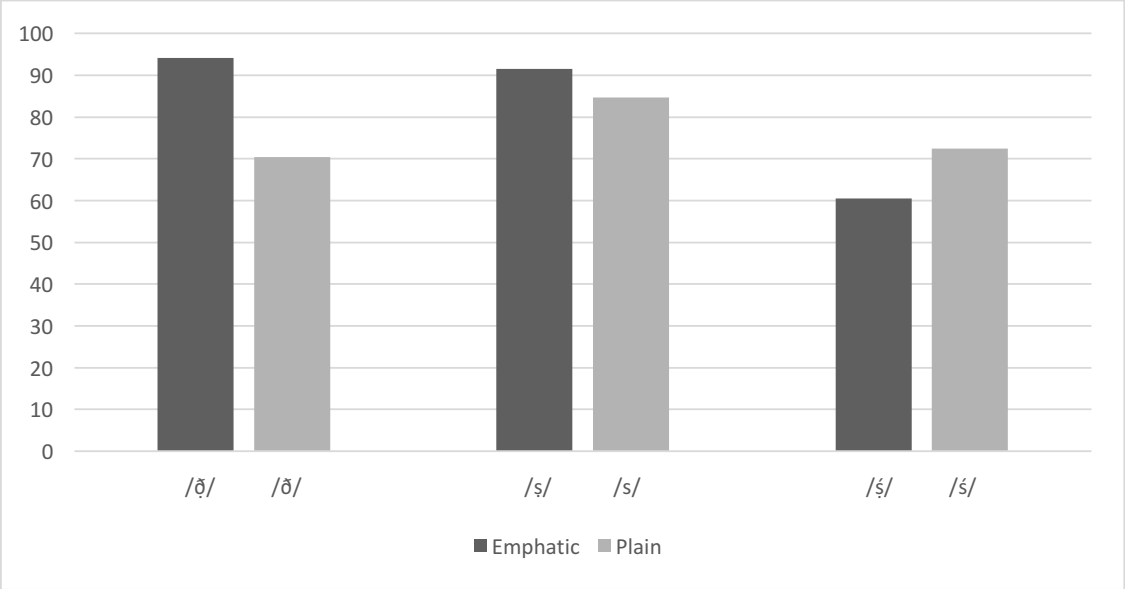


Table 8: Frication time (ms) of emphatic fricatives compared to that of their plain counterparts in utterance-initial position

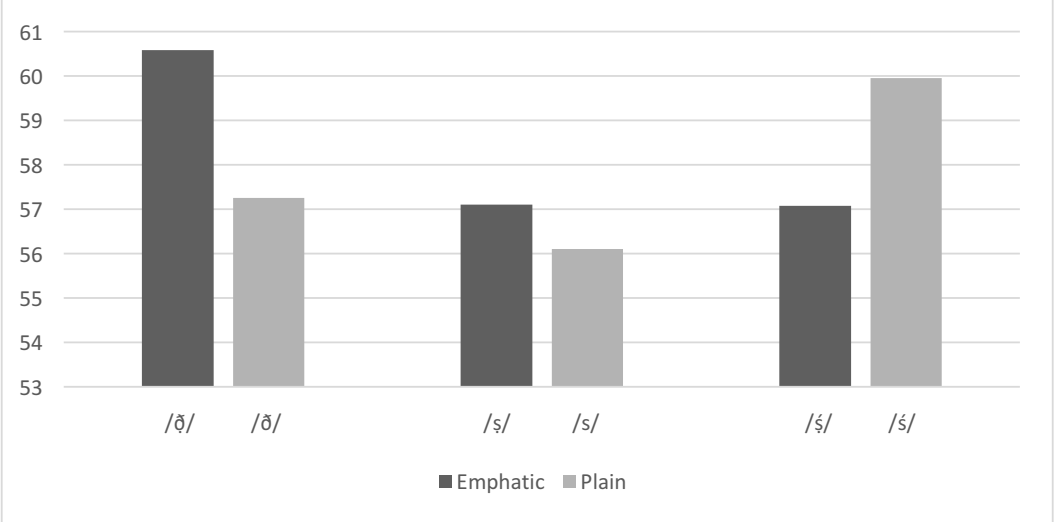


Table 9: values of frication noise intensity (dB) of emphatic fricatives compared to that of their plain counterparts in utterance-initial position

6. Conclusions

The role played by glottalization in Baṭḥari seems to be less salient than it is in Mehri. The only segment which regularly shows a fully ejective realization is /ḳ/ in all kinds of environments observed so far. Speakers' articulations vary between those of "stiff" and "slack" ejectives, as defined by Lindau (1984) and Kingston (1985): stiff

ejectives are characterized by a silent period between consonant release and vowel onset resulting in long VOT, fast rise time, relatively long total duration (closure duration + VOT), small closure/VOT ratio, relatively intense burst, high F0 at vowel onset and modal voice quality at vowel onset. Stiff ejectives are also found in Tigrinya (Kingston 1985). Slack ejectives have slightly different characteristics: relatively short VOT, creaky voice quality at vowel onset, slow rise times, shorter total duration and larger closure duration/VOT ratio, as happens in Hausa (Lindau 1984). More statistical evidence might prove the idiolectal nature of variation between these two kinds of articulation.

Also /ṭ/ occasionaly shows ejective realization (concurrent with pharyngealization), but with a higher degree of variation than for /ḳ/. In intervocalic position only pharyngealization takes place. As for emphatic fricatives, their status is surely different from that of Mehri fricatives. Sibilants are particularly problematic, in that high idiolectal variation requires the analysis of a wider set of tokens in order to have a better understanding of their phonemic status.

This paper tried to show a first insight over the relationship and interactions between glottalization and pharyngealization in the articulation of Baṭḥari emphatics. As said, both pharyngealization and glottalization are at play. Pharyngealization seems to be more salient than glottalization, which is still manifested in specific contexts. It is clear, though, that pharyngealization represents a preferred way for the expression of emphasis for the set of emphatic fricatives, while emphatic stops partially preserve clear ejective realizations. It is important to note that there were no traces of ejective affricate releases for the emphatic fricatives as happens in Mehreyyet (Ridouane et al. 2015), where glottalization plays a heavier role.

In conclusion, I argue that the label "emphatic" should be retained as a general cover term when referring to this contrastive phonological class rather than making use of either "ejective" or "pharyngealized". The ambiguous status of these segments and especially of emphatic fricatives does not allow to define the whole class using a narrower term.

The topic remains of high interest and studies of this kind should be conducted over the remaining MSAL (particularly Ḥarsūsi and Hobyōt, for which no such analysis has been made so far). A thorough comparison of the results may lead to a final typology of MSAL emphatics which would shed new light over the studies on emphasis in Semitic.

BIBLIOGRAPHICAL REFERENCES

Ali, H.A. & Daniloff, R. G. (1972), *A Contrastive Cinefluorographic Investigation of the Articulation of Emphatic–Non Emphatic Cognate Consonants*, in "Studia Linguistica" 26: 81-105

Bellem, A. (2007), *Towards a comparative typology of emphatics: across semitic and into Arabic dialect phonologymphatics*, PhD Thesis, London

Dufour, J. (2016), *Recherches sur le verbe sudarabique moderne*, mémoire d'habilitation à diriger les recherches, sous la direction de Gilles Authier, Paris, Ecole pratique des hautes études

Fallon, Paul D. (2002), *The Synchronic and Diachronic Phonology of Ejectives*, New York, Routledge

Gasparini, F. (Forthcoming), *Baṭḥari language of Oman,* PhD thesis, Università "L'Orientale" di Napoli

Ghazeli, S. (1977), *Back Consonants and Backing Coarticulation in Arabic*, University of Austin, Texas PhD Thesis
Giannini, A. & Pettorino, M. (1982), *On the Distinctive Function of the F1 Locus: the Emphatic Consonants in Arabic*, in "Fortschritte der Akustik", FASE/DAGA: 1001-1004, Göttingen
Johnstone, T.M. (1970), A definite article in the Modern South Arabian Languages, in "BSOAS" 33: 295-307
Johnstone, T. M. (1975), *The Modern South Arabian languages*, in "Afro-Asiatic Linguistics" 1(5)
Keating, P., M. Garellek, & J. Kreiman (2015), *Acoustic properties of different kinds of creaky voice*, in The Scottish Consortium for ICPhS 2015 (Ed.*), Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, UK: University of Glasgow
Kingston, J. (1985), *The Phonetics and Phonology of the Timing of Oral and Glottal Events*, Ph.D. dissertation, Department of Linguistics, University of California, Berkeley
Ladefoged, P. & Maddieson, I. (1996*), The Sounds of the World's Languages*, Oxford, Blackwell
Leslau, W. (1947), *Four Modern South Arabian Languages*, in "Word" 3: 180-203
Lonnet, A. & Simeone-Senelle, M-C (1997*), La phonologie des langues sudarabiques modernes*, In Kaye, A. S. (ed.), *Phonologies of Asia and Africa* 1, Winona Lake, Eisenbrauns: 337-371
Lonnet, A. (2009), *South Arabian, Modern*, in K. Versteegh et al. (eds), *Encyclopedia of Arabic Language and Linguistics* 4: 297-300
Maddieson, I. (1984*), Patterns of Sounds*, Cambridge, Cambridge University Press
Maddieson, I. (1998), *Why Make Life Hard? Resolutions to Problems of Rare and Difficult Sound Types*, in "Proceedings of the Annual Meeting of the Berkeley Linguistic Society" 24: 367-380
Maddieson, I. (2013), *Glottolized Consonants*, in Dryer, M. S. & Haspelmath, M. (eds.), *The World Atlas of Language Structures Online*, Leipzig, Max Planck Institute for Evolutionary Anthropology (Available online at http://wals.info/chapter/7, accessed on 19/05/2017)
Morris, M. (1983), *Some Preliminary Remarks on a Collection of Poems and Songs of the Batahirah*, in "Journal of Oman Studies", Vol. 6 (1): 129-144
Morris, M. (2007*), The pre-literate, non-Arabic languages of Oman and Yemen*, in "Journal of the British–Yemeni Society" 15 (12)
Moseley, C. (ed.) (2010), *Atlas of the World's Languages in Danger*, 3rd ed. Paris, UNESCO Publishing. Online version: http://www.unesco.org/culture/en/endangeredlanguages/atlas, visited on 12th of January 2017
Müller, D.H. (1902), *Die Mehri - und Soqoṭri - Sprache. Volume I: Texte*. Südarabische Expedition, Band IV. Kaiserliche Akademie der Wissenschaften in Wien, Wien: Alfred Hölder
Müller, D.H (1905), *Die Mehri - und Soqoṭri - Sprache. Volume II: Soqoṭri Texte*. Südarabische Expedition, Band VI. Kaiserliche Akademie der Wissenschaften in Wien, Wien: Alfred Hölder
Müller, D.H (1907), *Die Mehri - und Soqoṭri - Sprache. Volume II: Šhauri Texte*. Südarabische Expedition, Band VII. Kaiserliche Akademie der Wissenschaften in Wien, Wien: Alfred Hölder
Nöldeke, T. (1899), *Die Semitichen Sprachen. Eine Skizze,* 2nd ed., Leipzig
Naumkin, V. V. & Porkhomovskij, V. Y. (1981*), Ocherki po Etnolingvistike Sokotry*, Moscow
Obrecht, D.H. (1968), *Effects of the Second Formant on the Perception of VelarizationConsonants in Arabic*, The Hague, Mouton
Ridouane, R., Gendrot, C., & Khatiwada, R. (2015), *Mehri ejective fricatives: an acoustic study*, in "Proceedings of the 18th International Congress of Phonetic Sciences", Glasgow
Rubin, A. & Huehnergard, J. (2012), *Phyla and Waves: Models of Classification of the Semitic Languages*, in Weninger, S. (ed.), *The Semitic Languages: An International Handbook*, Berlin, Mouton De Gruyter: 259-278
Shosted, R.K., & Rose, S. (2011), *Affricating ejective fricatives: The case of Tigrinya*, in "Journal of the International Phonetic Association" 41(1): 41-65
Simeone-Senelle, M-C. (2011), *Modern South Arabian*, in S. Weninger, G. Khan, M. P. Streck, J.E.

Watson (eds), *Semitic languages. Handbücher zur Sprach und Kommunikations Wissenschaft [HSK 36]. An International Handbook, Berlin*, New York, Mouton De Gruyter: 1073-1113
Thomas, B. (1937), *Four Strange Tongues from South Arabia: The Hadara Group*, in "Proceedings of the British Academy", Vol. XXIII. 1, London, Humphrey Milford: 231-331
Watson, Janet C. E. & Bellem, Alex (2010), *A Detective Story: Emphatics in Mehri*, in "Proceedings of the Seminar for Arabian Studies", Starkey, Janet C. M. Archaeopress 40: 345-356
Watson, J. C. E. & Bellem, A. (2011), *Glottalisation and Neutralisation in Yemeni Arabic and Mehri: an acoustic study*, in Heselwood, B. & Hassan, Z. (eds), *Arabic Instrumental Phonetics*, Amsterdam, Benjamins: 436-4
Yeou, M. (2001), *Pharyngealization in Arabic: Modelling, acoustic analysis, airflow and perception*, in "Revue de La Faculté des Lettres El Jadida" 6, 51-70
Wright, R., Hargus, S. & Davis, K. (2002), *On the categorization of ejectives: data from Witsuwit'en*, in *Journal of the International Phonetic Association* 32: 43-77

FABIO GASPARINI • Is a PhD student from the University "L'Orientale" of Naples. His current project concerns the study of Baṭḥari language, the least known and documented among the Modern South Arabian Languages. His main research interests are Modern South Arabian Languages and Semitic languages in general, Phonetics and Phonology, Endangered Languages and Documentation. His previous works related to Neo-Aramaic studies.

MAIL • fabio.gsparini@gmail.com

LANGUAGE AND NATURE IN DHOFAR*

Janet C. E. Watson *& Abdullah Musallam* Al-Mahri

ABSTRACT • This paper addresses the relationship between language and nature in Dhofar. We begin by considering how erosion of the environment and the relationship people have with the environment can precipitate language loss. We consider how the relationship between language and nature is expressed, referring to transcribed oral texts to illustrate points and focussing on spatial and temporal reference terms, and discuss how decoupling of the human–nature relationship contributes to language attrition. We then examine figurative language and nature in the region; in section 3, we discuss grammaticalisation of Mehri *śaff*, Śḥerēt *śɛf* to the mirative particle *śaf* 'it transpired that' in Mehri, *śɛf* in Śḥerēt.

KEYWORDS • Modern South Arabian, Mehri, Śḥerēt, nature, figurative language

1. Introduction

Regions of the world with greatest biodiversity are shown to exhibit greatest linguistic diversity (http://www.pnas.org/content/109/21/8032.long), strongly suggesting that the relationship between Language and Nature is both symbiotic and spatially and temporally determined. Indigenous languages reflect the close relationship between people and their natural environment, embodying the complex relationship humans enjoy with landscape and seasons. These connections can be broken when indigenous languages are severed from the ecosystems in which they arose, a factor that can arise through replacement of indigenous languages by dominant lingua franca, through degradation of the ecosystem, through depopulation, or through forced or

* Many thanks to the Leverhulme Trust for funding this research through a project grant 2013-2016 RPG-2012-599. Thanks to Nasir al-Awdhi al-Mahri, Ahmad Hardan, Yahya Musallam al-Mahri, Naima al-Mahri, Nwēr Muhammad Bakrayt, Saeed al-Mahri, Musallam Hazmay al-Mahri, Saeed al-Awaid and Munira al-Azraqi for supplying and checking parts of the data in person, by phone and by WhatsApp and SMS; thanks to the speakers Azad Musallam al-Kathiri, Musallam Hazmay al-Mahri, Ahmad Xamis al-Mahri and Ubxayt Saeed Shel al-Mahri for providing the recordings; and thanks to Miranda Morris, Jon Lovett, James Dickins and Sarah Dickins for reading and commenting on the text.

voluntary removal of the indigenous language community from the local ecosystem. Language and nature exhibit a particularly tight symbiotic relationship in regions of the world such as Dhofar in which people have traditionally enjoyed a close relationship with the natural world. Work in recent years has examined the relationship between language and nature in different regions of the world, showing that 'linguistic structure is not only shaped by how speakers interact with each other and with the world they live in, but also by external forces that are outside the control of individual speakers or speech communities' (De Busser and LaPolla 2015). This paper is, however, the first to focus on the language–nature relationship within Dhofar. Our data is all first-hand fieldwork data, collected mainly in Dhofar and partially in the UK through digital audio and audio-visual recordings. Data was then checked with native speaker consultants in addition to the second author, in person, and through email, WhatsApp, SMS and phone.

In section 1 of this paper, we consider language erosion factors, and how erosion of the natural environment and the relationship people have with the environment can precipitate language loss. In section 2, we examine the language–nature relationship focussing on spatial reference terms, temporal reference terms and the measurement of time, and then considering figurative language. Figurative language is closely engrained into the natural environment (Macfarlane 2015): cross-linguistically, expressions of beauty relate to what communities of speakers find beautiful in nature: in eastern Saudi Mehri a beautiful girl may be referred to as *ṭōhī* 'large cumulus cloud' (Munira al-Azraqi, p.c.); in San'ani Arabic, a young girl is described as *xaḏra* 'green' due to her freshness, where 'green' in English would be interpreted as naive; in English, beauty is associated with spring in relation to the cold of winter, or to summer, as we see in Shakespeare's 'Shall I compare thee to a summer's day?' (Sonnet 18). We show that much figurative language in the languages of Dhofar cannot be understood fully once the human–nature relationship decays. Figurative language may in turn induce grammaticalisation, and in section 3 we discuss the grammaticalisation of Mehri *śaff*, Śḥerēt *śɛf* 'track' to the mirative particle *śaf* 'it transpired that' in Mehri, *śɛf* in Śḥerēt, and present two illustrative texts on tracking in Mehri in which concrete and figurative uses of the term *śaf(f)* occur.

1.1. Erosion of language and the environment

Language erosion in many parts of the world has been precipitated by social change, the collapse of traditional cultural activities, and a break with the relationship people have with the natural environment. We see this in the British Isles, where modernisation, urbanisation and communication have resulted in local terms often with concise nuances losing currency and being replaced by general and superregional cover terms: in Shetland, *feevl* 'snow falling in large flakes' contrasts with *flukra* 'snow falling in large, scale-lie flakes', and regional English terms for icicle include *aquabob*, *clinkerbell*, *dagger*, *daglet*, *ickle*, *shuckle*, *shackle* (Macfarlane 2015: 87-9). Classical Arabic exhibits a plethora of terms for 'to go' at various times of day and for various purposes;[1] in Modern Standard

[1] As we shall see below for Modern South Arabian.

Arabic, 'to go' is predominantly expressed with the cover term *ḏahaba* with an adverbial phrase to express time of day or manner of moving. In Dhofar, we are at a time when members of the generation who have experienced the pre-motorised past are still alive, and thus we are able to observe the effect erosion of the environment and the human–environment relationship is having on language at the present time.

1.2. Social change and urbanisation

There is an absence of accurate statistics for Dhofar (http://www.fao.org/ag/agp/agpc/doc/counprof/oman/oman.htm). However, we know from our ethnographic work conducted with people throughout the region that in the pre-Sultan Qaboos era (pre-1970) people had no motorised vehicles and transport was by foot, by riding animal or by boat, water was collected by individuals from natural sources, and people outside the small towns on the coast lived in caves or in brushwood or stone huts they constructed themselves. The ethnographic texts recorded through the Leverhulme-funded Documentation and Ethnolinguistic Analysis of Modern South Arabian project (http://www.leeds.ac.uk/arts/homepage/462/modern_south_arabian_languages) include first-hand accounts from people in their 50s and above of constructing huts and shelters for animals and people, fetching water from different sources, walking or riding great distances, producing tools and handicrafts from leaves, leather, bone, clay, stone and wood, and practising seasonal transhumance. We also know that multilingualism was widespread at this time, and that people of the interior enjoyed a symbiotic relationship with people of the small coastal towns. The common currency was Maria Theresa Dollars, but for many people a barter system operated, and townspeople would barter dried fish and imported goods for farm produce from the mountains. Frankincense would be bartered for food and clothing.

Today the region enjoys all the trappings of the modern age, with a hitherto overwhelmingly rural population becoming rapidly urban, and a significant nomadic population becoming almost wholly sedentarised. Many small towns around the gravel desert came into existence in the 1980s – Rabkut, the town from which two of the texts presented in this paper come, only came into existence in 1984. Since then there has been gradual sedentarisation of people, with, since 1976, the construction of *šaʕbīyāt* 'government-funded housing'. At various times, waves of people have abandoned life in the mountains and parts of the desert to settle in the regional centre, Salalah.

Through urbanisation and sedentarisation, life which used to be lived predominantly out of doors has become, for many, almost entirely indoor. As a result, younger generations no longer require, have, or understand the extensive knowledge and practical skills of their elders; much earlier expertise has been lost or is disregarded when imported plastic, metal and nylon alternatives replace locally produced items. Where items do continue to be produced locally, they are often made by the large migrant labour force from south-east Asia using imported artificial materials rather than the local natural materials used in the past. This we have observed particularly in the case of fishing equipment. When a society no longer discusses and passes on traditional skills, the older generation may forget, and the younger generation never need to learn, the relevant lexical items (Thomason 2015).

In an increasingly sedentary, urban way of life, traditional methods of natural resource and water management are no longer passed to the next generation, and significant degradation of the environment has occurred, with overgrazing and mismanagement of increasingly scarce water supplies, and severe overfishing on the coast. One result of environmental degradation is that plants and fauna that once played a significant role in everyday human life are now extinct or rare, and where they do remain extant they no longer play the essential role in human life they once played. This loss of traditional knowledge, skills and habitat is one of the key factors in language endangerment in all MSAL, particularly, but not exclusively, in the lexis: the loss of tools produced from the natural environment leads to loss of lexemes and of linguistic expressions relating to these objects.

1.3. A dominant national language

One of the principal ways in which the language–nature relationship can be severed is through (partial) replacement of indigenous languages through dominant lingua franca. This is an effect we are witnessing in the Modern South Arabian region. The Modern South Arabian languages lack any traditional script, which means that any script-based education or communication is conducted through Arabic. The introduction of schools to the region in the 1970s and the subsequent spread of Arabic have led to particularly frequent expressions being replaced by their Arabic equivalents. Thus, we observe Arabic terms such as *maṯalan* 'for example' and *nafs aš-šī* 'the same thing' in place of Śḥerēt *gens,* Mehri *gans*; *al-gīl al-gidīd* 'the new generation' in place of Śḥerēt *ṣaffɛt īdənut* and Mehri *ṣaffēt īdanōt*; *lākin* 'but' in place of Śḥerēt *duʰn* and *min duʰn* (cf. Rubin 2014: 257) and Mehri *lahinnah*; *yaʕnī* 'that is to say' and *ʕabārah* in place of Śḥerēt *yaxīn* and Mehri *(y)axah*; and *tamām* 'fine' in place of Śḥerēt *ḥayšōf* and Mehri *hīs taww ~ histaww*.

Numbers and colours are often among the first lexical items to be lost in endangered languages, and this we see in Dhofar: older generation speakers are still able to count in the local languages, but children and many urban dwellers in their twenties count beyond 10 using the Arabic numerals. Speakers of all ages of Mahriyōt, the eastern Yemeni variety of Mehri, Mehreyyet elsewhere in Yemen, and the Mehri spoken in eastern Saudi Arabia (Munira al-Azraqi, p.c.) rarely count beyond 10 in the local language. Throughout the region, telephone phone numbers are given exclusively in Arabic, possibly due to the lack of a single-word MSAL equivalent to Arabic *ṣufr* 'nothing'. Regarding colour terms, we observed that even children living in goat- and camel-herding villages would refer to animal colours using the Arabic terms in place of terms in the local languages: *aḥmar* 'bay' in place of Mehri *ōfar*, Śḥerēt *ʕofer*, *aswad* 'black' in place of Mehri *ḥōwar*, Śḥerēt *ḥōr*, and *abyaḍ* 'white' in place of Mehri *ūbōn,* Śḥerēt *lūn*. We believe this is at least in part due to the fact that children and young people rarely work with livestock today.

2. The language–environment relationship

The closer the relationship between people and the natural environment, the more linguistic expressions in the local languages refer to the environment and/or reflect the

human–environment relationship.[2] Within the MSAL domain, spatial reference points are based on topographic terms and differ according to language variety and place of the speaker. The language of quantification is frequently nature-based: daylight temporal reference points depend on the relative height and position of the sun, the passing of time is measured by reference to known traditional activities, verbs of movement differ according to the time of departure, and expressions of amounts and animal group sizes are dependent on the object of description. Here we consider spatial reference terms and measurement of time.

2.1 Spatial reference terms

The MSAL employ absolute spatial reference terms which correlate with topographic variation. The younger generation speakers are said by our consultants to generally recognise and employ the traditional latitudinal terms, but to replace the longitudinal terms by Arabic *šimāl* 'north' and *janūb* 'south'. Several of our older Mahriyōt consultants claim that they have known *šimāl* and *janūb* since they were young and employ these to the exclusion of local terms.

For latitudinal reference terms, *ṣalōt* or *nṣurɛ̄t* refers to the direction of the rising sun, and *ḳabəlɛ̣̄t* the direction of the setting sun in Śḥerɛ̄t; these correspond to *nṣarāt* and *ḳabalēt* in Mahriyōt (Nwēr Muhammad Bakrayt, p.c.), and *maśḳayṣ* and *ḳabalēt* or *mārēb ~ mārīb* in Omani Mehri. Three of these terms refer to fixed geographical locations: *ṣalōt* 'the eastern region of Dhofar', *ḳabalēt* (*ḳabəlɛ̣̄t*) 'the Qibla', and *mārēb ~ mārēb* 'Ma'rib' or 'the Ma'rib dam'.[3] For longitudinal reference, terms differ according to language and place of the speaker. In the mountains, the Jarbeeb (the coastal plain) and the town of Salalah, Mehri *rawram* 'sea', Śḥerɛ̄t *ramnam*, may describe a general southerly direction, and Mehri *nagd*, Śḥerɛ̄t *fagir* 'desert'[4] general north (Saeed al-Mahri, p.c.). Around Taqah and south of the mountains, Śḥerɛ̄t *ramnam* 'sea' indicates general south, and *śḥɛr* 'mountains' general north. In the mountains, *fagir* 'desert' indicates general north (Saeed al-Awaid, p.c.). More commonly, longitudinal directions are described in terms of the direction of the flood waters: *u-mṣaʔ* (< *la-mṣaʔ*) in Mehreyyet denotes the direction in which the flood waters run, and *ḥaḳ*, *l-ḥaḳ* the direction from which the flood waters

[2] For example, the parrot fish, known in eastern Yemeni Mehri, Mahriyōt, as *ġaṣabīt hibʃayt* 'disarmer of seven' is said to have been thus named following an incident where seven men made a vain attempt to catch it. A text describing this story, 20130414_MehriMo_M005_howGha SbiitHib'aytGotItsName, can be accessed from ELAR (http://elar.soas.ac.uk/deposit/0307). According to Miranda Morris (p.c.) it 'needed seven men to take hold of it' because it is notoriously slippy.

[3] In San'ani Arabic, *giblī* refers to north and *ʃadanī* 'towards Aden' to south (Watson 1993). In Faifi, spoken on the Saudi side of the Saudi–Yemeni border, *šāmin* 'Levant' refers to north of the Faifa mountains and *yamanin* 'Yemen' to south of the Faifa mountains (Alfaifi 2016).

[4] *fajir* in Eastern Śḥerɛ̄t, as in the text in 2.1.1.

emanate.[5] To the desert-side of the mountains, *u-mṣaʔ* roughly indicates 'north', since the flood waters flow in a northerly direction; seaward of the mountains, *mṣaʔ* roughly indicates 'south'.[6] Such differences in spatial reference terms that correlate to topographic variation, irrespective of how closely related the language varieties are, are found in other regions of the world in which languages have systems of absolute spatial reference (Palmer 2015; cf. also Rowley 1980 for Tirol; McKenzie 1997 for Sulawesi; Palmer 2002 for Oceania).

The following short text, 20131028_JibbaliEJ_J019_directions, recorded by the first author from Azad al-Kathiri in Jufa in eastern Dhofar describes spatial reference in Śḥerēt from the speaker's perspective. Azad is in his 50s. He used to herd goats, but now has a governmental job and lives in the new town of Jufa in eastern Dhofar beneath the dry mountains of Jabal Samhan. *d-ɛ̃̄ṣóʕ*, cognate with Mehri *u-mṣaʔ*, indicates the direction in which the flood waters run: in this case, seaward of the mountains *d-ɛ̃̄ṣóʕ* indicates roughly 'south' since the waters run into the sea, and *də-ḥāḳaʰl*, cognate with Mehri *(l-)ḥaḳ*, indicates roughly north. The Arabic terms in this text: *šimāl* 'north', *janūb* 'south' and *ʕabārah* 'that is to say', are probably used to ensure that the first author understands the text. In describing 'north', Azad also uses the adverb *ḥaṯ̱ih* 'up' and the place term *fɑ́jir* 'desert'; in describing 'south', he also uses the adverbial *b-āġɑ́ʰl* 'down', and says that this term may be used in addition to *də-ḥāḳaʰl* to describe 'south'.

We present this text and the following texts in broad phonemic transcription with the English translation following each speaker turn. The transcription system used is that of the *Zeitschrift für Arabische Linguistik*; in addition: superscript 'h' indicates preaspiration of phrase-final sonorants; above vowels, the tilde '~' indicates nasalisation; '/' indicates a brief pause in the speech; consonantal symbols between round brackets indicate inaudible, but phonologically present, consonants in pre-pausal position; superscript [A] indicates Arabic words or phrases. The text was recorded in the speaker's house in Jufa in WAV format, 44,000 Hz, 16 Bit, with a Marantz PMD661 solid-state recorder and Shure SM11 dynamic lavalier microphone. It was transcribed and translated by the first author, and checked in the UK in 2015 with Khalid Ruweya al-Mahri from the fishing village of Sadah.

2.1.1. Text 1: 20131028_JibbaliEJ_J019_directions

nṣurɛ̃̄t / min tɛl təfzɛ́z yuʰm / bi-ḳabəlɛ́t / m-tɛl tġ̃̄d yuʰm / də-ḥāḳaʰ(l) / [A]ʕabārah šimāl[A] / də-ḥāḳaʰ(l) / ḥaṯ̱ih / fɑ́jir / bi-[A]jinū(b)[A] / ḏ̱-ɛ̃̄ṣóʕ / b-āġɑ́ʰ(l) / ḏ̱anuh [A]janūb[A] b-yʕũ̄r yɔ̄ d-ɛ̃̄ṣó b-īʕũ̄r yɔ̄ aġaʰ(l) / min tɛl eḏ̱hēb yifɔ̄ṣ̱ yol eremnem /

[5] To say, 'I came from downstream', Central Omani Mehri has the phrase *nakak min ṣāwan* (Yahya al-Mahri, p.c.).

[6] In Mahriyōt, *mṣaʕ* means 'down; downstairs' and *ḥaḳ̱* 'up; upstairs'; in Omani Mehri, *l-ḥaḳ* also means 'inside'.

East / where the sun rises / and west / where the sun sets / inside / means north / inside / up / the desert / and south / south / down / that is south, people say *ḏ-ɛ̃ṣóʃ* and people say *b-āġaʰl* / where the flood waters flow towards the sea.

2.2. Temporal reference points and measurement of time

Several of our documented texts make reference to a lack of watches or clocks in the pre-Qaboos era, and time was judged traditionally by the position of the sun, the direction and length of shadows, and the degree of light. Day and night are described broadly as *nhīr* 'day' and *bi-ḥallay* 'night'. Some speakers use *k-aṣōbaḥ* more broadly in sense of *nhīr,* and *nhīr* can be used in the sense of *k-aṣōbaḥ*. For Omani Mehri, the main divisions of day and night are as follows:

fahag ~ k-affēgar 'first light; dawn'
śarḳ 'sunrise'
ṣūbīḥan 'early morning after sunrise'
k-aṣōbaḥ 'in the morning'
ḏwēlēban 'early mid-morning' – described by the second author as when you can see sun in front of you
ḏʔawban 'late morning'
nhūran 'midday'
ka-lʔasr ~ alʔasr 'mid-afternoon'
ġasrawwan ~ ġasarawwan 'late mid-afternoon'
ġasērēyan 'late afternoon'
kalʔaynī 'before sunset'[7]
kalāʔāni 'early evening'[8] – later than *kalʔaynī*
k-amgawzaʔ ~ agzēʔ 'once the sun begins to redden'[9]
ū-maġrāb 'sunset'
bi-ḥallīyēn ~ ḥāwēl ḏ-aʔāṣar 'first part of the night'
ḥallīw 'night'
bi-ḥallay 'at night'
faḳḥ ḏ-aʔāṣar 'middle of the night'
tōlī ḏ-aʔāṣar 'end of the night'

Of these terms, the diminutives *ṣūbīḥan* 'early morning after dawn before sunrise', *ḏwēlēban* 'early mid-morning', *ġasērēyan* 'late afternoon' and *kalāʔāni* 'early evening'

[7] The reference of *kalʔaynī* is not universally agreed. By some speakers, it is said to refer to when the livestock return to the homestead (Musallam Hazmay al-Mahri, p.c.). For others, *kalʔaynī* refers to the period from mid-afternoon to sunset (Ali al-Mahri, Nwēr Muhammad Bakrayt, p.c.).
[8] Some speakers describe *kalāʔāni* as coming before *kalʔaynī.*
[9] *agzēʔ* can also be used colloquially in the sense of 'sunset'.

describe times when the sun is lower in the sky than it is for their non-diminutive[10] cognates *k-aṣōbaḥ*, *ḏ?awban*, *ġasrawwan* and *kal?aynī*. The diminutive *bi-ḥallīyēn* 'first part of the night' describes a time when the darkness is less intense than for *ḥallīw* and *bi-ḥallay*; thus, diminution in height of the sun and in intensity of darkness is reflected linguistically through the diminutive morphological pattern.

The MSAL have several verbs to describe going and coming according to the time of departure. In Central Omani Mehri, traditional verbs of going are: *ġsūm* 'dawn to c. 7am', *ahhawgar* 'midday', *šūgūś* 'mid-afternoon', *śśōfaḳ* 'just before sunset', *abōṣar* 'twilight, early evening', *bār* 'night'. In western Omani Mehri, *haḳrawr* originally refers to going in the heat of the day.

Many of the nuances of these verbs have been lost in the speech of the younger generation, to be replaced by cover terms for 'to go', such as Central Omani Mehri and Mahriyōt *syūr* and *ghēm ~ jhēm*, Western Omani Mehri *haḳrawr*, and Śḥerēt *aġād*, with time of going expressed by time adverbials. Mehri *abōṣar* originally meant going after the sun had set and when the goer still had sufficient natural light to see. Today where *abōṣar* continues to be used, it commonly indicates going at any time at night. The first author's tentative hypothesis here is that the introduction of street lights now enables people to see at any time of night.

Traditional Omani Mehri verbs of coming are: *šaghūm* 'early morning', *ḳhēb* 'around midday', *watxaf* 'afternoon to sunset', *ṭwuh* 'night'. Today, *watxaf* remains in common usage, but otherwise *nūka* (*nūkaʃ* in Mahriyōt, *zḥam* or *nukaʃ* in Śḥerēt) is used to indicate coming at any time of day together with a time adverbial.

When the sun cannot be relied upon – for example, at times of thick fog during the monsoon period, time is judged by other natural features: flowers opening and continuing to move to face the sun even when the sun is obscured, the movements of dung beetles, and the sound of a large flying beetle that comes out at sunset (Miranda Morris, p.c.).

In measuring periods of time, reference is traditionally made to known activities and shared recognition of the time it would take to carry out these activities. Musallam Hazmay al-Mahri (from Dhahbun) in the following short extract from a text recorded with a Marantz PMD661 solid-state recorder and Shure SM11 dynamic lavalier microphone describes judging the length of time between sunset and evening prayers:

20141103_MehriDhahbun_M017_divisionofdayandnight

m-bayn agzēʔ / m-bayn aʃīśēʔ / wḳōn msēr ḏa-ġaśafīt min ḥawōdī / walā ġaśafītī ṯrayt / ḥābū ḏa-ʔāśaməh śī lā sāʃāt / ār ār ūṭōmah / walā ḥāzar ḏa-ḥalēb ḏ-ōśar bēr

Between sunset and supper [prayer time], it will be the time it takes to walk a bend in the wadi, or two bends. In the past, people didn't have watches. It was just that, or around the time it takes to milk ten camels.

[10] Or lesser diminutive – *ġasarawwan* is also diminutive (Lonnet 2003).

2.3. Figurative language and the environment

The de-coupling of language and the environment results in lexical attrition, as we have seen above. It also leads to lack of comprehension of non-literal language, such as metaphors, similes and metonymy. These are not peripheral linguistic phenomena, since a typical speaker employs around 5.88 non-literal expressions in five minutes of speech (Tosey, Sullivan & Meyer 2013), and many expressions we assume to be literal are, in fact, non-literal (Traxler 2011). In Modern South Arabian, metaphors and similes are frequently environment-specific: in Mehri, a man may be described as *axahēh sīmar* 'he looks like a *sīmar* [tree]',[11] encapsulating height, uprightness, slenderness and a shock of hair; a child may be compared to *śēḥaz* 'frankincense' because of her clinging nature; and exaggerations of amounts may be expressed through the simile *hīs abaṭḥ* 'like dust', e.g. *ḥābū hīs abaṭḥ* 'there are loads of/millions of people'. The importance of recognising directional terms appears in the Mehri figurative phrase: *yiġōrab mṣā min lḥaḳ lā* 'he doesn't know downstream from upstream'.

Poetry is famously rich in nature-figurative language: in Mehri, *xśawr* related to 'grue [colour]' has the sense of 'sea' only in poetry; and the Śḥerēt line *hɛz min ġarb / ṣarif d-irḥasɛn* 'wind from west / stones he licks' refers to a man from the west taking another's wife, who is so poor he has to lick stones. This allusion goes back to a custom in the past of people placing large stones on their stomachs to reduce the hunger pangs.

Terms are frequently introduced on first encounter of an object through extension in meaning: thus, *ḳalīfūt* in eastern Yemeni Mehri and *ḳalifōt* in Bathari (Miranda Morris, p.c.) have the secondary sense of 'spoon'. The original sense of 'bark [tree]' and the knowledge that tree bark was used in the recent past for stirring and as an eating implement is lost on many of the younger generation. In the second of our two texts below, we see *śaff* 'animal track' and *ḥaśśˀ* 'to track an animal' adopted to track an inanimate moving object – in this case a lorry.

Particles with a grammatical sense frequently emerge from an extension in function of words with a concrete sense. This process, commonly known as grammaticalisation typically involves a word with a lexical meaning, such as 'head', 'back' or 'mouth', coming to have a largely grammatical function, such as a preposition or uninflectable particle. Grammaticalisation is thought to have originated with Meillet (1958). Recent works on grammaticalisation include Hopper and Traugott (2003) and Fischer (1997). In Mehri, *axah* 'appears to be' in phrases such as *aġiggīt ḏīmah* ***axa****sēh bōkar* 'that girl has the stature of a young female camel' is a functional extension of the noun *xahh* 'mouth' (*xoh* in Śḥerēt) – the mouth may be the part of a person providing visual information from which first judgements are made; similarly, the mirative particle *śaf* in Mehri, *śɛf* in Śḥerēt in the sense of 'it transpired, as it happened, really' (Johnstone 1981, 1987; Watson 2012), is a metaphorical extension of the noun *śaff ~ śɛf* 'track, print'.

[11] *Boscia arabica*, a tree of the desert and drier mountains that looks like an opened umbrella (Miranda Morris, p.c.).

śaf ~ śɛf is a particle the first author had pondered for some time. It is translated by Johnstone (1981, 1987) as 'as it happened; it transpired that; probably', by Rubin (2010, 2014) as 'it happened/turned out that; as it happened/turned out', and analysed by Watson (2012) as a mirative particle. During a fieldwork session in the UK in 2015, Khalid Ruweya al-Mahri pointed out that the particle *śɛf* in Śḥerɛ̄t relates to the noun *śɛf* 'track'. The same interpretation for Mehri *śaf* was confirmed in conversation with several Mehri speakers during fieldwork conducted later in 2015 and 2016. Thus, we now interpret *śaf ~ śɛf* as resulting from grammaticalisation of the content word for 'track', *śaff* in Mehri, *śɛf* in Śḥerɛ̄t. Grammaticalisation frequently involves phonological reduction of the content word. Thus, Mehri *śaf* is reduced from the content word *śaff* through reduction of /ff/ > /f/: *śafs* 'it turns out that she' contrasts with *śaffas* 'her tracks' (cf. the text in 3.1 below).[12] Mehri *śaff* and Śḥerɛ̄t *śɛf* have the plural forms *śfūtan* and *ɛśfof* respectively, but as a particle *śaf ~ śɛf* shows no inflection. In both languages, the particle may occur in its bare form or with a pronominal suffix (Rubin 2010, 2014).

The grammaticalisation of *śaff ~ śɛf* to the particle *śaf ~ śɛf* can be understood when we consider that tracks, like fingerprints, reveal an indisputable identity that may otherwise not be recognised. From sight someone may believe they are following a camel from one herd, but on close examination of tracks discover they are tracking a camel from a different herd. The track, therefore, reveals the true identity of an animal or a person, and by semantic extension the particle *śaf/śɛf* is used where an action, object or event that was originally thought to be X turns out unexpectedly to be Y.

Many younger speakers use Mehri *śaf*, Śḥerɛ̄t *śɛf* accurately in sentences such as: *śīnak ḥaybīt akabs ḥaybaytī* ***śafs*** *ḥaybitk* 'I saw a camel I thought was mine, but it turned out to be your camel', but fail to appreciate the link between this particle and the vital social importance of tracking in the past. The track reveals identity, and the ability to track accurately may determine life or death. In a recent discussion, a teenage Śḥerɛ̄t speaker asked whether *śɛf* came from *śof* 'hair'. Ali Ahmad al-Mahri, a bilingual speaker of Mehri and Śḥerɛ̄t, explained this lack of awareness among young people living in Salalah as resulting from a lack of earth or sand paths in the town: tracks cannot be imprinted on solid asphalt, and therefore the activity and terminology of tracking is absent in a paved environment.

3. Texts on tracking

The Documentation and Ethnolinguistic Analysis of Modern South Arabian (DEAMSA) project has collected a number of texts on tracking. Some describe the art of tracking, and others present factual stories about tracking. Here we present two factual stories. The texts were transcribed by the first author, and checked several times with

[12] *śaffas* 'her tracks' may be realised as *śafs* when non-focussed (compare *xaffas* and *xafs* in the third turn from the end), but *śafs* 'it turns out that she' can never be realised as *śaffas*.

the second author. The sound files and corresponding ELAN files will soon become accessible through ELAR (http://elar.soas.ac.uk/deposit/0307). Ahmad Xamis al-Mahri's account is already accessible through the Semitic Sound Archive in Heidelberg (http://www.semarch.uni-hd.de/), and has been previously published in transcription and translation in *The Structure of Mehri* (Watson 2012).

The first text was recorded in Rabkut by Abdullah al-Mahri (M001) with his friend Ubxayt Saeed Shel al-Mahri (M008) in WAV format, 44,000 Hz, 16 Bit, using an Olympus LS11 digital recorder. Both interviewer and speaker are from the gravel desert in Rabkut, and are in their early twenties. The second text was recorded in Rabkut by the first author in WAV format using a Marantz PDM661 solid-state recorder and PG58 microphone. It describes the speaker's first encounter as a young child with a motorised vehicle. Ahmad Xamis al-Mahri is in his 70s, and today lives in the gravel desert village of Rabkut.

3.1. Text 2: 20140612_MehriRabkut_M008_cameltracking

20140612_MehriRabkut_M008_cameltracking describes the speaker's great uncle adventure after losing one of his camels. The speaker begins by saying how camel herders are able to retrieve their livestock through following their tracks. In the end, his quest takes him to the lost camel's foal who had inherited the same track features as her dam. There is probably an error in the time period discussed: the speaker claims the camel had been lost for over a year; however, for the camel to have given birth and the camel foal to be old enough to display the same track features as her dam, it is probable that the time period was around two years. We see use of the mirative particle *śaf* in the turn:

M008: *wa-ṭlōt* ***śaf*** *ḏikm ḥaybīt / [...] ār ḥaybīt ḏa-ḥaybith ahah / wa-ṭlōt l-śaffas /*

And it turned out that in fact that camel was just the daughter of his camel, yes, and it had the same tracks.

M001: *hinay ūbxayt bar saʃīd bar shēl / wa-mhaddal tay / ših ḳassēt / mhaddal tay bīs / aywah ūbxayt /*

I have Ubxayt bar Saeed bar Shel with me, and he is going to tell me, he has a story he will tell me about. Okay, Ubxayt.

M008: *aḥōm lāmēr hūk ṣarōmah fīh ḥābū bāṣ́ hām fḳawdam hibērīham walā ġṣawb līham / yišadlīl l-ḥaybīt bi-śaffas /*

I want to tell you know there are some people, if they lose their camels or the camels went away from them, they can find the camel by its tracks.

M001: *hmm /*

Okay.

M008: *yikawn yiġarbam tēs l-ḥōh lā lākin wat kūsam aśfūtan ḏa-hibērīham / yiġarbam tēsan / yikawn bār ḏa-ġrawb ḥābū ykūn šīham / ykūn šīham ḥanīt /* [A]*xibrah*[A] *min bād ḥibīham w-aʃimīham / ykawn ḏa-ġrawb bi-śaff ḏakmah /*

They won't know where it is, but when they find the tracks of their camels, they know them, they will know. People have, they have, what do you call it, knowledge from their fathers and grandfathers. They know from that track.

M001: *yiġarbam ḥaybīt min śaffas /*
They know the camel from its tracks.
M008: *yiġarbam ḥaybīt min śaffas / wa-šadəlīl bīs tā kasyam tēs /*
They know the camel by its tracks, and they are directed by it until they find it [the camel].
M001: *ahah /*
Yes.
M008: *ṭawr hōh kūṯōn līkam bi-ḳassēt ḏ-aʔōmī /*
Once, I'm going to tell you a story about my grandfather.
M001: *hmm /*
Okay.
M008: *aġā / ḏa- / ḥayb / ḏa-ḥāmay / hammah mḥammad ba.. / bar salīm /*
The brother of my mother's father. He was called Muhammad bar Salim.
M001: *hmm /*
Okay.
M008: *ṭawr ḏ-isyūr / ḥaybī(t) ḏikmah bār ġabarūt ḥawlas / al-ād kisīs lā /*
Once he was walking. That camel had been [lost] for a year and he still hadn't found it.
M001: *ahah /*
Okay.
M008: *hīs maxṭār ḏa-ḥarmah ḏ-īgayr yikays śaff ḏ-ībīt /*
Once while he was walking he found the tracks of a camel.
M001: *bār hīs bār hīs snēt /*
She'll have been, she'll have been [lost] for a year.
M008: *bār hīs ūḳōn snēt wa-zōyad /*
She'll have been [lost] for around a year or more.
M001: *ahah /*
Yes.
M008: *bār ġabarūt ḥawl / yikays śaff ḏa-ḥaybīt ḏikmah yiġarbas / hankūr ḏa-sēh ḥaybith linn aġayg yiġōrab aśfūtan ḏa-hibēr / syūr ġayg štaba štabays štabaʔ / bi-śaff ḏakm wa-hēh ḏ-ītabah ḏ-ītabah yitabah anhūrah / attā l-hīs kalʔay.. ḳarayb / anhūran ūṭōmah yikays / aśśaff ḏakm w-wīṣal* [A]*ʕazbah*[A] */*
She'd been [lost] for a year. He found the tracks of that camel, he recognised them. He thought it was his camel, because the man knew the tracks of camels. The man went on and followed and followed it, and followed those tracks, and he kept following them all day, and then by around evening [sic], [or] midday like that he found those tracks and came to a camel camp.
M001: *mbarīk ḏa-bēr /*
[That we would call] *mbarīk ḏa-bēr*.
M008: *mbarīk ḏa-bēr /*
mbarīk ḏa-bēr
M001: *hmm /*
Okay.
M008: *kūsa ḥābū ḥalakmah wa-ḳlūb līham salōm / wa-klūṯ līham bi-ḳassēt kallas / amūr hōh ḥaybīt / ḏa-fḳadak tēs / wa-bār šīs ūḳōn snēt wa-zōyad / amūr wa-tabak tah śaff ḏohm tā*

kisk tah hinīkam bawmah / amawr hēh ḥābū xayban slōb / tā wat kalaynī hibēr kluh / śnē hēt min śaff ḏōmah wa-ġlēḳ min ḥaybitk /

He found people there and greeted them. And he told him the whole story. He said, 'I [had] a camel that I lost, and it will be over a year ago'. He said, 'and I followed the tracks until I found them here where you are.' The men said to him, 'In that case, wait until evening when the camels return and you check those tracks and look for your camel.'

M001: *hmm /*

Okay.

M008: *amūr tamām / ḥgūr hīs kalʔaynī hibēr kluh / gruh bark aśfūtan ḏa-hibēr lyakm tā kisyah / wa-tabayh bark hibēr lyakm wa-tabayh wa-hēh ḏ-īgayr anhūrah anhūrah /*

He said, 'Okay', and waited until the camels returned in the early evening. He went around the tracks of those camels until he found it [i.e. the tracks he was looking for] and he followed them among those camels and followed them and kept on going.

M001: *hibēr mēkin /*

[There were] lots of camels.

M008: *hibēr mēkin hēh / wa-hēh ḏa-ġrūb ār śaff ḏa-hankarih ḏakm ḏa-yġarbah /*

[There were] lots of camels. And he recognised tracks that he thought were ones he knew.

M001: *ahah /*

Yes.

M008: *yitabah / tā wīṣal śaff ḏakm hāl ḥaybīt /*

And he followed them until those tracks came to the camel.

M001: *aywah /*

Yes.

M008: *amūr aḥ-ḥābū ḏīm sēh ḥaybaytī / wa-ḥaybīt xtalafūt l-ād sēh ḏayk ḏa-ʔāṣamis lā / ksīs ār ḳannitt / šxəbūr ḥābū min ḥaybīt ḏīmah amūr hōh kisk ḏīm sēh ḥaybaytī amawr xayban ḥaybīt ḏīmah / āṣamis bār nkatan / ē nkatan ḥaybīt /*

He said to the men, 'That is my camel, but the camel is different, it isn't the same as it was.' He found it was still young. He asked the men about that camel and said, 'I've found this to be my camel.' They said, 'Okay, that camel, a while ago it came to us, a camel came to us.

M001: *ahah /*

Okay.

M008: *wa-haḳṭawt / w-īs haḳṭawt ḥaybīt mtōt wa-xūfūt min sirīs ḥabrits ḏīmah /*

'And it gave birth, and when it gave birth, the camel died and left behind it its daughter.'

M001: *ahah /*

Yes.

M008: *wa-ṭlōt śaf ḏikm ḥaybīt /*

And it turned out that in fact that camel,

M001: *ār ḥabrīt ḏa-ḥaybith ḏayk ḏa-haġṣabīs /*

was just the daughter of the camel that he had lost.

M008: *ār ḥaybīt ḏa-ḥaybith ahah / wa-ṭlōt l-śaffas /*

was just the daughter of his camel, yes, and it had the same tracks.

M001: *wa-ġarbīs l-śaffas /*
And he recognised it from its tracks.
M008: *wa-ġarbīs min śaffas ḏikm /*
And he recognised it from those tracks.
M001: *subḥān allāh /*
Praise be to God!
M008: *ṣarōmah ḥābū yikawn yiġarbam li-hibērīham / wa-ḏīmah xubrah /*
Now people know their camels, and that is expertise.
M001: *bāṣ ār lā /*
But some don't.
M008: *yūrīṯs min sār ḥibīham min sār aʔamīham /*
They inherit [the knowledge] from their fathers and from their grandfathers.
M001: *lākin hēm ḥābū kall lā ʔār bāṣ min bāṣ /*
But not everyone, just some of them.
M008: *ʔār bāṣ min bāṣ /*
Just some of them.
M001: *yiġarbam aśśaff ḏa-ḥaybīt /*
They recognise the tracks of camels.
M008: *ḥasab ḏa-ṯafaritīsan / bāṣ yikawn məḳtarabūtan wa-śī yikawn ḏa-ratḥaḳam /*
According to the nails [on their pads]. Some are close together and some are far apart.
M001: *aywah /*
Yes.
M008: ^{A}*fīh*^{A} /
And there is ...
M001: *xaffas ykūn /*
Their pads are ...
M008: *wa-*^{A}*fīh*^{A} *xaffas ykūn [ḏ-]ārayṣ / wa-śī xafs ḳṭayn /*
And some have a wide pad, and some have a narrow pad.
M001: *ahah /*
Yes.
M008: *mēkin ḥanūtan tā yġarbam tēsan / b-aśfūtan /*
There are lots of things [they have to know] to recognise them by their tracks.

3.2. Text 3: My first experience of seeing a car

In the following text, we observe the use of environment-related metaphor: the speaker has not yet seen a large inanimate object move by itself and compares the tanker he sees with a hill or mountain, *karmaym*, due to its size. When the men then follow the tanker's tracks, the speaker adopts the term *ḥaṣṣ* 'to track prints or tracks' and *śaff* 'track, print' otherwise used to indicate human or animal tracks. In line 5, we observe use of both the mirative particle *śaf* 'it turned out to be' and the concrete noun *śaff* 'track' from which it was grammaticalised.

1. *awwal marrah fī ḥayātī śīnak sīyaryat /*
   The first time in my life I saw a car.

2. *b-amahrayyat / śīnak sīyaryat ūṭōh nḥah bātī ʔawraḥ b-aḳāʔ ḏa-bātī kākūt / aśōnī tankar / aśōnī tankar tasyūr bi-ṣayḥ / aśaynas man aġawf w-aʔōmar aḥ-ḥābū bark aḳābī ḏīmah karmaym taghūm / wa-yiṣṣak /*
   In Mehreyyet! I saw a car there, we were at Awrah in the area of Kābūt, I saw a tanker. I saw a tanker moving in the gravel desert. I saw it from above and said to people, 'I think that's a mountain moving!' And I was afraid.

3. *wa-bḳaśk l-hāl ḥābū la-ḥlawk / tā nakak hāl ḥaybī / wa-ḥābū / wa-mġōran ʔamūr karmaym tasyūr lā /*
   I ran to the people over there, until I reached my father and the other people. Then he said, 'A mountain doesn't move!'

4. *syawr ḥābū man ḥlakmah ḥõh ḥõh tā wiṣlan hāl amkōn man hāl śīnak tēs /*
   The people went from there, until we got to the place where I had seen it.

5. *ḥaybī ḥaśśś sīyaryat śaf tankar taghūm ḥaśśan aśśaffas / bi-ṣayḥ taghūm /*
   My father tracked the car's tracks. It turned out it was a lorry moving. We followed its tracks. It was moving in the gravel desert.

6. *ʔamūr aġayg la-hāl ḥaybī / ʔamūr ḏīmah hammas sīyaryat / wa-ḏīmah sēh karmaym lā taghūm hēt śīnak ār sīyaryat w-assīyaryat bār ghamūt /*
   A man told my father, he said, 'That is called a *sīyaryat* [car]. That isn't a mountain moving, you just saw a car!' And the car had gone.

4. Conclusion

Language enjoys a close relationship with the natural environment, particularly in regions where people share their everyday lives with nature. This relationship can be seen in the lexis of a language and in the use of figurative language. Dhofar until the recent past was a region in which people made, gathered, farmed or bartered materials for food, water, shelter, tools, medical treatment and clothing. Today no one relies entirely on the land and the sea. Here and in other regions of the world that have experienced rapid sedentarisation and urbanisation, a break in the human–natural environment relationship is reflected in a loss of use and/or of understanding of lexemes and figurative expressions. This paper addressed some of the ways in which the natural environment is reflected in language and in which a degradation of the environment can lead to a degradation of the lexis of a language. We plan to follow up this initial paper with further work examining the relationship between the Modern South Arabian languages and the natural environment.

BIBLIOGRAPHICAL REFERENCES

Alfaifi, A.H.J. (2016), *Aspects of the morpho-syntax of the Faifa dialect of Saudi Arabia: A descriptive study*, PhD thesis, Ulster University

Fischer, O. (1997), *On the status of grammaticalization and the diachronic dimension in explanation,* in "Transactions of the Philological Society" 95: 149–87

Hopper, P.J. and E.C. Traugott (2003), *Grammaticalization*, Cambridge, CUP

Johnstone, T.M. (1981), *Jibbāli Lexicon*, Oxford, OUP

Johnstone, T.M. (1987), *Mehri Lexicon and English-Mehri Word-List*, London, Routledge

Lonnet, A. (2003), *L'accumulation des déictiques: l'expression de «maintenant» en sudarabique moderne*, in J. Lentin & A. Lonnet (eds), *Mélanges David Cohen: etudes sur le langage, les langues, les dialectes, les littératures, offertes par ses élèves, ses collèges, ses amis*, Paris, Maisonneuve & Larose: 421-438

Macfarlane, R. (2015), *Landmarks*, London, Hamish Hamilton

McKenzie, R. (1997), *Downriver to here: Geographically spatial deictics in Aralle-Tabulahan (Sulawesi),* in G. Senft (ed.), *Referring to Space: Studies in Austronesian and Papuan languages*, Oxford, OUP: 39-51

Meillet, A. (1958), *L'évolution des formes grammaticales*, in "Linguistique historique et linguistique générale", Paris: Champion: 130-148

Morris, M. (in press), *The linguistic situation in the Central Oman Mountains: Baṭḥari*, in "The mountains of Oman. An illustrated reference to nature and society", Hildesheim, Georg Olms Verlag

Palmer, B. (2002), *Absolute spatial reference and the grammaticalisation of perceptually salient phenomena*, in G. Bennardo (ed.), *Representing space in Oceania: Culture in language and mind*, Canberra, Pacific Linguistics: 107-157

Palmer, B. (2015), *Topography in language: Absolute frame of reference and the Topographic Correspondence Hypothesis*, in R. De Busser & R.J. LaPolla (eds), *Language Structure and Environment: Social, Cultural, and Natural Factors*, Amsterdam, John Benjamins: 177-226

Rowley, A. (1980), *Richtungs und Ortsangabe in der Mundart von Florutz (Fierozzo) im italienischen Tirol*, in A. Rowley (ed.) *Sprachliche Orientierung I. Untersuchungen zur Morphologie und Semantik der Richtungsadverbien in ober-deutschen Mundarten*, Bayreuth, Sprach und Literaturwissenschaftliche Fakultät: 73-96

Rubin, A. (2010), *The Mehri Language of Oman*, Leiden, Brill

Rubin, A. (2014), *The Jibbali (Shaḥri) Language of Oman*, Leiden, Brill

Thomason, S.G. (2015), *Endangered Languages: An Introduction*, Cambridge, CUP

Tosey, P., W. Sullivan & M. Meyer (2013), *Clean Sources: Six metaphors a minute.* http://epubs.surrey.ac.uk/806935/–accessed 6th January 2016

Traxler, M.J. (2011), *Introduction to Psycholinguistics: Understanding language science*, New York, Wiley-Blackwell

Watson, J.C.E. (1993), *A Syntax of Ṣanʕānī Arabic*, Wiesbaden, Harrassowitz

Watson, J.C.E (2012), *The Structure of Mehri*, Wiesbaden, Harrassowitz

JANET C.E. WATSON • Studied Arabic & Islamic Studies at the University of Exeter; she moved to SOAS, London to study Linguistics and completed a PhD on the phonology and morphology of Yemeni Arabic dialects in 1989. She has held academic posts in the UK at the Universities of Edinburgh, Durham and Salford. She has also held visiting posts at the universities of Heidelberg (2003-4) and Oslo (2004-5). On 1st May 2013, she took up the Leadership Chair for Language@Leeds at the University of Leeds, and was elected Fellow of the British Academy in July 2013. Her main

research areas are the phonetics, phonology and morphology of modern Arabic dialects and Modern South Arabian, and the documentation of Modern South Arabian. She has recently received funding for a research network to examine the relationship between language and nature in southern and eastern Arabia with partners from Qatar University.

E-MAIL • j.c.e.watson@leeds.ac.uk

LANGUAGE AND IDENTITY AMONG THE 'ARABS OF THE COAST' IN IRAN AND THE ARAB GULF STATES

Dénes Gazsi

ABSTRACT • This paper explores the efforts of auto-documentation and outreach of the 'Arabs of the Coast', an understudied segment of Arab population in Iran's Persian Gulf coast and the Arab Gulf states. The presence of a large number of Arab tribes on the Iranian shore had been well-documented by British officials until the early 20th century CE. Afterwards, a general consensus among the public in the region emerged that these Arabs (widely referred to as *Hōla*, *Hula*, *Huwala* or *Hawala*) had all relocated to the Arabian Peninsula, from Iran. Using procedures of Discourse-Centered Online Ethnography (DCOE), study analyzes the social media activity and face-to-face interviews with 'Arabs of the Coast' with the aim of demonstrating how they began to work toward a general recognition of their presence in Iran and the Arab Gulf states, and how they maintain their Arab identity, culture, traditions and language in the Persian State. The paper also sheds light on the hardships they endure during the reintegration process after they move to the United Arab Emirates and Qatar.

KEYWORDS • Persian Gulf, Arabic dialects, Arabs of the Coast, Hula Arabs, Huwala Arabs, bilingualism, social identity, Arabs in Iran, social media

"*Identification with a language and loyalty to it are aims of nationalism... Language is the medium which makes the nation as an 'imagined community' imaginable*"
Suleiman 2003: 29-30

1. Introduction

Hingām, a calcareous island located south of Qišm Island in the Islamic Republic of Iran, is a popular tourist destination for scuba diving, dolphin-, turtle- and shark-watching trips as well as history aficionados[1]. The island, merely 6 km long and 5 km

[1] For easier identification on maps, the most common English spelling of geographical names appearing in this study are listed at the end of the paper.

wide, had a turbulent past with continued struggles by the Portuguese, the British and the Sultanate of Muscat to gain control over its territory[2]. Today, English harbor buildings along with a coal storage, Portuguese shipwrecks, a crocodile farm, a shrimp aquaculture and the silver beach constitute the attractions that provide livelihood for the 417 people who live on the island in 110 households[3]. Almost a century ago, however, the allegiance of the island's population was shaped by different factors, as evident from a confidential report titled *Position and Rights of His Majesty's Government in the Island of Henjam, The Sheikh of Henjam and his Affairs* drafted by the India Office of the British Government on September 26, 1928[4]. Paragraph 3 reads:

> The island, the only place of importance on which is the village of Henjam, an Arab settlement dependent principally on pearl fishing, is inhabited by Arabs belonging to the Beni Yas tribe of the Arabian coast. The Beni Yas are said to have migrated to it from the Arabian coast about a century ago, with the permission of the Sultan of Muscat, at a time when Henjam was definitely claimed and administered by that State. They are reported consistently to have refused to acknowledge Persian sovereignty or to accept Persian nationality, and to have made it clear that they would rather leave Henjam than become Persian subjects. They have intermarried throughout with the Beni Yas of the Arabian coast; their commercial dealings and their pearling ventures are all connected with that tribe; their Sheikh is the father-in-law of the Trucial Sheikh of Dabai, and the ruling family of Dabai and their subjects claim to have 3 1/2 lakhs of rupees sunk in the concerns of the Henjam branch of their tribe.

This account elucidated that the island dwellers at the time were members of the Banī Yās, a powerful tribal confederation that had ruled over a vast domain from Dubai to Qatar. They were reluctant to accept Persian sovereignty, and instead maintained extensive cultural, religious and trade ties with their kin on the Arabian Peninsula[5]. Arab tribes were not limited to this tiny island, but their presence had a well-documented past on the Iranian coastline of the Persian Gulf. According to travelogues from the 18th to the 20th century CE and British archival materials dating back to the British Residency of the Persian Gulf, Arab tribes inhabited most fishing and pearling villages on the Iranian shore[6]. Towns with great strategic importance (Bandar ʿAbbās, Bandar Linga)

[2] Lorimer 1970: II A, 629-633. Floor 2010: 56, 145, 150.

[3] *Saršumārī* 2011-2012. The main modern settlement on the northeastern tip of the island is Hingām-i Jadīd 'New Hingām', but 25 people in a total of 5 households still inhabit two tiny villages, Hingām-i Qadīm 'Old Hingām' and Qīl, on the southern and western coastline respectively.

[4] Schofield 1990: 545.

[5] The leaders of this confederation are the Āl Nahyān from the Āl Bū Falāḥ tribe, current rulers of the Emirate of Abu Dhabi. The second significant branch of the Banī Yās is the Āl Maktūm from the Āl Bū Falāsa tribe, current rulers of the Emirate of Dubai.

[6] Eminent travel accounts of Arabia include Carsten Niebuhr: *Beschreibung von Arabien* (1772) and *Reisebeschreibung von Arabien und anderen umliegenden Ländern* (1774-78). The most recent

and a dozen islands dotted across the area had become strongholds of Arab culture. Specifically valuable is Lorimer's monumental work on his travels along Iran's Persian Gulf coast, where he visited every settlement, and listed all Arab tribes and the number of their members residing there. In contrast with the majority Shiite Persians of Iran, these tribes were Sunni Arabs, who had migrated from Central Arabia to the eastern coast of the Peninsula, and after having crossed the waterways to the Iranian side, formed Arabic-speaking communities. Major tribes involved in this migration were the Qawāsim, Marāzīq, Āl Ḥaram, Āl ʿAlī, Āl Naṣūr, al-Mālikī, ʿAbādila, Banī Tamīm, Banī Ḥammād and Banī Bišr[7]. They extended their influence over the adjoining regions, forming so-called Emirates along the Iranian shore. A prominent example of Arab influence is Bandar Linga: at the turn of the 19th and 20th century CE, the town and surrounding lands were controlled by the Qawāsim and Marāzīq, who earned their livelihood from charcoal and firewood trade throughout the Persian Gulf region[8]. The Qawāsim had originally moved to the Iranian coast and three islands from the present-day Emirate of Ras al-Khaimah and Sharjah, which they rule to this day[9]. In 1903, Bandar Linga's population of around 12,000 had approximately 5,000 Arabs and 5,000 Persians. According to Lorimer, Hūwala Arabs (his generic term for Arabs on the Iranian coast) numbered approximately 14,000, spreading from the hinterland of the town of Kangān in the north as far as Bandar ʿAbbās in the south.

Since Lorimer's extensive descriptions, barely any information has been gathered and published about Arabs on the Iranian Persian Gulf coast. It may seem that their glorious past abruptly vanished, because the settlements they once inhabited now bear Persian names. This paper does not discuss the history of Arabs on the Iranian shore during the remainder of the 20th century CE, but it rather sheds light on how they are sending life signals to the outside world through social media, and how their identity and language are being shaped in the 21st century. First, the study briefly outlines the reasons for reverse migration of Arabs to the Arabian Peninsula. Second, it describes the terminology and methodology of the research. Third, it offers a classification of the Arab population in question. This is followed by the main part of the article that provides a detailed analysis of how 'Arabs of the Coast' have turned social media into a tool that makes the world aware of an understudied segment of ethnic composition in the Gulf region. The paper highlights how they take advantage of social media to nurture their language, culture, heritage, traditions, and the themes they debate online and in face-to-face interviews. Observations on their language use and linguistic attitudes will also be made.

---

publication on this topic is Al-Anṣārī, J. Ḥ (2014), *Qabāʾil ʿArab al-Hawla fī ʿUyūn al-Raḥḥāla wa al-Mustašriqīn*, Bairūt, al-Dār al-ʿArabīya lil-Mausūʿāt. Archival materials include Burdett, A. L. P., Seay, A. (Eds.) (2000), *Iran in the Persian Gulf 1820-1966*, 6 Vols, Cambridge Archive Editions, and *Islands and Maritime Boundaries of the Gulf* edited by R.N. Schofield.

[7] Bosworth 2012. Niʿma 2001: 7-54. Ṣiddīq 1994: 21-28.

[8] Lorimer 1970: II B, 1098.

[9] The three disputed islands are Abū Mūsā, Greater Tunb (Arabic: *Ṭunb al-Kubrā*, Persian: *Tunb-i Buzurg*) and Lesser Tunb (Arabic: *Ṭunb al-Ṣuġrā*, Persian: *Tunb-i Kūčak*).

2. Reverse Migration of Arabs to the Arabian Peninsula

The vibrant Arab culture began to fade when, during the 20th century CE, political and economic trends contributed to the reconfiguration of the long-standing ethnic composition of the Iranian coastline. At least four major economic and political factors can be identified:

1. Lucrative job opportunities in the developing oil industry lured 'Arabs of the Coast' to the Arabian coast after vast oil reserves had been discovered in Bahrain in 1932, in Kuwait and Saudi Arabia in 1938, in Qatar in 1939 (natural gas fields were discovered in 1971), and in the UAE in the 1960s[10]. This process coincided with the gradual decline of the pearling industry throughout the region.

2. In January 1936, the era called *Kašf al-Ḥijāb* 'Uncovering of the Veil' began with a decree issued by Reza Shah Pahlavi (1925-1941) banning all types of female veils and several varieties of traditional male garments[11]. The ban alienated conservative Iranian women and minorities such as Arabs, who had been proudly wearing their traditional attire, the long white robe (*ṯaub*). Although the official measures were slightly relaxed under Mohammad Reza Shah Pahlavi (1941-1979), the enforcement of the ban drove many Arabs to the opposite coast of the Persian Gulf, where wearing this type of clothing as well as headscarves had been the norm[12].

3. The British Empire terminated its protectorate over Kuwait in 1961, and announced its decision in 1968 to withdraw from the Persian Gulf, paving the way for the independence of Bahrain, Qatar, the United Arab Emirates in 1971. Subsequently, 'Arabs of the Coast' returned in waves to their original homeland on the Arabian Peninsula, where they could easily adapt to the dominant Arab culture and receive schooling in their mother tongue.

4. The "Persianization" policy during the reign of Reza Shah Pahlavi and Mohammad Reza Shah Pahlavi, which emphasized and propagated Iran's Persian ethnic identity to the detriment of minorities. Their pan-Persian nationalist policy lead to the centralization of power in Tehran, from where Persian-speaking governors and mayors were dispatched to non-Persian-speaking provinces[13]. This was particularly heavy on Arabs, towards whom Persians had cultivated a sense of antagonism. The culmination of the anti-minority sentiments came after the Islamic Revolution in 1979, which took a heavy toll on Arabs along with the growing anti-Sunni stance of the government.

[10] Family members of several informants I interviewed had moved to the Arab Gulf states to join the workforce in the oil industry.
[11] Katouzian 2000: 335-337.
[12] A video showing the aftermath of the ban and the resistance of local Arabs in ʿAināt and Kalāt villages, augmented with the narrative of eyewitnesses, is available on YouTube: https://www.youtube.com/watch?v=d7eeDndvO1g
[13] Haddadian-Moghaddam, Meylaerts 2015: 858.

Article 15 in Chapter II of the Constitution of the Islamic Republic of Iran asserts that Persian is the official language of the country, and "official documents, correspondence, and texts, as well as text-books, must be in this language and script. However, the use of regional and tribal languages in the press and mass media, as well as for teaching of their literature in schools, is allowed in addition to Persian[14]." Although the Article makes provisions for the minority languages of Iran, these languages are deprived of formal status and are not officially regulated by authorities[15]. Multilingualism is rather deemed as a form of threat to Iran's national unity and territorial integrity, while a non-translation policy is in effect for all government and administrative institutions[16]. The need to restrict minority languages is bundled into the policy of advancing the hegemony of the Persian language, which falls into the purview of the Academy of Persian Language and Literature (*Farhangistān-i Zabān wa Adab-i Fārsī*). The Academy is one of the primary government agencies tasked with Language Planning, and its mandate is to promote the Persian language in all aspects[17]. Article 16 of the Constitution makes the teaching of Arabic, "the language of the Qur'an and Islamic texts and teachings" mandatory "in all classes of secondary school", resembling the education policies of the Pahlavi period. Arabic language education, however, solely focuses on Classical religious texts, and not communication skills in Modern Standard Arabic or modern Arabic literature.

3. 'Arabs of the Coast' in the 21st century

The most vital element for 'Arabs of the Coast' to define and promote their ethnic identity is the promotion of Arabic language. This is supplemented with the dissemination of Sunni Islam and encouragement of Arab customs and traditions such as clothing, food, poetry recitals, songs, wedding celebrations. Diglossia is attested in both Arabic and Persian. The high (H) and low (L) variety for Arabic are Modern Standard Arabic (MSA) and Colloquial Arabic respectively, in our case Gulf Colloquial Arabic (GCA). For Persian, these are Modern Standard Persian (MSP) and Colloquial Persian (CP). CP is the standardized form of the dialect of Tehran used throughout Iran in informal oral and written communication. 'Arabs of the Coast' speak various dialects of Gulf Colloquial Arabic (GCA) that varies from village to village depending on the area of the Arabian Peninsula from which the tribe had originally migrated to Iran. During my interviews and on social media, I have witnessed dialects close to the Kuwaiti, Qatari and Emirati vernaculars. 'Arabs of the Coast' regard Persian as a language imposed on them by 'foreigners' (*ajānib*, whereby they mean Persians), though in private conversations they

[14] The Constitution is accessible in English at http://www.servat.unibe.ch/icl/ir00000_.html.
[15] Riazi 2005: 108.
[16] Sheyholislami 2012: 21. Haddadian-Moghaddam, Meylaerts 2015: 859-861.
[17] Riazi 2005: 110.

frequently expressed their fascination with and respect for the millennia-old Iranian culture. Codeswitching is a salient characteristic of bilingual Arabs, where the nature of codeswitching is situational and transactional, both inter- and intra-sentential. The choice of language and variety is determined by conversation topic, the interlocutors' identity and their relationship to each other. In the spoken and written code, 'Arabs of the Coast' often engage in a tetra-glossic switching between Modern Standard Arabic (MSA), Gulf Colloquial Arabic (GCA), Modern Standard Persian (MSP), Colloquial Persian (CP) or any of its local vernaculars. Individuals I followed online or interviewed in person are coded according to their current country of residence: Iran (IR1-10), United Arab Emirates (UA1-4), Qatar (QA1). The number representing the speaker follows a two-letter combination. The juxtaposition of Arabic and Persian lexemes or sentences will occur in the texts, but to avoid rendering the analysis incomprehensible, an Arabic-based phonological transcription is followed. Recorded materials are transcribed, while online threads remain in their original format and spelling.

A widespread term to designate Arabs on the Iranian coast, and especially those who returned to the Arabian Peninsula, is *Hōla* (in Arabic هولة, variously referred to as *Hula, Huwala or Hawala*[18]). A vibrant discussion about the etymology of the term is underway on social media, some claiming the word to be a derivation of the Arabic root *ḥ-w-l* in the meaning 'those who have moved', while others point to a possible origin in the Haul desert in Iraq[19]. The Iranian Arab informants I interviewed for my research seemed reluctant to identify themselves as *Hōla*, and preferred the terms 'Arabs of the Coast' (*ʿArab al-Sāḥil*), 'Arabs of the Eastern Coast' (*ʿArab al-Sāḥil al-Šarqī*), 'Arabs of the Persian Land' (*ʿArab Barr Fāris*) or 'Arabs of Persia' (*ʿArab Fāris*). A local historian from the village of Fawāris (IR1) expounds on the recent shift in the ethnic composition of the region, and refutes the validity of the term *Hōla*[20]. Following is the transcript of a video interview with him in GCA conducted by UA1, whose advocacy for the 'Arabs of the Coast' will be examined later in this paper[21]:

> UA1 [GCA]: *ʿadkum muṣṭalaḥ əl-hōla hinā?*
> IR1 [GCA]: *lā, hinā muṣṭalaḥ əl-hōla mā miš. hād əs-sāḥil fīh yiskinūna qabāyil əl-ʿarab. min kangūn u bnak li-tūṣal linya hād əs-sāḥil wi l-jizir baʿd illi tinzil hād əs-sāḥil jizīrət əš-šēḥ, ḥatta jizīrət ḫārg illi fi šamāl būšahar muqābil gināwa ʿād yiskinūhā nās ʿarab, jizīrat ət-tīn ṣōb gināwa. əs-sāḥil hū sikkān əs-sāḥil əl-aṣlīyīn əl-ʿarab. bass hād əl-iḫtilāṭ illī ṣār, fi murūr əz-zamān, fi nās fi ṣifit yaʿnī ʿummāl jau, bugau hinā min əl-ʿajam. fi nās fi ṣifit riyā əl-ġanam, ḥaddāda, mā adri, hāy əl-ašyā*

[18] Al-Anṣārī 2011: 48. Al-Dailami 2014: 301-302.
[19] The possible connection to Iraq was first raised by the Ruler of the Emirate of Sharjah, Sheikh Sultan bin Muhammad Al-Qasimi, a Gulf historian.
[20] Fawāris is located in the Kūškinār district of Pārsiyān sub-province (formerly called Gāwbandī sub-province) in Hurmuzgān Province, and has 279 residents in 69 households (*Saršumārī* 2011-2012).
[21] The video can accessed on this link: https://www.youtube.com/watch?v=SUZy-oVmufk.

yaʿnī. u baʿd ham zādat yaʿnī ʿalašān əl-ḥīn əš-šarikāt wa l-ašġāl u ġēru li-hād əl-manṭiqa jau nās. wa hāy əl-manṭiqa miyya fi l-miyya gabul yaʿnī ḫamsīn sana mā tgūl yaʿnī akṯar killa sikkān hāy əl-manṭiqa ʿarab.
UA1 [GCA]: *ʿarab, qabāyil ʿarabīya.*
IR1 [GCA]: *qabāyil ʿarabīya.*
UA1 [GCA]: *mā simaʿtum min yaʿnī ajdādkum əl-hōla walla šai?*
IR1 [GCA]: *lā, hāy kalimat əl-hōla hinā fī ha l-manṭiqa hinā mā simaʿnāhā wa lā ḥadd ḏikarhā. hinā qabāyil banī timīm, qabāyil əl-mawālīč, əl-manāṣīr, əl-ḥaram, la-sāmān əš-šarg əl-ḥammādīya, ʿibēdil, qabāyil yaʿnī, killhā qabāyil ʿarabīya.*

UA1: 'Do you have the term *Hōla* here?'
IR1: 'No, the term *Hōla* isn't present here. This coast is inhabited by Arab tribes. From Kangān and Banak[22], till you arrive to [Bandar] Linga, this coast and also the islands that flank this coast, Šēḫ [Šuʿaib] island [= Lāwān Island] as far as Ḫārg Island, which is to the north of Būšihr, opposite Gināwa, are still populated by Arab people. Tīn Island towards Gināwa. The coast, the original inhabitants of the coast are Arabs. But with the mingling that happened as time passed, there are types of workers who came, Persians who stayed here[23]. They are types of sheepherders, ironsmiths, I don't know, these things. Afterwards it increased, because of the companies and jobs, etc. people came to this region. But this region beforehand, hundred percent, I would say not more than fifty years ago, all inhabitants of this region were Arabs.'
UA1: 'Arabs, Arab tribes.'
IR1: 'Arab tribes.'
UA1: 'Didn't you hear, from your ancestors, about the *Hōla* or something like that?'
IR1: 'No, we didn't hear the word *Hōla* in this region, nor did anybody mention it. Here [we have] the Banī Tamīm tribe, the Mālikī tribe, the Manāṣīr, the Ḥaram, as far as the Eastern Territory[24] of the Ḥammādī tribe, the ʿAbdalī. Tribes, I mean, all of them Arab tribes.'

The video, produced in 2009, provides a clear affirmation that the Arabic language did not completely vanish from Iran's Persian Gulf shore. The question to pose is, to what extent does Arab culture still permeate the once flourishing Arab settlements? The most straightforward method to explore this question would be expansive fieldwork, but the topic's sensibility in Iran has hampered efforts to visit Arab villages and assemble ethnographic data as Lorimer did. Although this remains my long-term goal, linguistic

[22] Bandar Kangān, home of a major natural-gas processing facility, and the adjacent village of Banak lie 200 km south of Būšihr, 80 km north of ʿAsalūya. These settlements were under the rulership of the Āl Naṣūr tribe, of whom some remain there.
[23] 'Arabs of the Coast' call Persians *Furs* or *ʿAjam* [GCA *ʿAyam*], referring to "the name given in medieval Arabic literature to the non-Arabs of the Islamic empire, but applied especially to the Persians" (Bosworth 2011).
[24] 'Arabs of the Coast' divide their homestead into two territories, the Eastern Territory (*Sāmān Šarg*) and the Western Territory (*Sāmān Ġarb*) with the dividing line running at Pārsiyān. The Persian lexeme *sāmān* means 'city, country, territory' (Dihḫudā).

data can be directly collected from Arabs who have moved back to the Arabian Peninsula, and indirectly from their emerging online activity. The methodology of this research follows the procedures of Discourse-Centered Online Ethnography (DCOE), "a combination of systematic observation of online activities and interviews with online actors, which was developed as a complement to the linguistic analysis of log data[25]." The ethnographic framing of the linguistic analysis of online log data enables the researcher to contextualize the recorded data. Since 2012, I have been monitoring the social media activity of 'Arabs of the Coast', and have followed the progress of their auto-documentation. Through social networking, micro-blogging and video-sharing sites, 'Arabs of the Coast' have managed to reach out to other Arabs in the region, most notably their fellow tribesmen on the Arabian Peninsula[26]. We are witnessing a steadily growing movement that aims to substantiate the presence of Arab culture in Iran's two provinces on the Persian Gulf: Būšihr and Hurmuzgān. This endeavor was enhanced by the events of the Arab Spring, where the use of social media played a key role in organizing demonstration mobilizing people to share ideas, or to report events that may not have been adopted by the mainstream media. I am fully aware that online sources need to be handled cautiously as they can be politically or ethnically biased. Nevertheless, they provide an insight into how Iranian Arabs as an ethnic, linguistic, and religious minority preserve their culture. The interfaces I have studied are Facebook, Twitter, Instagram, YouTube, WhatsApp, Telegram in addition to community websites in Iran and the Arab Gulf states[27]. The timespan of my focus extends from the inception of their online activity in the mid-2000s until the present moment. Simultaneously, I recorded language data from both the resettled and visiting Iranian Arabs in the United Arab Emirates in 2011, 2013, 2015, 2016, Qatar in 2013 and Kuwait in 2016. The interviews span over 45 hours, and interviewees were selected based on the criteria that they self-identified as Arabs, and either themselves or their immediate family members (father, grandfather) were born in Iran. 40 individuals agreed to provide spoken language data in the age range of 19-75 years. The interviews were semi-structured, face-to-face, and centered around conversations on family, tribal history, village life as well as Arab culture and heritage on the Iranian coast. In some cases, follow-up questions on WhatsApp and lexical-semantic questionnaires were provided to the informants in person. For the recorded data I exclusively had access to men. The written data is taken from the public social media accounts of 46 individuals of both genders in the same age range. Five people overlap in the two data sets.

[25] Androutsopoulos 2008: 1.

[26] Lee 2011, 111.

[27] It should be noted that legal access to social media in Iran has been hampered by efforts of Iranian authorities since 2000 to block popular blogs and websites such as YouTube, Facebook, Twitter, and Google Plus (Srebnerny, Khiabany 2010: 70). However, many Iranians are still able to access these sites by changing the settings of proxy servers. Twitter was partially unblocked in late 2016, see https://www.iranhumanrights.org/2017/02/briefing-twitter-third-wave-in-iran/.

4. Classification of 'Arabs of the Coast'

The interviews I conducted as well as the information gleaned from online activity allow for the categorization of 'Arabs of the Coast' into four groups, based on three criteria: birthplace, current residence and citizenship.

Group 1. Arabs born in Iran, currently residing in Iran, and holders of Iranian citizenship. The younger generations of these Arabs are bilingual in Arabic and Persian, Persian being the language of education and administration. Classical Arabic, solely for religious purposes, is formally taught to children in local schools. Gulf Colloquial Arabic (GCA) is restricted to the household, and serves as a vehicular language among Arab residents. Persian is spoken by and with administrators and government officials appointed to Arab-dominant villages from Shiraz, Tehran or other parts of Iran. Older generations of Arabs without formal schooling have limited Persian knowledge, and rely on their children or grandchildren for assistance with interpretation in official settings.

Group 2. Arabs born in Iran, currently residing in Arab Gulf states, and holders of Iranian citizenship[28]. These individuals tend to be bilingual, but their proficiency in Persian varies according to the number of years they spent in Iran, and at what age they moved to Arabia. Their children will likely only be Arabic speakers, yet unable to obtain Gulf citizenship if the father moved from Iran after the window for citizenship registration closed[29]. A subgroup sharing the three criteria should be added under this heading for those whose ancestors were originally Persians from the Zagros mountains in Iran, but entered the patronage of Arab tribes in coastal settlements, and were subsequently Arabized.

Group 3. Arabs born in Iran, currently residing in Arab Gulf states, and holders of Gulf citizenship. This group comprises individuals who moved to the Arabian coast in the early days of the newly established Arab Gulf states when rapid population growth was in the interest of the ruling families. They were mainly Arab tribal leaders who, along with their immediate families, sought to reunite with the tribal elite in Arabia. They may be bilingual based on the age they left Iran. Their children hold Gulf citizenship and are Arabic speakers.

Group 4. Arabs born in the Arab Gulf states, currently residing there, and holders of Iranian citizenship. This group encompasses the children of people in Group 2. They were raised and educated as Arabs, are almost entirely Arabic-speaking, but live with the local stigma of being Iranian citizens. When women from this group marry Gulf citizens, their children automatically inherit Gulf citizenship.

A potential **Group 5** could be dedicated to descendants of people in Group 3. These individuals, as Gulf citizens educated in Arabic, perceive the presence of their ancestors

[28] Many Arabs in this group let their Iranian documents expire. They become stateless people (*Bidūn*), but hope to turn eligible for Gulf citizenship by purchasing citizenship from the Union of the Comoros.

[29] With the exception of Kuwait, Gulf citizenship is passed down from father to child. If a woman from the Iranian coast marries a Gulf citizen, their children will be Gulf citizens.

on Iranian soil as a matter of mere historical curiosity. When Arabs from Group 1 continue to emigrate to the Arab Gulf states, they will belong to Group 2 without much hope of acquiring Gulf citizenship (unless they are women married to Gulf citizens). Their children will join Group 4. People from Groups 1, 2 and 4 will be represented in conversation threads and recorded materials in this study[30].

5. Language and Identity among 'Arabs of the Coast'

5.1. Village Life in Iran

The video-sharing website YouTube has become one of the most powerful tools to stream information to the world without having to handle more than a cell phone and its built-in camera. An 'Arab of the Coast' in Qatar launched a channel in 2008 to share his videos of friends and relatives reciting Arabic poetry, telling stories of local villages, and explaining their heritage by showing traditional occupations such as the *kawwāḥ*, the person who builds the *kōḫa* 'hut' to hunt falcons[31]. The channel has 27 videos, 180 subscribers and over 2,840,000 views[32]. Additionally, the videos show a wide range of cultural elements tied to the Iranian coast, but equally relevant in the Arab Gulf states: bagpipe (*habbān*) music from Kūškinār[33], weaving of the *ḥaṣīr* 'matt', types of fishing nets (*ġazūl* for larger fish, *digga* for smaller fish) and the fishing basket (*gargūr*). Five videos were recorded in the village of Hāla (Arabic: al-Ḥālla) on occasion of a mosque opening, which was funded by a member of the Āl Bū Sumaiṭ tribe of Qatar[34]. On a separate channel, the same Qatari resident of Iranian Arab origin posted videos of *nabaṭī* poetry and a video with an elder from the abandoned village of Nāyband (Arabic: Nāband)[35]. The elder (IR2) recounted the story of Sheikh Muḥammad al-Ḥaramī and his tribe's

[30] In an upcoming article, I extensively discuss Arabic-Persian codeswitching among Arabs from Group 3 (Gazsi 2016).
[31] The YouTube channel is available on this link: http://www.youtube.com/user/azoz221
[32] The numbers here and in subsequent instances represent viewership in April 2017.
[33] Kūškinār is a small town located in the Pārsiyān sub-province of Hurmuzgān Province, 23 km southwest of Pārsiyān city, the sub-provincial capital. During the 2006-2007 census (*Saršumārī*), Kūškinār had a population of 2,902 people in 663 households, and the surrounding district has been the stronghold of the Āl Ḥaram tribe (Daryāʾī 2012: 126).
[34] Hāla is a tiny village in the Čāh Mubārak district in Būšihr province, 20 km southeast of ʿAsalūya. The village had strong ties to the State of Qatar with the continued presence of the Āl Bū Sumaiṭ and Banī Tamīm tribes. The ruling family of Qatar, the Āl Ṯānī belong to the latter. The name was replicated from a village on the far south side of Muharraq in Bahrain, that used to be a separate village until 1949 CE, see link:
http://www.alayam.com/Article/alayam-article/88845/تاريخ-كبير-..«الحالة»-لمنطقة.html (published on May 29, 2014).
[35] Nāyband lied in the Čāh Mubārak district, 2 km west of Hāla, and served as the entry point to

movement along the road leading from Sūrbāš to Nāband and ʿAsalūya through the villages of Ḫiyārō, Aḫand and Dihnau[36]. The village elder engages in intra-sentential codeswitching between MSA and GCA in the video, of which a short excerpt is included here[37]:

> IR2 [GCA]: *wa huwa intaqal ilā ḥāllat nāband wa sakan fī nāband wa ʿammar nāband. kānat nāband ḫarba, nāband əl-qadīma ġēr nāband əl-jidīda illī hū kān ʿāš fīhā. fa-nazal fī nāband wa banā la galʿa šamālī ʿan nāband əl-jidīda. kān fī jiwāra aḥmad bin ḥaidar wa ibrāhīm muḥammad əs-saʿīd wa aḥmad zabārtī illī hum byikūn āl ḥamad illī hum aṣəlhum nāṣir aḥmad yābir u haḏōl āl ḥamad əl-maujūdīn. wa jiwāra əṯ-ṯānī mullā yūsif wa ḫālid, mullā ḫālid wa mullā jumʿa illī kān ḫaṭīb fī masyid əl-jāmiʿ illī banāh šēḫ muḥammad ...*
>
> 'He moved to Ḥālla of Nāband, and lived in Nāband and populated Nāband. Nāband was dilapidated: Old Nāband, not New Nāband where he used to live. So he came to Nāband and built a fort in it, north of New Nāband. On his side was Aḥmad bin Ḥaidar and Ibrāhīm Muḥammad al-Saʿīd and Aḥmad Zabārtī, who belong to the Āl Ḥamad tribe, whose ancestor is Nāṣir Aḥmad Jābir. These were present from the Āl Ḥamad tribe. On his other side was Mullā Yūsif and Ḫālid, Mullā Ḫālid and Mullā Jumʿa, who was the preacher at the Friday Mosque built by Sheikh Muḥammad ...'

The engine of the online community for 'Arabs of the Coast', the driving force and most active member of their online activity is a young man from Basātīn (UA1)[38]. He is a Sunni Muslim, and belongs to the Naṣūrī sub-branch of the Jabūr branch of the Banī Ḫālid tribe[39]. UA1 completed his elementary and secondary education in Čāh Mubārak, and relocated to the United Arab Emirates in 2011[40]. His Iranian citizenship notwithstanding, he set up residence in al-Mirfaʾ, 160 km west of Abu Dhabi in the

---

the Iranian coast for Arab tribes arriving from Bahrain and Qatar. It was abandoned during the mid-20th century due to drought and the subsequent plague outbreak. Its surviving residents moved to Hāla or other neighboring villages, while some made their way to Al-Faw in Iraq. From there they moved to Kuwait, where a family with the name al-Nābandī still resides. Nāyband area news are available on this website: http://nayband.ir/. Link to the channel: https://www.youtube.com/channel/UCrsBVhfppPzM-2W06XNW0QA

[36] Sūrbāš, ʿAsalūya, Aḫand and Dihnau have been home to the Āl Ḥaram tribe (Daryāʾī 2012: 96).

[37] Link to the video: https://www.youtube.com/watch?v=JOXvSUZP2Y8

[38] Basātīn village is located in Čāh Mubārak district, ʿAsalūya sub-province in Būšihr province. It lies 3 km east of Hāla, and had a population of 638 living in 108 households during the 2006-2007 census (*Saršumārī*).

[39] For more information on the tribe, see Daryāʾī 2012: 19-24. Al-Anṣārī 2011: 66.

[40] Čāh Mubārak is located 16 km east of Basātīn. It had a population of 2,343 during the 2006-2007 census (*Saršumārī*), and has been the homestead of the Banī Tamīm tribe. The website of the District Council (Persian *baḫšdārī*): http://bakhshdarichm.ir/, Village Council (*dihyārī*): http://dehyarichm.ir/, and the city blog: http://chcity.blogfa.com/.

Western Region Municipality. Al-Mirfaʾ had been a fishing and pearling village with direct connections to the Iranian side of the Persian Gulf before Sheikh Zayed, the founder of the UAE, implemented extensive development projects from the 1970s. His goal with al-Mirfaʾ was to house 'Arabs of the Coast' arriving in multiple waves to the UAE. I interviewed UA1 eleven times between 2013 and 2016 in al-Mirfaʾ, Abu Dhabi and Dubai, during which he proudly revealed that he had been raised in a completely Arabic-speaking environment. His father knows some Persian, while his mother allegedly does not speak a "single word" of Persian, and needs an interpreter when interacting with Persian government officials. UA1 is bilingual in Arabic and Persian. Although he speaks Persian fluently and considers it his second language, he speaks it with a distinct Arabic accent. What follows is an extract from his reminiscences of daily life in Basātīn, as recounted to me in 2016 in GCA. The text includes several locally used Persian lexemes (*čāmī, kiškī, kara*) that he explains in Arabic, and Arabic technical terms (*rōb, gabgūb*), for which he provides a Persian translation. He talks about daily morning chores, meals they prepare and evening activities with friends at the seaside.

UA1: [GCA - MSA] *lākin qura miṭil basātīn əl-ḥīn baʿṭīk ʿala sabīl miṭāl yaum ana kint əhnāk kēf itkūn əl-ḥayā l-yaumīya, ḥayāt əl-yaumīya, agʿad əṣ-ṣubḥ badrī əs-sāʿa sitt yaʿnī sitt inta gāʿid min əṣ-ṣubḥ našīṭ u gaʿd badrī bitgūm bitrūḥ əl-mazraʿa maṭalan. illī ʿanda šuġl yirūḥ əš-šuġl. u ana maṭalan əṣ-ṣubḥ agʿad atrayyag ruyūgī. šū yikūn ruyūgī? yikūn ašyā freš yaʿnī maṭalan əl-wālda illī ummī msawīya čāmī au kiškī yisammūna šai abyaḍ min əl-bugar yaʿnī taḥlab əl-bugar, ḥalīb əl-baqar, u ʿugub ḥalīb əl-baqar ysawwī lak laban, rōb illī hū māst, ysawwī lak kara illī hū zibda bi l-ʿarabīya. u kaḏālik anwāʿ əl-aklāt əl-labanīyāt hāḏa ʿandna. yisawwī lak ər-ruyūq titrayyaq, u ḫabəz yiḫabbəz fi t-tannūr ...*

'But the villages such as Basātīn now, to give you an example, when I was there, how was daily life? The daily life [was], I get up early in the morning, at six o'clock. I mean from six o'clock in the morning you are active. You get going early to the farm for example. Whoever has a job goes to work. But, for example, in the morning, I eat my breakfast [GCA *atrayyag ruyūgī*]. What is my breakfast? *Fresh* [English word] things, for example my mother, who is my mother [UA1 used a synonym], prepares *čāmī* or *kiškī*, a white thing from the cow [GCA *bugar*], I mean you milk the cow [GCA *bugar*], cow milk [MSA *baqar*]. And after cow milk [MSA *baqar*] she prepares you dairy, yogurt which is *māst* [MSP], and she makes butter [MSP *kara*] for you, which is *zibda* in Arabic. Similarly, these are the types of dairy we have. She prepares the breakfast for you to eat [MSA pronunciation of GCA lexemes: *ruyūq titrayyaq*], and bakes bread in the oven ...'

UA1 [GCA - MSA]: *ana hnāk ḥayātī gabl mā ayī, arūḥ ʿala ṭūl lailīyan arūḥ wain? əl-baḥr, agūl muġarb ʿugub ṣalāt əl-ʿišā. mā nitʿašša fi l-bēt əngūm maʿa š-šabāb illī hū grūp mālna au šabāb rifjānna aṣdiqāna nirūḥ ʿal-baḥr nigʿad niḥaṭṭ farša.* [...] *wa nidḫil niṣīd əs-samak mil-baḥr wa nṣīd əl-gabgūb illī hū ḫarčang, š-isammūna, ḫarčang yisammūna əl-gabgūb əl-ʿayam. nṣīd əl-gabgūb u nṣīd ən-naġar ...*

'My life there before I came, every night I would go straight to, I would go where? To the sea. I'd say at sunset, [or] after the evening prayer. We don't eat supper at home, we set out with the guys who are our *group* [English word] or they are guys [who are] our pals

> [GCA *rifjānna*], our friends [MSA *aṣdiqāna*]. We go to the sea, we sit down, put down a mat. [...] We enter [the water] to catch fish from the sea, and we catch crab, which is *ḫarčang* [MSP]. What do they call it? The Persians [GCA *ʿayam*] call the crab *ḫarčang*. We catch crab and cuttlefish ...'

A recurring topic I heard from 'Arabs of the Coast' in the UAE and Qatar is their yearning for the unadulterated joy they felt leading a simple life in the villages surrounded by pristine nature. Even though they chose to leave Iran for feeling marginalized as Arabs, the prosperous but highly urbanized life and arid landscape in the Arab Gulf states make them feel isolated and somewhat alienated. UA1 expressed similar sentiments in a 2016 interview with me.

> UA1 [GCA]: *fa l-ḥayā əhnāk jiddan mumtaʿ wa l-ḥayā r-rīfīya jamīla. jamīla ʿašān čī ana mištāg ilhā? ʿašān čī mištāg ila t-turāṯ? ʿašān čī mištāg ila l-ḥayā l-awalīya maʿa anna t-taṭawwur killa illī ḥaṣal? lākin yaumīyan aštāg bi-šakl faḍīʿ ila l-ḥayā illī kint fīhā. Ḥayā ṭabīʿīya jiddan jiddan jiddan jiddan jiddan. mā fī ayy šai taṣannaʿ, mā fī ayy šai, mā fī ayy šai min əl-, yaʿnī, ṭabīʿa ʿalā ṭabīʿa ʿalā ṭabīʿa min kill jawānib ...*
>
> 'So life is very pleasant there, and rural life is beautiful. Beautiful, why do I long for it? Why do I long for the heritage? Why do I long for the primitive life despite all the development that happened? Yet every day I long terribly for the life I had. [It was] a very, very, very, very, very natural life. There is nothing artificial in it, there is nothing, there is nothing of, I mean, nature on nature on nature from all sides ...'

5.2. Outreach on Social Media

UA1 is the founder and administrator of the most comprehensive website dedicated to the 'Arabs of the Coast', titled *Muntadayāt ʿArab al-Sāḥil, mauqaʿ ʿArab al-Juzur wa al-Sāḥil al-Šarqī li al-Ḫalīj al-rasmī* 'Forums of the Arabs of the Coast, the official site of the Arabs of the Islands and East Coast of the Gulf'.[41] The site has over 43,000 posts from 10,555 members in more than 3,500 topics, ranging from 'Our Lands and Villages' (*Diyārnā wa Qurānā*), 'Our Sheikhs and Tribes' (*Šuyūḫnā wa Qabāʾilnā*), 'Our Braveries and Glories' (*Buṭūlātnā wa Amjādnā*), 'Our Heritage is the Gate to our Culture' (*Turāṯnā Bawwābat Ḥaḍāratnā*) to 'Marine History' (*al-Tārīḫ al-Baḥrī*). Each topic has dozens of forums where participants record their own heritage with stories and photos, and gather all relevant information they can access from academic and non-academic sources. Examples of forums under the topic 'Our Lands and Villages' include 'Meaning of Village Names of the Arabs of the East Coast' and 'Village of Banak'. Forums under the topic 'Our Sheikhs

[41] The website is available at: http://arabalsahel.net/vb/.

and Tribes' incorporate 'al-ʿAbādila (MSA *ʿUbaidil*, GCA *ʿIbēdil*) in Light of Historical Sources', 'Āl Naṣūr between Rulership of the Past and Marginalization of the Present'. Forums under the topic 'Our Heritage is the Gate to our Culture' are 'Old Terminology of the Arabs of the Coast' and 'Weapons of the Arab Sheikhs of the Coast'. UA1 administered a Facebook page between October 2010 and August 2016 that served as a patchwork of information on 'Arabs of the Coast'. The profile has over 3,300 friends, primarily Arabs from Iran and the Arab Gulf states, with or without relations to the opposite shore. Posts and comments in the conversation threads are exclusively in Arabic (MSA and GCA), and discuss the rich heritage of local people and former Arab tribal chiefs from the villages by posting pictures, poems, videos and reports on local events. This resulted in an often randomly organized amalgamation of materials, where photos from 50 years ago are posted next to pictures on the renovation of a traditional Arab home on Kīš Island. UA1 also endeavored to provide historical documents proving the existence of Arabs on the Iranian coast. Moving to the UAE offered him a boost in expanding his mission to document Arab heritage with less restricted access to the internet than in Iran. Another valuable project to represent the traditions of his homeland is his YouTube channel titled 'Arabs of the Islands and East Coast of the Arabian Gulf', with a description that overtly intends to showcase 'the forgotten Gulf population on the islands and coast opposite the Gulf'[42]. Established in 2008 and active until 2013, the channel had over 1,500 subscribers and 1,300,000 views of 113 videos. The videos feature Arab cultural events, dances, marriages, poetry recitals, historical documentaries (both professional and personally created), television discussions and Arabic folk tales from Iran. The channel integrates video footage of street scenes in Arab settlements, local Arabs taking a tour around their villages and talking about the history of buildings, forts or palm tree plantations[43].

UA1's recent large-scale initiative is a daily newsletter published since 2014 on WhatsApp and Telegram, two instant messaging applications for smartphones. As messages on both platforms are encrypted, the newsletters avert the watchful eyes of Iranian authorities who dislike the association of their side of the Persian Gulf coast with Arabs. The newsletter with the heading 'Newscast of the House of Knowledge for the Regions of the Arabs of the East Coast' (*Našrat ʿUlūm al-Dār li-Manāṭiq ʿArab al-Sāḥil al-Šarqī*) is published via the 'News Center of the Council of Arabs of the Coast' (*Markaz Mailis ʿArab al-Sāḥil al-Iḫbārī*)[44]. The newsletter consists of the following sections: 1. News reports on local events from villages and cities; 2. Accidents and deaths; 3. Historical accounts titled 'Emirates of the Arabs of the Coast'; 4. Religious section on Sunni Islam titled 'Islam and Life'. Seasonal sections include 'Book of the Week', sports news, environmental issues, reports on Islamic festivities in villages, weddings, dialectal section on Arabs terms used on the East Coast of the Persian Gulf. The daily news updates

[42] The channel is accessible at: http://www.youtube.com/user/fareso0o
[43] UA1 directed most of the video clips on the channel.
[44] MSA *majlis* > GCA *mailis* 'council'. The word is spelt with a *yāʾ* in the newsletter's heading.

UA1 incorporates into the newsletter come from his vast network of local Arabs in virtually every village on the Iranian shore between Banak in the north and Bandar Linga in the south. The newsletter is associated with several Instagram sites, where photos of the news reports are shared and discussed by participating members[45]. From the several mottos the newsletter has had over the years, the one that best captures the essence of the project reads: 'Our clothing .. Our customs = Our identity' (*Libāsnā .. ʿĀdātnā = Huwīyatnā*) referring to the predilection of 'Arabs of the Coast' for wearing the long white robe (*ṭaub* or *kandūra*), white or red headdress (*ġitra*), headband (*ʿigāl*) for men, and face mask (*burgaʿ*) for women.

Twitter is a modestly successful platform to publicize Arab culture on the Iranian shore as it has been fully or partially blocked in the Islamic Republic in recent years. 'Arabs of the Coast' now residing in Arab Gulf states, however, have used the social networking service to raise awareness among their tribes about fellow tribesmen on the opposite coast. One Twitter account titled 'The East Coasts of the Gulf' (*Sawāḥil al-Ḫalīj al-Šarqīya*) was active between March 2012 and November 2015, had almost 19,000 followers and approximately 13,000 tweets about local customs, games, traditional food, clothing, flora and fauna[46]. Examples of more localized accounts and channels have also been established, where the creators focused exclusively on their and surrounding settlements. Kīš Island is represented on Twitter with over 6,000 tweets and 13,700 followers by a local Arab from the Āl ʿAlī tribe now living in the United Arab Emirates[47]. The villages of Bandar Kung and Lāft on Qišm Island are shown on a YouTube channel dedicated to the region around Bandar Linga, formerly under the rulership of the Qawāsim[48]. Street scenes on Eid al-Adha and other religious events from Chāh Mubārak are available on a YouTube channel, where an example of video title is 'Eid al-Adha in the village of the Banī Tamīm, Jahəmbārak, Arab tribes on the coast of Persia, Arabs of the East Coast and Islands'[49]. A series of wedding celebrations for the 'Arabs of Persia' from Dihnau village has been recorded on a separate YouTube channel[50].

[45] Accompanying Instagram sites are as follows: 1. https://www.instagram.com/explore/tags/عرب_الساحل/ (active since November 2012, over 6,000 posts); 2. http://instagram.com/arabalsahel/ (established in November 2015, over 1,100 posts and 13,000 followers); 3. https://www.instagram.com/oloomaldar.news/ (active since August 2016, over 900 posts and 6,200 followers).

[46] The account is still available online: http://twitter.com/arabKhaleej.

[47] The link to his account is: https://twitter.com/jazerat_Qais.

[48] The Qāsimī tribe has ruled various areas on both coastlines since the 18th century CE, and currently rule the Emirate of Ras al-Khaimah and the Emirate of Sharjah within the UAE. The channel is accessible at: https://www.youtube.com/channel/UC5zWC2iLPq-nhA_FnfADtlw.

[49] The YouTube channel is available on this link: http://www.youtube.com/user/alwsem03.

[50] The channel can be reached on this link: https://www.youtube.com/user/dhno23.

5.3. Expressions of Language and Identity

Discussions on social media may evolve into heated exchanges over language and identity, as witnessed in the following Facebook post. It opens with a photo showing elderly men from the village of Naḫl Jamāl[51]. The conversation thread began in June 2010, and ended in September 2011. The participants are from Naḫl Jamāl, Bandar Linga, Bandar ʿAbbās and Basātīn. Their family names speak volumes about their identity as they write in GCA, yet have Persian names: *Malikzād* 'king's offspring, prince', *Aḥmadzāda* 'Aḥmad's offspring', *Ātašdāman* 'fire-fanner', *Ḫadang* 'poplar tree'. Following the European model, last names were officially given to Iranian citizens during the Pahlavi era. This measure took Iranian Arabs by surprise, due to the fact that the new names were chosen randomly based on the individual's occupation or personal skills. Simultaneously, names representing tribal affiliation were erased.

IR3: واللـه بهالصـوره دكرتنـــي بيدي - اللـــه يرحمه -
nice pic bro
[...]
IR4: هذين الشياب من قرية نخل جمال الله يحفظهم
UA1: شكرا لكم جميعا
[...]
IR5: صور جميله /الله يحفظهم / بارک الله فيک اخوثی [name] على الاهتمام
IR6: الجالس فی شمال الصوره هو خالی
IR6: الله يحفظه
IR7: شو اسم خالک ؟
[...]
IR8: شوتبون من الماضي غير القحط والتعب ماسمعنامنهم شي شوفوا الحاظر
UA1: اللي ماله اول ماله تالي ،،،،صمودهم و صبرهم ايام القحط و بطولاتهم و تاريخهم و امجادهم العربيه في طرد المحتل و تشكيل امارات عربيه على ضفاف الساحل الشرقي للخليج العربي يستحق التدوين و التوثيق و الحاضر بيكون اجمل يوم تدخله من بوابة الماضي ،،، تحياتي و احترامي لك ...

IR3 [GCA]: *wallāhi bi-ha ṣ-ṣūra ḏakkartnī bi-yaddī – allāh yarḥama. nice pic bro.*
[...]
IR4 [GCA]: *hāḏain əš-šiyāb min qaryat naḫl jamāl allāh yaḥfaẓhum.*
UA1 [MSA]: *šukran lakum jamīʿan.*
[...]
IR5 [MSA]: *ṣuwar jamīla / allāh yaḥfaẓhum / bārak allāhu fīk aḫūṯī [= aḫūyī]* [name] *ʿalā l-ihtimām.*
IR6 [GCA]: *əl-jālis fī šimāl əṣ-ṣūra hū ḫālī.*
IR6 [GCA]: *allāh yaḥfaẓa.*
IR7 [GCA]: *šū ism ḫālik?*

[51] Naḫl Jamāl has 105 households with 461 residents, predominantly Sunni Arabs from the Banī Ḥammād tribe. They originated from the Khor Al-Udaid region in southeast Qatar (*Saršumārī* 2011-2012). The village is located in the Bandar Muqām sub-district, Šībkūh district, Bandar Linga sub-province of Hurmuzgān province. Social media presence of the village is a Telegram account (http://instidy.com/nakhl.jamal), a YouTube channel (https://www.youtube.com/user/albokhlf), an Instagram account (https://www.instagram.com/albnoudi/), and a photoblog and informational page (http://www.albnoudi.loxblog.com/).

[...]
IR8 [GCA]: *šū tabbōn min əl-māḍī ġēr əl-qaḥṭ wa t-taʿb mā simaʿnā minhum šai šūfū l-ḥāẓir.*
UA1 [GCA – MSA]: *illī mā la awwal mā la tālī ,,,, ṣumūdhum wa ṣabrhum ayyām əl-qaḥṭ wa buṭūlāthum wa tārīḫhum wa amjādhum əl-ʿarabīya fī ṭard əl-muḥtall wa taškīl imārāt ʿarabīya ʿalā ḍifāf əs-sāḥil əš-šarqī li l-ḫalīj əl-ʿarabī yastaḥiqq ət-tadwīn wa t-tauṯīq wa l-ḥāḍir byikūn ajmal yōm tadḫula min bawwābat əl-māḍī ,,, taḥīyātī wa ḥtirāmī ...*

IR3: By God, with this picture you reminded me of my grandfather – may God have mercy upon him. nice pic bro.
[...]
IR4: These old men are from the village of Naḫl Jamāl, may God protect them.
UA1: Thank you all.
[...]
IR5: Nice pictures /may God protect them/, may God bless you, my brother Bū Ḫālid, for your care.
IR6: The [person] sitting on the left of the picture is my uncle.
IR6: May God protect him.
IR7: What's your uncle's name?
[...]
IR8: What do you want from the past, other than drought and hardship? We didn't hear anything [else] from them. Look to the present.
UA1: Who doesn't have a past, doesn't have a future,,,, Their defiance, their steadfastness during the time of drought. Their bravery, their history and their Arab splendor in the expulsion of the occupiers [= Persians], and in forming Arab emirates on the banks of the East Coast of the Arabian Gulf, deserves recording and documentation. The present will be nicer when it enters through the gate of the past,,, My greetings and respect ...

The conversation revolves around a topic that frequently surfaced during my interviews with Iranian Arabs as well as in social media threads. It highlights the yearning for the past glory the ancestors of 'Arabs of the Coast' are said to have enjoyed. They constantly question whether they should reminisce about the olden days or look to the future instead. The past reminds them of the harsh living conditions on the Iranian coast their parents and grandparents had grown up in, and rekindles the sights of underdeveloped villages with no running water, no asphalt roads, no schools, no medical facilities. Yet this was also the time when their Arab culture was unblemished, and connections with fellow tribesmen on the Arabian side of the Persian Gulf were flourishing. The last quarter of the 20th century has seen official efforts to uproot their traditions and integrate them into the modern, revolutionary Shiite Persian state, but it also brought far-reaching development projects to the region. The counterargument is that the Arab heritage should be remembered and nourished, and not let it become diluted among the surrounding Persian population who may frown upon the display of Arab identity in this politically sensitive region. It also signals an effort to make Arab voices heard, at a time when Shiite mosques and *ḥusainīyas* are being built in villages to counterbalance the religious view of local Sunni Arabs[52].

[52] An illustrative example of the expanding Shiite influence is the naming of the main mosque

A similar dialogue unfolds in another Facebook post among members of the Banī Ḥammād tribe wo have a strong presence on social media. The Banī Ḥammād, a Qaḥṭānī tribe, had moved to the Iranian coastline from present-day Qatar, and established residence in Bandar Muqām (pronounced *al-Mugām* or *al-Magām* in the local GCA[53]). They spread to the neighboring villages Naḫīlū, Čīrūya, Kalāt, Rastāq, Murbāġ and Lāwān Island[54]. The last sheikh of the Banī Ḥammād moved with his kins to the United Arab Emirates in the 1970s, where they grew so close to the ruling Āl Nahyān family that his sons were appointed to the highest levels of the Emirati government, including the incumbent Minister of Education. Two Facebook profiles are partially dedicated to Bandar Muqām, one active between August 2010 and April 2016[55], the other active between April 2010 and June 2014[56]. An Arab from Bandar Muqām, now living in Dubai, created the profiles. The latter profile had 152 friends, and was a hodgepodge of personal reflections on world events in Arabic and Persian, interspersed with old and recent photos of Bandar Muqām and its inhabitants. A YouTube channel with four videos has been dedicated to the settlements formerly under the control of Banī Ḥammād sheikhs[57]. One video caption expresses the disenfranchised grief of Arabs for losing their former glory on the Iranian coast: 'Territories and villages of Banī Ḥammād's rule on the East Coast of the Arabian Gulf before the odious Persian occupation (*al-iḥtilāl al-fārisī al-baġīḍ*) of Arab territories'.

The discussion below from Facebook was initiated by UA2, a UAE resident born and raised in Bandar Muqām. He posted a photo of the village with the caption 'From the City of Dreams al-Muqām', and three people joined the conversation to offer their thoughts: an Emirati citizen with the last name Al-Ḥammādī (UA3), an Arab from Bandar Linga working in Bandar Muqām (IR9), and one from Bandar Muqām residing in Bandar ʿAbbās (IR10). The conversation is conducted in GCA. The Al-Ḥammādī from the UAE expresses his amusement at his fellow tribesmen's longing for and loyalty to the village they hail from, despite the welfare and social services they receive in the UAE. UA2 explains that they must cultivate this connection and emphasize the Arab character of the village, since outsiders – *ʿAyam* 'Persians' – are trying to infiltrate the region by imposing their cultural and religious perceptions on local Arabs.

in Hingām-i Jadīd on Hingām Island after Imām Riḍā, the eighth Imām. The mosque in the almost completely abandoned village of Qīl, however, holds the name ʿUmar Fārūq, named after the second caliph ʿUmar ibn Al-Ḫaṭṭāb. He is viewed negatively in Shiite literature as a usurper of ʿAlī's right to the Caliphate.

[53] Al-Anṣārī 2011: 65. Bandar Muqām had a population of 7,640 people in 1,679 households during the 2011-2012 (*Saršumārī*) census, chiefly Sunni Muslims.

[54] Al-Anṣārī 2014: 236-237, 239, 287.

[55] Accessible at the following link: https://www.facebook.com/moqam.

[56] Available at: http://www.facebook.com/bandar.moqam.

[57] The channel can be reached on this link: https://www.youtube.com/channel/UCwF0em RVRLSL8lYd_GCqsSw. More videos from Bandar Muqām can be found on the previously mentioned YouTube channel 'Arabs of the Islands and East Coast of the Arabian Gulf'.

UA2: من مدينة الاحلام المقام
UA3: ياخي انتوا عايشين هنيه وخيركم كله من هنيه .. وللحين عندكم ولاء لهناك
UA2: علي الاقل هاي منطقة عربية من مناطق بني حماد مب مثل بعض الناس ايحطون صور غريبة وماندري عيم يا شيعة وينسبونها لينا مشكلتك يا [name] انت حمادي بس مولود أهني اللة يهديك بس
UA2: لا يعجبك العجب ولا الصوم في رجب
IR9: داری
IR10: داری ولا ارضی بدار غیر داری...

UA2 [MSA]: *min madīnat al-aḥlām al-muqām.*
UA3 [GCA]: *yā aḫī intu ʿāyšīn hnīya wa ḫērkum killa min hnīya .. wa li l-ḥīn ʿandkum wilāʾ li-hināk.*
UA2 [GCA]: *ʿalā l-aqall hāy manṭiqa ʿarabīya min manāṭiq banī ḥammād mub miṭl baʿḍ ən-nās iḥaṭṭōn ṣuwar ġarība wa mā nidri ʿayam yā šīʿa wa yinsibōnhā līnā muškiltik yā* [name] *inta ḥammādī bass maulūd ihnī allāh yihdīk bass.*
UA2 [GCA]: *lā yiʿjibik əl-ʾajab wa lā əṣ-ṣōm fī rajab.*
IR9 [MSA]: *dārī.*
IR10 [MSA]: *dārī wa lā arḍī bi-dār ġēr dārī.*

UA2: 'From the City of Dreams, al-Muqām.'
UA3: 'My brother, you (pl.) live here and your welfare is all from here .. but until now you still have loyalty to there.'
UA2: 'At least this is an Arab region, from the regions of the Banī Ḥammād tribe, unlike the way some people put strange photos [= on Facebook]. We don't know [whether they are] Persians or [Persian word *yā*] Shiites, and they attribute those photos to us. Your problem, [name], is that you are a Ḥammādī, but you were born here [= the UAE]. May God guide you.'
UA2: 'A surprise should not surprise you, neither should fasting in the month of Rajab.'
IR9: 'My home.'
IR10: 'My home. I have no ground in a home that's not my home.'

An intriguing element in the identity of 'Arabs of the Coast' is how their history intertwined with the Persian population surrounding them, and how the two ethnicities intermingled. An illustrative example of the dual identity and background is a former Iranian citizen, now living as a *Bidūn* 'stateless person' in the UAE (UA4), whom I interviewed twice in 2013 in Sharjah and Ajman. UA4 was born and raised in Bandar Muqām before relocating to Sharjah in the mid-2000s. His family had originated from the city of Ḫunj, a Persian-speaking city in the Zagros mountain ranges, 250 km north of Bandar Muqām. The city lies in the historical Lāristān region where, together with the cities of Iwaz, Girāš and Lār, the population is largely Sunni Persian. His Persian ancestors descended to the coastal plains of Bandar Muqām in the first half of the 20th century, and blended in with the local Arabs. The Banī Ḥammād adopted them into their tribe, and they were gradually Arabized, both linguistically and culturally. UA4 considers himself an Arab and no longer a Persian, but having completed primary and secondary education in Persian as well as Arabic and Islamic studies in the holy city of Qom, he is a bilingual speaker of both languages. When speaking, he transitions between Arabic and Persian effortlessly with no foreign-sounding Persian accent, and repeats the same sentence in both languages frequently. This phenomenon is attested in his posts and comments on Facebook. The following two posts bemoan the identity crisis of 'Arabs of

the Coast', a crisis I observed on many occasions during my interviews. The first ascribes the entire Persian Gulf region to Arabs, and places Persians within the mountains overlooking the waterways, while the second laments the hardships of assimilation on either side of the Persian Gulf.

العروبه والبحر والخليج هو لعرب فارس اهل الجزر والسواحل اما بهرام وشهرام ليس لهم الا پشت كوه اى وراء الجبل

UA4 [MSA]: *al-ʿurūba wa al-baḥr wa al-ḫalīj huwa li-ʿarab fāris ahl al-juzur wa as-sawāḥil ammā bahrām wa šahrām laysa lahum illā pušt-i kūh ai warāʾ al-jabal.*
'Arabdom and the sea and the Gulf belongs to the Arabs of Persia, people of the islands and coasts. Bahrām and Šahrām do not belong to them, only [to those] *Pušt-i Kūh* meaning "behind the mountain".'

القهر اللي في قلبي ان لبسنا ولبس عجايزنا في بر فارس نفس اللي عند عرب الخليج بس يقولون لنا عجم واللي اكثر يقهرني انا لما تروح عن. العجم يقولون لنا أنتم عرب

UA4 [GCA - MSA] *əl-qahr illi fi qalbī ann libisnā wa libs ʿajāyiznā fi barr fāris nafs illi ʿand ʿarab əl-ḫalīj bass yigūlōn lanā ʿajam wa-lli akṯar yaqharnī ana lamma trūḥ ʿan əl-ʿajam yigūlōn lanā ʾantum ʿarab.*
'The oppression in my heart is that our clothing and the clothes of our elders on the Persian Land is the same as with the Arabs of the Gulf, but they tell us "you are Persians". What subdues me even more is when you go to the Persians, they tell us "you are Arabs".'

5.4. The Stigma of Citizenship

The stigma of being an Iranian citizen is a serious emotional trauma for children of 'Arabs of the Coast' who were born in Arab Gulf states, raised and educated as Arabs, but are not entitled to hold Gulf citizenship. Although these people, members of Group 4, are exclusively Arabic-speaking and Sunni Muslims, their passport represents a country to which they have no cultural connections. Iranian citizenship prevents them from benefiting from the wealth of the countries they call 'home', and hinders their free movement around the world. From this group, I interviewed QA1 in Doha at three different occasions in 2013. He is in his late 20s, and has lived in Qatar all his life. His father belongs to the al-Mālikī tribe, and originates from Tunbū village. He had immigrated to Qatar before his children were born[58]. The complex web of intertribal marital relations is attested in the somewhat confusing account of his family background:

QA1 [GCA]: *ummī min kiškinār baʿd. [...] ana jaddī māḫiḏ arbaʿa. ʿandī yaʿnī ʿandī ḫāl maṯalan fī bistānō wa ʿandī umm maṯalan waḥda bass fī kiškinār wa l-bāgī fī bistānō. twazzaʿū yaʿnī, fī kiškinār u fī bistānō. [...] abūy fī tunbō. [...] ana jaddī llī māt allāh yirḥamu kān mitzawwij arbaʿa au ṯalāṯa,*

[58] Tunbū splits into two adjoining villages, Eastern Tunbū and Western Tunbū. The two villages lie 10 km NW of Kūškinār.

arbaʿa ḥarīm yaʿnī aḫaḏ waḥda min bistānō, ḫawālī min bistānō, aḫaḏ waḥda min kiškinār illī hī ummī, wa aḫaḏ waḥda min fawāris. aḫaḏ iṯnēn min kiškinār, wāḥid min fawāris u waḥda min bistānō.

QA1: 'My mother is also from Kūškinār. [...] My grandfather had four [wives]. I mean, I have an uncle for example in Bustānū[59], and my mother is the only one from Kūškinār. The others are from Bustānū. They dispersed, I mean, in Kūškinār and Bustānū. [...] My father is from Tunbū. [...] My grandfather who passed away, may God have mercy upon him, was married to four or three, four women, I mean. He married one from Bustānū, my uncles are from Bustānū. He married one from Kūškinār, who is my mother, and married one from Fawāris. He married two from Kūškinār, one from Fawāris and one from Bustānū.'

QA1's family and many 'Arabs of the Coast' in Qatar, have submitted paperwork for citizenship. The country's rulers, however, requested the al-Mālikī tribe to join the Āl Ḥaram upon granting citizenship, which infuriated tribal members and halted the application process. They regard this move as an assault on their tribal pride and integrity, and refuse to partake in the effort to increase the number of Āl Ḥaram tribesmen in Qatar. QA1 narrated this dilemma the following way:

QA1 [GCA]: *bass ḥālīyan fī mašrūʿ əl-ḥīn, bass, rāḥ yiḫallūn, aiš, əl-ḥaram. iḏā yaʿṭūn əl-basbōrt əl-qaṭarī, gālō, anḍummkum maʿa l-ḥaram, maʿa qabīlat əl-ḥaram. abūy zaʿlān yaʿnī liʾanna ḫiyāra əl-mālkī, lēš akūn maʿa l-ḥaram?* [...] *ḥinni miš ʿandinā jawāz ḫalāṣ bi l-istimrār. Fa-iḏā tabbūn əl-jawāz, yaʿnī, iḏā yaʿṭīkum əl-jawāz əl-qabīla titġayyar itkūn əl-ḥaramī. maʿa l-ʿarab əl-ḥīn, əl-ḥīn fī mašrūʿ hina fī qaṭar, ʿarab əs-sāḥil jamīʿ jamīʿ əl-manāṭiq fī ʿarab əs-sāḥil illī maujūdīn fī qaṭar rāḥ yikūnū maʿa qabīlat əl-ḥaram ʿašān tikbar əl-qabīla wa yaʿṭūnhum jinsīya qaṭarīya. abūy ana rāfiḍ yigūl inn-ana, šisma, šisma, ana mālkī lēš akūn əl-ḥaramī? lēš atġayyar əl-family? wa nimšī nāḫiḏ əl-jinsīya ḫalāṣ ʿašān nigdar nisāfir u-nrūḥ.* [...] *fa-əl-ḥīn iḏā tabbī l-basbōrt əl-qaṭarī lāzim əl-qabīla tkūn aiš, əl-ḥaram.*

QA1: 'At present there is a plan now, that they will keep what? The Āl Ḥaram tribe. If they give the Qatari *passport* [= citizenship; English word], they said, join the Āl Ḥaram tribe. My father is angry, I mean, because his choice is the Mālikī tribe. [He'd say:] Why should I be with the Āl Ḥaram? [...] We continuously don't have a passport at all. So if you want the passport, I mean, if they give you the passport, the tribe changes, and it will be Āl Ḥaram. With the Arabs now, there is a plan here in Qatar that the 'Arabs of the Coast' [from] all regions of the [East] Coast, who live in Qatar, will be [affiliated] with the Āl Ḥaram tribe in order to grown the tribe. And they will give them Qatari citizenship. My father refuses [because] he says that I am, what's it called, what's it called, I am Mālikī, why should I be Āl Ḥaram? Why should I change the *family* [English word]? We should [just] go and take the citizenship, that's it, so that we can travel and move around. [...] So now if you want the Qatari *passport* [= citizenship; English word], the tribe has to be what? Āl Ḥaram.'

[59] Bustānū village is located on the coast 12 km south of Pārsiyān.

The intricacies of tribal politics had a negative effect on these young Arabs. It planted a sense of disenfranchisement in them for being sidelined by a government that bestowed citizenship on fellow tribesmen who had resettled from Iran just 15-20 years prior to them. Although they have no reason not to call Qatar their 'home country', the notion of marginalization shifts them away from speaking in the Qatari dialect.[60] QA1 expressed his fascination with the Saudi Arabian dialect of Riyadh, and he deliberately dropped typical phonological features of Qatari Arabic in favor of the dialect he called 'the purest'. Examples from his speech are *jadd* (~ GCA *yadd*) 'grandfather', *gālō* (~ GCA *gālau*) 'they said' and *tabbūn* (~ GCA *tabbōn*) 'you (pl.) want'. This phenomenon clearly demonstrates how the collusion between tribal politics and group identity influences language use.

6. Conclusions

During the past ten years, online technological advances and broader internet accessibility have offered unprecedented resources for an understudied segment of Arabs on the Iranian side of the Persian Gulf and Arab Gulf states to document their language, culture and heritage. The 'Arabs of the Coast' have launched a surging social media activity to send life signals to their fellow tribesmen in the Arab Gulf states and the rest of the world. The younger generations of these Arabs are bilingual in Arabic and Persian, while the older generations may only understand Arabic. The 'Arabs of the Coast' constitute a diverse community, where residents of every Arab village on the Iranian coast may have a different tribal background and a diverging set of ties to the Arabian Peninsula. The level of Persianization among them is very low, as 'Arabs of the Coast' refuse to associate themselves with the Persian national and Shiite religious identity of the Islamic Republic of Iran. They feel that their culture and language are firmly rooted in Arabia, and would rather see themselves as an integral part of the societies of the Arab Gulf states. However, this is often restricted by strict citizenship regulations in those countries. At the same time, online discussions and my informants complained that Gulf Arabs have an inclination to consider 'Arabs of the Coast' as *ʿAjam* 'non-Arabs, Persians', judging from the fact that they were born in Iran and also speak Persian. 'Arabs of the Coast' claim that this approach disregards tribal and religious aspects, which should play a significant role in defining the predominantly Sunni Gulf Arab's stance toward Shiite Persians[61].

Despite the existence of a seemingly vibrant Arab cultural life in certain parts of the Iranian side of the Persian Gulf, the contributors to the above-analyzed websites,

[60] For a thorough analysis of the label Qatari Arabic, see Bettega 2014.

[61] Suleiman (2003: 57) noted: "ʿArabi must be treated as a term of positive and specific inclusion, while, in contrast, ʿAjami must be viewed as a term of exclusion, residual inclusion or inclusion by default. Looked at from a different angle, ʿArabi is a term of in-groupness, while ʿAjami is a term of out-groupness."

channels, social media forums occasionally magnify the importance of contemporary Arabs by blending older footage with more recent ones. Reliable statistics on the number of Arabs in Iran are difficult to acquire as Iranian statistical yearbooks are cautious about publishing data on the ethnic composition of Būšihr and Hurmuzgān provinces. With the expansion of the oil and tourism industry across the Iranian shore of the Persian Gulf, the region has seen decade-long efforts by Iranian authorities to Persianize the local population, and to broaden the influence of the country's dominant Persian national identity. Sunni Arabs have an inclination to alienate themselves from the Persian-speaking Shiite majority population. This prompts the younger generations of 'Arabs of the Coast' to lead a concerted effort to maintain their distinctive Arab character. The pillars of their identity thus constitute a strong advocacy for the Arabic language and Sunni Islam. However, the same younger generations who wish to receive higher education must move out of Arabic-speaking villages, and settle down in larger Persian-speaking towns, or migrate to the affluent Arab Gulf states.

'Arabs of the Coast' speak and write in multiple varieties of Gulf Colloquial Arabic (GCA). The Arabic dialects on Iran's Persian Gulf coast have been underrepresented in academic studies. Therefore, a large-scale collection, documentation, and thorough linguistic analysis would help preserve the cultural heritage of ethnic Arabs in Iran. My hope is that this research will also contribute to the field of interactional sociolinguistics, computer-mediated discourse analysis and linguistic anthropology, in addition to opening up new horizons in discussions on cultural and religious identity in a region where Arabs, Persians and many other ethnic groups share a common history.

Appendix: Geographical Names in Iran

Geographical names are organized into two groups that reflect their location in two Iranian provinces, and are listed from northwest to southeast.

Bushehr Province (Būšihr): Bandar Ganaveh or Genaveh (Gināwa), Kharg Island (Ḫārg), Banak (Bnak), Bandar Kangan (Kangān), Asaluyeh (Persian: ʿAsalūya, Arabic: Aḥsilō), Dehnow (Dihnau), Akhand (Persian: Aḫand, Arabic: Ḫand), Kheyaru (Persian: Ḫiyārū, Arabic: Ḫiyārō), Bostanu (Persian: Bustānū, Arabic: Bistānō), Nay Band (Persian: Nāyband, Arabic: Nāband), Haleh (Persian: Hāla, Arabic: al-Ḥālla), Besatin (Basātīn), Zobar (Zubār), Chah Mobarak (Persian: Čāh Mubārak, Arabic: Jahəmbārak)

Hormozgan Province (Hurmuzgān): Tombu (Persian: Tunbū, Arabic: Tunbō), Farsi (Persian: Fārsī, Arabic: Fawāris), Surobash (Persian: Sūrbāš), Koshkonar or Kushk-e Nar (Persian: Kūškinār, Arabic: Kiškinār), Parsian (formerly Gāwbandī, Persian: Pārsiyān), Lavan Island (Persian: Lāwān, Arabic: Šaiḫ Šuʿaib), Bandar-e Moqam (Persian: Bandar Muqām, GCA: al-Mugām or al-Magām), Nokhaylo (Persian: Naḫīlū, Arabic: Naḫīlō), Nakhl-e Jamal (Naḫl Jamāl), Chiruiyeh (Persian: Čīrūʾīya, Arabic: Šīrū), Morbagh (Murbāġ), Rostaq (Rastāq), Ainat (ʿAināt), Kalat (Kalāt), Kish Island (Persian: Kīš, MSA: Qais, GCA: Gēs), Bandar Lengeh (Persian: Linga, Arabic: Linja or Linya), Qeshm Island (Qišm), Bandar Kong (Bandar Kung), Laft (Lāft), Hengam Island (Hingām), Bandar Abbas (Bandar ʿAbbās)

BIBLIOGRAPHICAL REFERENCES

Al-Anṣārī, J. Ḥ. (2011), *Tārīḫ ʿArab al-Haula wa al-ʿUtūb*, Bairūt, Al-Dār al-ʿArabīya li al-Mausūʿāt

Al-Anṣārī, J. Ḥ. (2014), *Qabāʾil ʿArab al-Haula fī ʿUyūn al-Raḥḥāla wa al-Mustašriqīn*, Bairūt, Al-Dār al-ʿArabīya li al-Mausūʿāt

Androutsopoulos, J. (2008), *Potentials and Limitations of Discourse-Centred Online Ethnography*, in "Language@Internet", 5, article 9, accessed on 4 April 2017, http://www.languageat internet.org/articles/2008/1610/androutsopoulos.pdf

Bettega, S. (2014), *The Use of Dialectal Tanwīn in Qatari Arabic*, in "Ricognizioni, Rivista di Lingue, Letterature e Culture Moderne", Vol. 1, No. 2: 15-25

Bosworth, C. E. (2012), *Linga*, In: *Encyclopaedia of Islam, Second Edition*. Brill Online, 2012, University of Iowa, accessed on 18 December 2016, http://referenceworks.brillonline.com.proxy.lib.uiowa.edu/entries/encyclopaedia-of-islam-2/linga-SIM_4669

Bosworth, C. E. (2011), *ʿAjam*, in *Encyclopædia Iranica*, I/7: 700-701; an updated version is available online at http://www.iranicaonline.org/articles/ajam (accessed on 25 April 2017)

Daryāʾī, ʿA. (2012), *Umarāʾ ʿArab al-Haula. Āl Naṣūr wa Āl Ḥaram*, Bairūt, Al-Dār al-ʿArabīya li al-Mausūʿāt

Al-Dailami, A. (2014), *"Purity and Confusion": The Hawala between Persians and Arabs in the Contemporary Gulf*, in L. G. Potter (Ed.), *The Persian Gulf in Modern Times. People, Ports, and History*, New York, NY, Palgrave Macmillan: 299-326

Dihḫudā, ʿA. A. (2006), *Luġatnāma-yi Dihḫudā*, accessed on 18 December 2016, http://www.parsi.wiki/

Floor, W. (2010), *The Persian Gulf. The Rise and Fall of Bandar-e Lengeh*, Washington DC, Mage Publishers

Gazsi, D. (2016), *From Arabia to Persia and Back: Code-switching among the Āl ʿAlī Tribe in the UAE and Iran*, unpublished manuscript

Haddadian-Moghaddam, E., Meylaerts, R. (2015), *What about Translation? Beyond "Persianization" as the Language Policy in Iran*, in "Iranian Studies", Vol. 48, No. 6: 851-870

Katouzian, H. (2000), *State and Society In Iran. The Eclipse of the Qajars and the Emergence of the Pahlavis*, London, New York, I. B. Tauris

Lee, C. K. M. (2011), *Micro-Blogging and Status Updates on Facebook: Texts and Practices*, in Thurlow, C., Mroczek, K., *Digital Discourse: Language in the New Media*, New York, Oxford University Press

Lorimer, J. G. (1970), *Gazetteer of the Persian Gulf, ʾOmān, and Central Arabia*, Volume II A & II B, Westmead, Gregg International Publishers

Niʿma, ʿA. Ḥ. (2001), *al-Haula al-Qabīla al-ʿĀmirīya*, al-Dauḥa

Riazi, A (2005), *The Four Language Stages in the History of Iran*, in A. M. Y. Lin, P. W. Martin (Eds.): *Decolonisation, Globalisation: Language-in-Education Policy and Practice*, Clevedon, Buffalo, Toronto, Multilingual Matters Ltd: 98-114

Saršumārī-yi ʿumūmī-yi nufūs wa maskan, Ustān-i Hurmuzgān (2011-2012 – 1390 Solar *Hijrī*). Available on https://www.amar.org.ir/

Saršumārī-yi ʿumūmī-yi nufūs wa maskan, Ustān-i Hurmuzgān (2006-2007 – 1385 Solar *Hijrī*). Available on https://www.amar.org.ir/

Schofield, R. (Ed.) (1990), *Islands and Maritime Boundaries of the Gulf*, Volume 7, 1920-1930, Oxford, Archive Editions

Sheyholislami, J. (2012), *Kurdish in Iran: A Case of Restricted and Controlled Tolerance*, in "International Journal of the Sociology of Language", 217: 19-47

Ṣiddīq, ʿA. M. (1994), *Ṣahwat al-Fāris fī Tārīḫ ʿArab Fāris*, Al-Šāriqa, Maṭbaʿat al-Maʿārif

Srebnerny, A., Khiabany, G. (2010), *Blogistan: The Internet and Politics in Iran*, London, I.B. Tauris

Suleiman, Y. (2003), *The Arabic Language and National Identity: a Study in Ideology*, Edinburgh, New York, Edinburgh University Press

DÉNES GAZSI • Worked as an Assistant Professor of Middle Eastern languages and cultures, and Co-Director of Arabic Studies at the University of Iowa. As a cultural linguist, his research focuses on the intersection of ethnolinguistics and Arabic-Persian language/cultural contact. He explores the dynamics of Arabic-Persian language contact in Classical Persian literature as well as present-day southwestern Iran (Persian Gulf coast, Khuzestan) and the Arabian Peninsula. As an Arabic and Persian dialectologist, he specializes in the vernaculars of the Persian Gulf and the broader Mesopotamian region.

E-MAIL • gazsidenes@gmail.com

SOURCES OF GULF PIDGIN ARABIC FEATURES

Andrei A. Avram

ABSTRACT • This paper is an attempt at identifying the sources of a number of selected features of the phonology, syntax and vocabulary of Gulf Pidgin Arabic. It is shown that these features can be traced back to a variety of sources: the substrate languages; the superstrate (input from specific Arabic dialects); the Arabic Foreigner Talk register; incipient grammaticalization; convergence of several factors. Also discussed are some implications of the findings with reference to the "feature pool hypothesis".

KEYWORDS • Gulf Pidgin Arabic, substrate, Gulf Arabic, Foreigner Talk, grammaticalization

1. Introduction

The aim of the present paper is to identify the origin of a number of selected features of Gulf Pidgin Arabic[1] (henceforth GPA).

The data examined are from general descriptions of GPA (Smart 1990, Wiswall 2002, Avram 2013, 2014a, 2016a) as well as from studies and a corpus collected from 25 websites, illustrative of GPA as used in specific territories: Saudi Arabia (henceforth SA) – Hobrom (1996), Almoaily (2008), Al-Azraqi (2010), Albakrawi (2012), Almoaily (2013), Alghamdi (2014), Almoaily (2014), Al-Shurafa (2014), Al-Zubeiry (2014), online sources; Kuwait (K) – Salem (2013); UAE – Smart (1990), Yammahi (2008), online sources), Bahrain (B) – online sources), Qatar (Q) – Bakir (2010); Oman (O) – Næss (2008), Alshuaimi (2011), online sources.

To ensure homogeneity, all examples are transliterated in a uniform system. Relevant items in examples appear in boldface. The length of quotations has been kept to a reasonable minimum. All quotations are accompanied by glosses and their translation.

[1] Alternative names include "Arabic Gulf Pidgin", "Urdu Pidgin Arabic", and "Asian Pidgin Arabic".

The structure of the paper is as follows. Section 2 illustrates the influence of the substrate languages. Section 3 deals with the dialectal Arabic input. Section 4 focuses on the role of the Arabic Foreigner Talk register. Section 5 is concerned with incipient grammaticalization. Section 6 considers the potential convergence of several factors. Section 7 is an assessment of the explanatory adequacy of the "Feature Pool Hypothesis" in light of the GPA data analyzed. Section 8 summarizes the findings.

2. Substrate languages

A number of phonological, morphological, syntactic and lexical features of GPA can be assigned to the influence of the substrate languages. These include, among others, Bengali, Hindi, Indonesian, Javanese, Kannada, Malayalam, Nepali, Pashto, Punjabi, Persian, Sinhalese, Tagalog, Tamil, Thai, and Urdu.

2.1. Phonology

Generally, the phonology of GPA is characterized by a significant reduction of the inventory of vowel and consonant phonemes of Arabic and in the number of phonological contrasts.

As shown by Næss (2008: 42), Salem (2013: 107), Avram (2014a: 15, 2016a: 62, 2016b: 87-88), in GPA vowel length is not distinctive. Although phonetically short and long vowels both occur, these alternate rather randomly, as in (1a-c), and do not play any phonological role.

(1) a. *bād**e**n* 'then' Q (Avram 2014a: 16)
b. *bad**e**n* 'then' SA (Avram 2014a: 16)
c. *g**u**l* ~ *g**ū**l* 'to say' O (Næss 2008: 42)
d. *kat**i**r* 'a lot' SA (Avram 2014a: 16)
e. *kab**ī**r* 'big' SA (Avram 2014a: 16)

The marked consonant phonemes of Arabic – e.g. the velar and pharyngeal fricatives as well as the so-called "emphatics", involving a secondary articulation[2] – are either replaced or lost (Hobrom 1996: 57-58, Almoaily 2008: 36-37, Næss 2008: 30-43, Salem 2013: 107, Avram 2014a: 15, 2016a: 62, 2016b: 87):

(2) a. /ḫ/ → [h]: *wæ**h**īd* 'one' SA (Almoaily 2008: 37)
b. /ḫ/ → [k]: ***k**ubus* 'bread' SA (Hobrom 1996: 43)
c. /ḥ/ → [h]: *rū**h*** 'to go' (Bakir 2010: 207)
d. /ḥ/ → [k]: *a**k**san* 'best' O (Næss 2008: 39)

[2] Variously described in the literature as velarization or pharyngealization.

e. /ġ/ → [g]: ***g**isli* 'to wash' Q (Bakir 2010: 209)
f. /ʕ/ → Ø: *araf* 'to know' K (Salem 2013: 107)
g. /ṣ/ → [s]: *hala**s*** 'finished' K (Salem 2013: 107)

Consonant germination, although occasionally attested, is not phonemic. Consonants frequently undergo degemination (Næss 2008: 36, Avram 2014a: 15, 2016a: 62), as in the following example:

(3) *si**t**a* 'six' K (Salem 2013: 107)

Also replaced are other consonants which are not attested in the L1 of the users of GPA. This is conducive to both intra-speaker and, in particular, inter-speaker variation (see Avram 2013b, 2014a: 15-16, 2016: 87). For instance, Javanese, Indonesian, Sinhala, and Tagalog speakers (may) realize /f/ as [p]:

(4) *na**p**ar* 'person' SA (Avram 2014a: 17)

Indonesian and Javanese speakers (may) realize /z/ as [s] or [ʤ]:

(5) ***s**ēn ~ **j**ēn* 'good' O (Næss 2008: 32, 34)

2.2. Syntax

Although GPA has SVO word order (Alghamdi 2014: 122), there are occasional instances of patterns typical of SOV. For instance, direct objects may occur in pre-verbal position:

(6) a. *badēn* ***sābūn hattēti*** Q (Avram 2014a: 25)
then soap put
'then [I] put soap'
b. *ana* ***čiko sūp*** O (Avram 2014a: 24)
1SG child see
'I [will] see [my] children'

Almoaily (2013: 154) reports that 12.3% of the old speakers and 8.7% of the new speakers [= recently arrived users of GPA exhibit SOV word order, while Alghamdi (2014: 122) finds that 18% of the sentences in his sample have SOV word order.

In attributive possession constructions the possessor[3] may precede the possessee:

[3] Note that GPA tends to use independent pronouns exclusively to encode the possessor; see also 4.3.

(7) a. ***ana mama*** *w aku* SA (Alghamdi 2014: 15)
1SG mother and brother
'my mother and my brother'
b. ***ana sadiki*** *bēt* O (Næss 2008: 61)
1SG friend house
'my friend's house'

(8) ***ana māl bint*** *tālīm arabi* O (Næss 2008: 63)
1SG POSS daughter learn Arabic
'my daughter is learning Arabic'

Adverbs may occur in pre-verbal position:

(9) ***sem sem*** *kalām* SA (Avram 2014a: 25)
same speak
'they speak in the same way'

Also attested are postpositions:

(10) a. *itnēn sana* ***badēn*** SA (Almoaily 2008: 105)
two year after
'after two years'
b. *zamal* ***fōk*** O (Avram 2014a: 25)
camel above
'on top of the camel'

Finally, also attested are pre-nominal relative clauses:

(11) ***taʔšira māl umān*** *nafarāt* O (Avram 2014a: 25)
visa POSS Oman persons
'People who have Omani visas'

Since the substrate of GPA includes many SOV languages, e.g. Bengali, Hindi/Urdu, Malayalam, Punjabi, Persian, Sinhala, this accounts for the occurrence of patterns typical of this word order.

2.3. Lexical items

GPA is an Arabic-lexifier language and therefore shares most of its vocabulary with Gulf Arabic. However, the vocabulary of GPA includes a number of loanwords from Hindi/Urdu and Persian (see also Smart 1990: 114).

(12) a. *ğaldi ğaldi* SA (Online 2014)
b. *ğeldi ğeldi* Q (Online 2016)
c. *ğildi ğildi* UAE (Smart 1990: 96)
'quickly'

The GPA variants *ğaldi ~ ğeldi ~ ğeldi* can be traced to Urdu/Hindi *jaldī* 'quick'.

(13) *ana* *ača* UAE (Online 2007)
1SG fine
'I'm fine'

GPA *ača* comes from Hindi/Urdu *aччhā* 'good, very well'.

(14) *ruh* ***sīda*** UAE (Online 2014)
go straight
'[I] go straight'

The GPA word *sīda* is from Urdu/Hindi *sīdha* 'straight'. Note that the word is also attested in Gulf Arabic, which may have borrowed it not directly from Urdu, but rather from GPA (Almoaily 2013: 184).

3. Superstrate: Dialectal Arabic input

Given the similarities between the dialects of Gulf Arabic it is rather difficult to identify in GPA features which can safely be attributed to influence from a specific variety. Nonetheless, a number of such cases exist.

3.1. Existentials

Salem (2013: 109) illustrates the use of existential *aku/maku* in GPA as spoken in Kuwait:

(15) *Ana* ***maku*** *šay.* K (Salem 2013: 109)
1SG NEG EXIST thing
'I don't have anything.'

Existential *aku/maku* is of Kuwaiti Arabic origin. According to Holes (2007: 615), "*āku* is the local form, which Kuwait shares with southern Iraq". This explains why the same use of *aku/maku* is documented in Romanian Pidgin Arabic[4]:

(16) a. ***Aku*** *fulus* *la* *inte?* (Avram 2010: 25)
EXIST money PREP 2SG
'Have you got any money?'
b. ***Maku*** *saa* *la* *ani* (Avram 2010: 29)
NEG-EXIST watch PREP 1SG
'I don't have a watch.'

[4] Formerly spoken in Iraq.

On the other hand, GPA as spoken in Kuwait also makes use of an alternative form:

(17) *Ana* ***fe*** *itnen* *paččа* K (Salem 2013: 108)
1SG EXIST two child
'I have two children.'

Again, this variant can be traced to Kuwaiti Arabic in which "*fīh* is also common" (Holes 2007: 615).

3.2. Lexical items

Both Smart (1990: 114) and Næss (2008: 36) attest the occurrence of the word *čiko* 'child, baby' in GPA as spoken in Oman:

(18) *bādēn* ***čiko*** *yiği* *marīd* O (Næss 2008: 36-37)
then baby come ill
'Then when the baby comes, [it's] ill.'

Smart (1990: 114) states that in GPA this is "the only word from other languages [= other than English, Persian, Urdu] is تشيكو شيكو *tshīko*, said to be from Portuguese *chico*". As shown in Behnstedt and Woidich (2011: 44, map 19 Baby), *šīko* ~ *čīko* is attested only in Oman. As for its etymology, Brockett (1985: 135) compares these forms with Classical Arabic *šakw* 'small lamb/camel" and Spanish *chico*. However, as noted by Behnstedt & Woidich (2011: 46), "erwarten würde man eigentlich eher eine Entlehnung aus dem Portugiesichen". To conclude, GPA *čiko* 'child, baby' is from Omani Arabic, in which it is probably a loanword.

3.3. Polysemy

As illustrated by the examples below, the verb *iğlis/yiglis/yiğlis* has become polysemous in GPA as recorded in Oman by Næss (2008):

(19) a. *awwal* ***yiğlis*** *andel* *sandūg* *māl* *cash* (Næss 2008: 65)
first sit PREP box POSS cash
'At first I was sitting at the cash register'
b. *mafi* ***iğlis*** (Næss 2008: 74)
NEG rest
'[We] can't rest'
c. ***yiglis*** *bēt* (Næss 2008: 74)
sit house
'stay at home'

The polysemy of *iğlis/yiglis/yiğlis* in GPA as used in Oman could be traced back to Omani Arabic (Maria Persson, p.c. November 2013).

4. Arabic Foreigner Talk

As is well known, the Foreigner Talk register (henceforth FT) is believed to have played a role in the genesis of pidgin languages (see e.g. Ferguson and DeBose 1977, den Besten et al. 1995: 95-97, Mühlhäusler 1997: 96-102, Siegel 2008: 26-27). According to Ferguson (1971/1996: 121), for instance, "the foreigner talk of a speech community may serve as an incipient pidgin". In his view, "the initial source of the grammatical structure of a pidgin is the more or less systematic simplification of the lexical source language which occurs in the foreigner talk register of its speakers (Ferguson 1971/1996: 121).

There appears to be a correlation between the role of the FT input and the developmental stage of pidgins. Mühlhäusler (1997: 102) writes that "the importance of foreigner talk in Pidgin formation appears to be restricted to relatively early stages of development". With respect to GPA, Næss concludes that it is "a variety on the way to becoming conventionalized and unified as a first-generation contact language"[5]. Avram (2016b: 95) shows that it exhibits most of the features typical of jargons[6] (in the classification suggested by Mühlhäusler 1997: 6). In other words, the developmental stage of GPA is that in which FT influence is to be expected[7].

The potential role of Arabic FT in the emergence of GPA has been mentioned in the literature. Tosco and Manfredi (2013: 510), for instance, state that "certainly the influence of foreigner talk was important in the genesis of GPA", but do not make reference to any specific structural features. As shown below, several such features may arguably be traced to the Arabic FT input (see also Avram 2014b). The Arabic FT data are from the following varieties: Saudi Arabic FT (Al-Ageel 2016), Kuwaiti Arabic FT (Wiswall 2002, Dashti 2013), and Omani Arabic FT (Brockett 1985, Næss 2008).

4.1. 'Two' + singular noun

As illustrated by the examples below, GPA has not retained the Arabic dual on nouns, which is replaced by a noun phrase made up of the numeral 'two' and a singular noun (see e.g. Almoaily 2008: 52, Alghamdi 2014: 120, Avram 2014a: 17):

(20) a. ***tanēn mara*** SA (Alghamdi 2014: 14)
two time
'twice'

b. ***itnen pačča*** K (Salem 2013: 108)
two child
'two children'

[5] The term "conventionalized" suggests that GPA has not reached the stabilization stage (in the sense of Mühlhäusler 1997: 138).

[6] Also known as "minimal pidgins" or "pre-pidgins" in the literature on creole and pidgin languages.

[7] The possibility that Arabic FT is influenced by GPA is explored in Avram (2014b).

c. ***tinēn usbū*** Q (Avram 2014a: 17)
two week
'two weeks'

d. ***isnēn sana*** O (Avram 2014a: 17)
two year
'two years'

This parallels, for instance, the situation in Kuwaiti Arabic FT, in which, according to Dashti (2013: 78), "the cardinal number [...] followed by the singular noun" is used "even if a dual [...] is required":

(21) *ṭalʔi* ***aṯnēn*** ***diǧāǧa*** (Dashti 2013: 19)
get out two chicken
'prepare two chickens'

4.2. Omission of definite article

GPA does not make use of the definite article of Arabic:

(22) a. **Ø** *kafīl fī sawwi ǧinǧāl* Q (Bakir 2010: 217)
sponsor FI make quarrel
'The sponsor quarrels [with me]'

b. **Ø** *Muškil eš?* O (Alshuaimi 2011)
problem what
'What's the problem?'

The omission of the definite article in can also be illustrated with examples from Saudi and Kuwaiti Arabic FT, in (23) and (24) respectively:

(23) *kīs* **Ø** *ǧīb* (Al-Ageel 2016: 167)
bag bring.IMP.2SG.M
'Bring the bag'

(24) *Tati, eš fīč dāḫil* **Ø** *seyāra.* (Dashti 2013: 73)
Tati what in 2SG.F inside car
'Tati, what's wrong with you, it's in the car.'

4.3. Masculine singular form of adjectives

In GPA adjectives do not agree in gender and number with the nouns they modify. As stated by Hobrom (1996: 81), "adjectives retain the masculine singular form in most cases", i.e. the etymologically masculine singular is used as a default form:

(25) *mumkin hiya* ***tābān*** O (Næss 2008: 41)
maybe 3SG.F tired
'Maybe she's tired.'

Both the lack of gender and number agreement of adjectives and the use of the masculine singular form are attested in Saudi and Kuwaiti Arabic FT. The examples of Saudi Arabic FT in Al-Ageel (2016: 176-177) are illustrative of the tendency towards using the masculine singular form of adjectives.

(26) ***aṣfar*** *ǧubnah* *wāḥid* (Al-Ageel 2016: 176)
yellow cheese one
'one yellow cheese [sandwich]'

As for Kuwaiti Arabic FT, Dashti (2013:80) writes that "when Kuwaitis communicate with their domestic workers, they usually use the masculine singular [...] only" (Dashti 2013: 80).

(27) *šīli* *hāḏa* *ǧiwāti* ***qadīm*** (Dashti 2013: 80)
take DEM shoes old
'keep all these old shoes'

4.4. Exclusive use of independent pronouns

GPA tends to use independent pronouns exclusively (Næss 2008: 52, Al-Azraqi 2010: 171, Avram 2016a: 64). These replace the pronominal suffixes of Arabic and are used both as object pronouns (Avram 2014a: 17-18) and to encode the possessor (Hobrom 1996: 81, 25-26, Avram 2014a: 25-26):

(28) a. *māmā* *yabi* ***anā*** Q (Avram 2014a: 17)
Madam wants 1SG
'Madam wants me.'
b. *kulo* *fulus* ***anta*** S (Avram 2014a: 26)
all money 2SG
'all your money'

Similar uses of independent pronouns are attested in e.g. Kuwaiti Arabic FT:

(29) *hāḏa* *ǧawāz* *māl* *inte* (Dashti 2013: 75)
DEM passport POSS 2SG
'This is your passport.'

4.5. Masculine singular form of demonstratives

As in the case of adjectives, the only form of demonstratives found in GPA is that of the masculine singular (Hobrom 1996: 82):

(30) a. ***Haza*** *nafarat* *zen* K (Salem 2013: 108)
DEM men good
'These men are good.'

b. *asān* ***hāda*** *mama* *kalām* *arabi* O (Avram 2014a: 29)
because DEM madam speak Arabic
'Because the madam [only] spoke Arabic'

There is evidence that the masculine singular is also the default form in Saudi and Kuwaiti Arabic FT. The samples of the former in Al-Ageel (2016) include such instances and Dashti (2013: 83) explicitly mentions, with reference to the latter, that "[Kuwaitis usually use *hāða* exclusively". Two such examples, from Saudi Arabic FT (31) and from Kuwaiti Arabic FT (32) are provided below:

(31) *Ana* ***hāḏa*** *ġurfa* *147* (Al-Ageel 2016: 180)
1SG DEM.SG.M room-SG.F 147
'I'm in room 147.'

(32) *ǧībi* ***hāḏa*** *ǧanṭə* *killə* *māl* *āna* (Dashti 2013: 77)
bring DEM.SG.M bag-SG.F all POSS 1SG
'bring my bags'

4.6. Invariant form of verbs

As already shown by Bakir (2010: 206), in GPA "the common verb form [...] is the Gulf Arabic 3rd person singular masculine imperfect form [which is] used with all subjects, regardless of their person, gender, or number". In addition, verb forms derived etymologically from Gulf Arabic imperatives are also frequently found (Bakir 2010: 207-208). Moreover, Bakir (2010: 208) notes that "we may find the two form occurring in the speech of the same speaker):

(33) *ʔanā* *fī* *lēl* *gūm* *yabi* *rūh* *yišrab* *māy* Q (Bakir 2010: 208)
1SG in night stand want go drink water
'At night, I got up to go and drink some water.'

A similar situation is documented for Saudi Arabic FT. Al-Ageel (2016: 167) writes not only that "the use of verbs is unstable", but also that "two forms of a verb can be used alternatively within the same conversation by the same speaker":

(34) *kīs* ***ǧīb*** *baʕdēn* *zabādī*
bag bring.IMP.SG.M then yogurt
yiǧīb *enti* (Al-Ageel 2016: 167)
3SG.M-bring 2SG
'Bring the bag, then bring the yogurt.'

Consider also the following examples from Kuwaiti and Omani Arabic FT, in (35) and (36) respectively:

(35) *wen inti* ***yiḥuṭ*** *ams āna* ***yaʔṭīč****?* (Dashti 2013: 71)
where 2SG 3SG.M-put yesterday 1SG 3SG.M-give-2SG.F
'Where did you put the bag I gave you yesterday?'

(36) *mā fī* ***yaʕraf*** (Brockett 1985: 25)
NEG FI 3SG.M- know
'I don't know.'

4.7. Light verb 'make' + verb

As first noticed by Bakir (2010: 220), "the verb *sawwi* 'do' [...] is used with a following [...] verb":

(37) *sawwi tiktib* Q (Bakir 2010: 221)
make write
'write'

Light verb 'make' + verb structures are attested in Kuwaiti Arabic FT. Interestingly, Wiswall (2002) even reports an overuse of the *sawwi* + verb construction in this register, i.e. such structures occur more frequently in Kuwaiti Arabic FT than in GPA.

(38) *anta* ***sawwi fakkar*** (Wiswall 2002)
2SG make think.3SG.M
'you think'

4.8. Time adverbials to indicate tense and aspect

Since the verbal inflections of Arabic has not been retained, in addition to contextual clues, tense and aspect marking is expressed by means of time adverbials such as *alhīn* 'now', *ʔamis* 'yesterday', *awwal* 'before', *bādēn* 'later', *bāčir/bukra* 'tomorrow' (Hobrom 1996: 84, Almoaily 2008: 40, Næss 2008: 85, Bakir: 211-213):

(39) a. *tadrib* ***awwal*** *šwayy* SA (Albakrawi 2012: 129)
practice first a little
'I practiced a little.'

b. ***amis*** *anā yabi ...* Q (Bakir 2010: 206)
yesterday 1SG want
'Yesterday, I wanted ...'

The same holds for e.g. Kuwaiti Arabic FT. In the following example it is the time adverbial which indicate the past time reference.

(40) *anā yigūlič* ***misāʕ*** (Dashti 2013: 71)
1SG 3SG.M-tell-2SG.F from hour
'I told you an hour ago'

4.9. Multi-purpose preposition māl

In GPA *māl* often functions as a multi-purpose preposition (see Avram 2013b, 2014a: 23-24), covering a variety of meanings which include 'at', 'in', 'for', and 'to'. Two examples illustrative of its use are given below:

(41) a. *sawwi* **māl** *ʔāna muškil* Q (Avram 2014a: 23)
make of 1SG problem
'[she] makes a problem for me'
b. *binti fi āti* **māl** *walad* O (Avram 2014a: 24)
daughter FI give PREP son
'[My] daughter gives [it] to my son'

Similar uses of *māl*, although not explicitly mentioned in the literature, are attested in the Arabic FT register. Consider the following example from Omani Arabic FT:

(42) *hāḏa abūy ġawāzīk u-l-maʕāš* **māl**
DEM father-1SG.POSS money -2SG.POSS and-DEF-salary PREP
ḫamsa šuhūr (Simone Bettega, p.c. February 2016)
five month.PL
'this, my dear, is your money, the salary for five months'

5. Incipient grammaticalization

Since GPA emerged as early as the 1960s (Smart 1990: 83) and grammaticalization is assumed to proceed at a more rapid pace in pidgin and creole languages, grammaticalization is yet another potential factor in the emergence of GPA features.

Consider, for instance, the case of *kalas*/*kalās*, which appears to function as a completive aspect marker (Hobrom 1996: 84, Avram 2014a: 27), as illustrated below:

(43) a. *inta* **kalās** *waddi fulūs?* Q (Avram 2014a: 27)
2SG COMPL send money
'Have you sent the money?'
b. *atbuk* **kalās** Q (Avram 2014a: 27)
cook COMPL
'[I] have cooked.'

However, *kalas*/*kalās* is infrequently used and only by some speakers of GPA. Moreover, there is inter-speaker variation with respect to the position occupied by *kalas*/*kalās*: pre-verbal, as in (43a), or post-verbal position, as in (43b). Therefore, *kalas*/*kalās* is illustrative of what may be regarded as incipient grammaticalization[8].

[8] Cf. the similar use of *kalas* in Pidgin Madam (Bizri 2010: 126-127), spoken in Lebanon by Sri

6. Convergence of factors

It has been shown so far that specific structural features of GPA can be attributed to a particular source, such as the influence of the substrate languages, dialectal Arabic, the Arabic FT register or incipient grammaticalization.

However, some GPA features are arguably the result of the convergence of various factors. A few such cases are discussed in what follows.

6.1. Substrate languages, Arabic Foreigner talk

6.1.1. Light verb 'make' + noun/adjective

GPA makes extensive use of light verb 'make' + noun/adjective structures (Bakir 2010: 220-221).

(44) a. *huwa* ***sawwi*** ***talīm*** SA (Avram 2014a: 19)
3SG.M make learning
'he learns'

b. *lēš* *māmā* *māfī* ***sawwi*** ***tabdīl*** Q (Bakir 2010: 219)
why Madam NEG FI make change
'why doesn't Madam change [it]'

c. *ana* ***sawwi*** ***nadīp*** O (Næss 2008: 91)
1SG make clean
'I clean'

It is tempting to account for such structures by pointing to the Hindi/Urdu so-called "denominative verbs" noun/adjective + *karnā* 'make' and to the Persian so-called "compound verbs" noun/adjective + *kardan* 'make' (Bakir 2010: 221). However, the use of light verb 'make' + noun/adjective structures is also attested in several varieties of Arabic FT (see Avram 2014a: 36, 2014b), including Kuwaiti Arabic FT:

(45) *āna* ***yisawwi*** ***talifūn*** *ams* (Dashti 2013: 72)
1SG 3SG.M-make telephone yesterday
'I phoned yesterday.'

6.1.2. Lexical items

Some GPA words may also be traced back to both substrate languages and to the Arabic FT register. A case in point is the following example:

Lankan female domestic workers and their employers, and of *halas* in Romanian Pidgin Arabic (Avram 2010: 26).

(46) a. ***paččа*** K (Salem 2013: 108)
b. *bačča* UAE (Yammahi 2008: 37)
c. *bečā* O (Næss 2008: 27)
'child'

As shown in Behnstedt and Woidich (2011: 38, Map 18b Kind(er), *bačča* "Kind" is attested only in Uzbekistan Arabic; this obviously rules out an Arabic source for this GPA lexical item. Behnstedt and Woidich (2011: 40) specify that this form is etymologically derived "von pers. *bačče* "Kind". According to Næss (2008: 27), "the [GPA] word *bečā* [...] appears to be derived from Urdu *bačča* "child", which seems more likely, on phonetic grounds. One source of the GPA forms is therefore Hindi/Urdu *baččā*. On the other hand, Næss (2008: 27) observes that in GPA as spoken in Oman *bečā* 'child' is also "used by a Javanese and a Tagalog native speaker". Næss (2008: 27) further writes that, since these speakers "are stay-at-home maids, who one must assume had minimal contact with native Urdu speakers", this may indicate that "foreigner talk to non-Arabs, even within the home, is quite common among the Gulf Arab employees".

6.2. Substrate languages, Arabic FT, grammaticalization

6.2.1. Predicative copula fi

The occurrence of *fi* as a predicative copula is widely attested in GPA (Al-Azraqi 2010: 169-171, Bakir 2010: 216, Al-Shurafa 2014: 18, Avram 2014b: 20-21, Bakir 2014: 420, Potsdam and Alanazi 2014: 16, Avram 2016a: 66), but its use is not obligatory (Avram 2012):

(47) a. ***fi ahsan*** SA (Avram 2012: 20)
FI good
'it's alright'
b. *enta ma* ***fi zēn*** UAE (Online 2014)
2SG NEG FI good
'You are not good.'
c. *inta fī maǧnūn?* Q (Bakir 2014: 420)
2SG FĪ crazy
'are you crazy?'
d. *bilād* ***fi zeyn*** O (Alshuaimi 2011)
country FI good
'Hometown is good.'

Overt copulas (not only predicative, but also existential, equative[9] and locative) occur in many of the substrate languages. Hobrom (1996: 66), for instance, notes that

[9] Also called "specificational", see Bakir (2014: 419).

"the informants' native languages (Tamil, Kannada, and Malayalam) generally require a copula" and that "Hindi/Urdu [...] also requires a copula" and concludes that "we can justify the overuse of *fi* as a copula in the speakers' pidgin".

However, *fi* as a predicative copula is also found in Arabic FT. Consider the following example from Saudi Arabic FT:

(48) ***kwayes*** *mā* ***fī*** *baʕdēn* (Al-Ageel 2016: 171)
good NEG FĪ then
'It won't be good then'

Similarly, in Omani Arabic FT *fī* occurs "with adjectives" (Brockett 1985: 24):

(49) ***fī*** ***ḫalāṣ*** (Brockett 1985: 24)
FĪ ready
'[when] that's finished'

Brockett (1985: 25) notes that this use of *fī* is typical of the Arabic spoken by the Indians and Pakistanis in Oman, and suggests that "perhaps this is imitated by Omanis when talking to them and to other foreigners, thinking that by doing so they will be more easily understood".

Yet another possibility is that the use of *fi* as a predicative copula is being reinforced by a still ongoing grammaticalization process: COPULA, LOCATIVE > COPULA, EQUATIVE, itself "part of a more extended pathway, namely LOCATIVE > EXIST > COPULA" (Heine and Kuteva 2002: 99).

6.2.2. Verbal predicate marker fi

As reported in the literature on GPA (Hobrom 1996: 63-65 and 70, Al-Azraqi 2010: 166 and 171, Bakir 2010: 217, Al-Shurafa 2014: 19, Avram 2014a: 22-23, Bakir 2014: 422-424, Potsdam and Alanazi 2014: 14-16, Avram 2016a: 67), *fi* also occurs in sentences containing main verbs, in pre-verbal position. These *fi* + verb structures express a wide range of meanings and are compatible with a variety of temporal (present, past, future reference) and aspectual (continuous, habitual) values. Therefore, *fi* appears to function as a verbal predicate marker (Avram 2012, 2013)[10]. Note that this use of *fi* is optional (Avram 2012).

[10] Similar generalized predicate markers are attested in other pidgins with various lexifier languages, see e.g. Sieger (2008: 15-16) on Pidgin Fijian *sa*. Note, however, that the status of *fi* is a matter of debate in the literature on GPA: it is considered to be a particle (Al-Azraqi 2010: 169), a predication marker (Bakir 2010: 215-219; Al-Shurafa 2014: 19; Bakir 2014: 433-434) or a copula (Hobrom 1992: 62-66, Potsdam and Alanazi 2014: 28).

(50) a. *ana* ***fe gul*** *inta taal bet* SA (Avram 2014a: 23)
1SG FI say 2SG come house
'I told you to come [to my] place.'
b. ***Fi kalam*** *arabi?* B (Online 2016)
FI speak Arabic
'Do you speak Arabic?'
c. *inta* ***fi yaskit*** Q (Bakir 2010: 217)
2SG FI be silent
'You keep quiet.'
d. *ma* ***fi malum*** UAE (Online 2014)
NEG FI know
'I don't know.'
e. *ana* ***fi šugul*** *hamstašar sana* O (Alshuaimi 2011)
1SG FI work fifteen year
'I've been working for fifteen years.'

This peculiar use of *fi* may reflect substratal/adstratal influence, from languages such as Hindi/Urdu and Persian, which use the verb 'to be' as an auxiliary to form the continuous and the habitual aspect.

On the other hand, it is worth mentioning that similar uses of *fi* have been documented for Arabic FT as well. In one such variety, Kuwait Arabic FT, *fi* + verb structures are even said to occur more frequently than in GPA (Wiswall 2002).

(51) *anta* ***fi*** *fakkar* (Wiswall 2002)
2SG FI think
'you think'

Another variety is Omani Arabic FT, in which *fī* is used "with verbs and verbal nouns" Brockett (1985: 24):

(52) a. *baʕdayn* ***fī šill*** *fir-rās* (Brockett 1985: 24)
then FI take in-DEF head
'then he takes it to the head'
b. *mā* ***fī yīgi?*** (Brockett 1985: 25)
NEG FI 3SG.M-come
'hasn't he come?'

Finally, two grammaticalization chains may also have contributed to the emergence in GPA 0f the verbal predicate marker *fi*: COPULA, LOCATIVE > CONTINUOUS (see Heine and Kuteva 2002: 97) and CONTINUOUS > HABITUAL (Heine and Kuteva 2002: 93). Furthermore, "progressive markers may develop into presents and imperfectives" and "the result is a gram [= grammatical morpheme] of very general meaning" (Bybee et al 1994: 158)[11]. The

[11] Cf. Siegel (2008: 15) on the origin of the predicate marker *sa* in Pidgin Fijian.

GPA verbal predicate marker *fi* appears to be on its way to becoming such a grammatical morpheme, the end result of an extended grammaticalization chain: COPULA, LOCATIVE > CONTINUOUS > HABITUAL > PREDICATE MARKER.

7. The "Feature Pool Hypothesis" and GPA

The so-called "Feature Pool Hypothesis" (see e.g. Mufwene 2001, Aboh and Ansaldo 2006, Aboh 2009, Lim and Ansaldo 2016) is an attempt at explaining the formation of creole languages. According to Mufwene (2001: 4), the so-called "feature pool" is "the "arena" where features associates with the same or similar grammatical functions came to compete with each other", i.e. "while interacting with each other, speakers contribute features to a pool". Aboh and Ansaldo (2006: 44) define the "feature pool" as being "the population of utterances OR features available to speakers in a contact environment". On this view, creole languages "emerge from the recombination of linguistic features from different languages" (Aboh 2009: 317) in the multilingual contact situation. Proponents of the "feature pool hypothesis" contend that the structures of creole languages thus result from a process of competition and selection of features, driven by the following factors: syntax-discourse prominence, markedness/transparency, frequency, salience, and typological (dis)similarity (Mufwene 2001: 57, Aboh and Ansaldo 2006: 44, Aboh 2009, Lim and Ansaldo 2016: 88).

The explanatory adequacy of "feature pool hypothesis" as a mechanism accounting for the genesis of creole languages has been questioned (see a.o. McWhorter 2012, Plag 2013, Lefebvre 2015: 268-271). As put by Parkvall (2015), a legitimate question would be, then, "could it be that the pool theory works better when applied to pidgin languages?". Thus, it is instructive to assess the applicability of the "feature pool hypothesis" to GPA in light of the data discussed in the preceding sections. The factors which are supposed to drive the competition and selection of features can, indeed, account for several characteristics of GPA. For instance, markedness may account for e.g. the loss of the marked phonemes of Arabic. Further, transparency and salience may account for e.g. the non-occurrence of the Arabic definite article: its transparency is obscured by its phonologically conditioned allomorphy, and it is not salient either. Finally, typological dissimilarity may account for e.g. variability in word order patterns: GPA exhibits word order patterns typical of SVO and of SOV languages, which reflect the contribution of such languages to the feature pool.

However, none of the factors said to drive competition and selection can account for the lack in GPA of a number of features, such as plural endings, grammatical gender, any inflections on verbs. Their absence is inexplicable in terms of the "feature pool hypothesis", given that a majority of the languages contributing to the feature fool and, significantly, a majority of the speakers in the contact situation, have these features.

The "feature pool hypothesis" does not include the FT register. However, this has been shown to be the possible sources of GPA features which cannot be accounted for in terms of the "feature pool hypothesis", such as the lack of gender and number agreement, the use of invariant forms of verbs, the use of time adverbials to indicate tense and aspect.

Consider also the potential role of simplificatory processes typical of interlanguages in untutored adult second language acquisition (see e.g. McWhorter 2001, Siegel 2008: 28-30, McWhorter 2012). As shown by e.g. Næss (2008), Almoaily (2008, 2013), Avram (2014a, 2016a, 2016b), the morphosyntax of GPA is characterized by considerable simplification in comparison to that of Gulf Arabic, its lexifier language. The simplification of morphosyntax in GPA is comparable, in many respects, with the so-called "Basic Variety" of immigrant languages in Europe (Klein and Perdue 1997). In their longitudinal study of 40 adult immigrants with Arabic, Finnish, Italian, Spanish, Turkish as L1 who were acquiring Dutch, English, French, German, Swedish, Klein and Perdue (1997) show with respect to morphosyntax that all the learners went through the so-called "Basic Variety" stage and, moreover, that approximately one third did not progress beyond this stage. According to Klein and Perdue (1997), the morphosyntactic features of the "Basic Variety" include (i) the absence of inflections; (ii) categorial multifunctionality; (iii) the use of invariant forms (but also of some inflected forms); (iv) the occurrence of only a few quantifiers; (v) the use of time adverbials to indicate tense and aspect; (vi) the use of a single negator; (vii) the occurrence of a few prepositions; (vii) the absence of complementizers; (viii) the absence of L1 influence, except occasionally for word order patterns. These are also features characteristic of GPA. Therefore, even though no comparable equivalent study of second language acquisition of Arabic by learners with various L1s exists to date, it may be surmised that similar simplificatory processes have contributed to the emergence of the morphosyntactic features of GPA[12].

Inclusion of these two additional factors – i.e. the FT register and simplificatory processes typical of interlanguages in untutored adult second language acquisition – would accord with the simplicity of pidgins in general (see e.g. McWhorter 2001, Parkvall 2008, McWhorter 2012) and with that of GPA. According to Trudgill (2011: 40), "simplification will occur in sociolinguistic contact situations only to the extent that untutored, especially short-term, *adult* second language learning [...] dominates". Trudgill (2011: 101) further states that "simplification is most likely to be found in communities which demonstrate *high contact* (of the post-critical threshold type), *social instability*, and *large size*". These sociolinguistic correlates of simplicity are characteristics of the circumstances under which GPA emerged. The evaluation of the "feature pool hypothesis" leads Plag (2013: 143-144) to the conclusion that the feature pool "may contain variants from all language varieties involved, i.e. from all first languages (superstrate, substrate, adstrate, etc.), all interlanguages (at all levels), and all L1 learner varieties (at all levels)". Except for the latter, all other components appear to have contributed to the emergence of GPA features.

Finally, it is suggested here that incipient grammaticalization may be an additional source of some GPA features.

[12] For a general discussion of commonalities of pidgins and immigrant languages, see Romaine (1989).

8. Conclusions

The structural features of GPA are the outcome the complex interplay of a variety of sources. These include the substrate languages; the superstrate (dialectal Arabic); Arabic FT; incipient grammaticalization. In some cases, GPA features appears to be the outcomes of convergence of several factors, which thereby reinforce one another.

An additional factor contributing to the emergence of GPA features is simplification typical of untutored adult second language acquisition

Summing up, GPA constitutes further proof of the necessity of assuming an "extended" feature pool to account for the features of pidgin languages.

BIBLIOGRAPHICAL REFERENCES

Aboh, E. O. (2009), *Competition and selection: That's all!*, in E. O. Aboh and N. Smith (eds.), *Complex Processes in New Languages*, Amsterdam/Philadelphia, John Benjamins: 31-344

Aboh, E. O., Ansaldo, U. (2006), *The role of typology in language creation: A descriptive take*, in U. Ansaldo, S. Mattews and L. Lim (eds.), *Deconstructing Creole*, Amsterdam/Philadelphia, John Benjamins: 39-66

Al-Ageel, H. (2016), *Requesting Behaviour of Saudi Atabian Women in Contemporary Arabic-Speaking Situations*. Unpublished PhD thesis, Royal Melbourne Institute of Technology

Al-Azraqi, M. (2010), *Pidginization in the Eastern region of Saudi Arabia: Media presentation*, in R. Bassiouney (ed.), *Arabic and the Media. Linguistic Analyses and Applications*, Leiden, Brill: 159-174

Albakrawi, H. (2012), *The linguistic effect of foreign Asian workers on the Arabic Pidgin in Saudi Arabia*, in "Research on Humanities and Social Sciences", 2 (9): 127-133

Alghamdi, E. A. (2014), *Gulf Arabic Pidgin: A descriptive and statistical analysis of stability*, in "International Journal of Linguistics", 6 (6): 110-127

Almoaily, M. (2008), *A Data-based Description of Urdu Pidgin Arabic*, Unpublished MA thesis, Newcastle University

Almoaily, M. (2013), *Language Variation in Gulf Pidgin Arabic*, Unpublished PhD thesis, Newcastle University

Alshuaimi, Suad (2011), *Why call it broken Arabic?*, http://suad47.blogspot.com/2011/07/why-call-it-broken-arabic.html. (20 August 2013)

Al-Shurafa, Nuha S. (2014), *Linguistic functions of fii/maafii in Hejaz Gulf Pidgin Arabic*, in "King Khalid University Journal for Humanities", 21 (1): 11-27

Al-Zubeiry, H. Y. A. (2014), *Linguistic analysis of Saudi pidginized Arabic as produced by Asian foreign expatriates*, in "International Journal of Applied Linguistics & English Literature", 4 (2): 47-53

Avram, A. A. (2010), *An outline of Romanian Pidgin Arabic*, in "Journal of Language Contact", 3: 19-37

Avram, A. A. (2012), *On the functions of fi in the verbal system of Arabic pidgins*, in "Romano-Arabica", XII: 35-58

Avram, A. A. (2013), *Multifunctionality in Gulf Pidgin Arabic: The case of fi*, Paper presented at "Linguistics in the Gulf 4", Qatar University, 17-18 March 2013, Doha

Avram, A. A. (2014a), *Immigrant workers and language formation: Gulf Pidgin Arabic*, in "*Lengua y Migración*", 6 (2): 7-40

Avram, A. A. (2014b), *On the relationship between Arabic Foreigner Talk and Pidgin Arabic*, Paper presented "ArCo 2014 – Arabic in Contact: Linguistic and Sociolinguistic Perspectives", Università degli Studi "L'Orientale", 15-17 December 2014, Naples

Avram, A. A. (2016a), *On the periphery of the periphery: Gulf Pidgin Arabic*, in M. F. Al Hamad, R. Ahmad, H. I. Alaoui (eds.), *Lisān al-Arab: Studies in Contemporary Arabic Dialects. Proceedings of the 10th International Conference of AIDA, Qatar University, 2013*, Vienna, LIT Verlag, Neue Beihefte zur Wiener Zeitschrift für die Kunde des Morgenlandes: 61-74

Avram, A. A. (2016b), *On the developmental stage of Gulf Pidgin Arabic*, in G. Grigore and G. Biţună (eds.), *Arabic Varieties: Far and Wide. Proceedings of the 11th International Conference of AIDA, Bucharest, 2015*, Bucharest, Editura Universităţii din Bucureşti: 87-97

Bakir, M. J. (2010), *Notes on the verbal system of Gulf Pidgin Arabic*, in "Journal of Pidgin and Creole Languages", 25 (2): 201-228

Bakir, M. J. (2014), *The multifunctionality of fii in Gulf Pidgin Arabic*, in "Journal of Pidgin and Creole Languages", 29 (2): 416-442

Behnstedt, P. and Woidich, M. (2011), *Wortatlas der arabischen Dialekte*, vol. I, *Mensch, Natur, Fauna und Flora*, Leiden · Boston, Brill

den Besten, H, Muysken, P. and Smith, N. (1995), Theories focusing on the European input. In J. Arends, P. Muysken and N. Smith (eds.), *Pidgins and creoles. An introduction*, Amsterdam, John Benjamins: 87-98

Bizri, F. (2010), *Pidgin Madame. Une grammaire de la servitude*, Paris, Geuthner

Bybee, J. L., Perkins, R. and Pagliuca, W. (1994), *The Evolution of Grammar. Tense, Aspect, and Modality in the Languages of the World*, Chicago, Chicago University Press

Dashti, A. A. (2013), *Interacting with domestic workers in Kuwait: Grammatical features of Foreigner Talk. A case study*, in "International Journal of the Sociology of Language": 224: 63-84

Ferguson, C. A. (1971/1996), *Absence of copula and the notion of simplicity: A study of normal speech, Baby Talk, Foreigner Talk and pidgins*, in T. Huebner (ed.), *Charles A. Ferguson, Sociolinguistic Perspectives. Papers on Language in Society, 1959 – 1994*, New York, Oxford University Press: 115-123

Ferguson, C. A. and DeBose, C. E. (1977), *Simplified registers, broken language, and pidginization*, in A. Valdman (ed.), *Pidgin and Creole Linguistics*, Bloomington, Indiana University Press: 99-125

Heine, B. and Kuteva, T. (2002), *World Lexicon of Grammaticalization*, Cambridge, Cambridge University Press

Hobrom, A. (1996), *An Analysis of the Arabic Pidgin Spoken by Indian Workers in Saudi Arabia*, Unpublished MA thesis, Kansas University

Holes, C. (2007), *Kuwaiti Arabic*, in K. Versteegh, M. Eid, A. Elgibali, M. Woidich, A. Zaborski (eds.), *Encyclopedia of Arabic Language and Linguistics*, vol. II, Leiden, Brill: 608-620

Klein, W. and Perdue, C. (1997), *The Basic Variety (or: Couldn't natural languages be much simpler?)*, in "Second Language Research", 13: 301-347

Levebvre, C. (2015), *Relabeling in Language Genesis*, New York, Oxford University Press

Lim, L. and Ansaldo, U. (2016), *Languages in Contact*, Cambridge, Cambridge University Press

McWhorter, J. (2001), *The rest of the story: Restoring pidginization to creole genesis theory*, in "Journal of Pidgin and Creole Languages", 17: 1-48

McWhorter, J. (2012), *Case closed? Testing the feature pool hypothesis*, in "Journal of Pidgin and Creole Languages", 27 (1): 171-182

Mufwene, S. S. (2001), *The Ecology of Language Evolution*, Cambridge, Cambridge University Press.

Mühlhäusler, P. (1997), *Pidgin and Creole Linguistics*, expanded and revised edition. London, University of Westminster Press

Næss, U. G. (2008), *"Gulf Pidgin Arabic": Individual Strategies or a Structured Variety? A Study of Some Features of the Linguistic Behaviour of Asian Migrants in the Gulf Countries*, Unpublished MA thesis, University of Oslo

Parkvall, M. (2008), *The simplicity of creoles in a cross-linguistic perspective*, in M. Miestamo, K. Sinnemäki and F. Karlsson (eds.), *Language Complexity: Typology, Contact, Change*, Amsterdam/Philadelphia, John Benjamins: 265-285

Parkvall, M. (2015), *The feature pool in pidginisation*, Paper presented at the Tenth Creolistics Workshop "Innovations" – with Special Attention to Parallels between Creole and Sign Languages", Aarhus University, 8-10 April 2015, Aarhus

Plag. I. (2013), *Creolization and admixture: Typology, feature pools, and second language acquisition*, in P. Bhatt and T. Veenstra (eds.), *Creole Languages and Linguistic Typology*, Amsterdam/Philadelphia, John Benjamins: 141-161

Potsdam, E. and Alanazi, M. (2014), *Fi in Gulf Pidgin Arabic*, in "Kansas Working Papers in Linguistics", 35: 9-29

Romaine, S. (1989), *Pidgins, creoles, immigrant, and dying languages*, in N. C. Dorian (ed.), *Investigating Obsolescence. Studies in Language Contraction and Death,* Cambridge, Cambridge University Press: 369-383

Salem, A. A. M. S. (2013), *Linguistic features of Pidgin Arabic in Kuwait*, in "English Language Teaching", 6 (5): 105-110

Siegel, J. (2008), *The Emergence of Pidgin and Creole Languages*, Oxford, Oxford University Press.

Smart, J. R. (1990), *Pidginization in Gulf Arabic: A first report*, in "Anthropological Linguistics", 32 (1-2): 83-119

Tosco, M. and Manfredi, S. (2013), *Pidgins and creoles*, in J. Owens (ed.), *The Oxford Handbook of Arabic Linguistics*, Oxford, Oxford University Press: 495-519

Trudgill, P. (2011), *Sociolinguistic Typology. Social Determinants of Linguistic Complexity*, Oxford, Oxford University Press

Wiswall, A.-Q. (2002), *Gulf Pidgin: An expanded analysis*, http://oak.cats.ohiou.edu/~aw321500. (1 December 2002)

Yammahi, G. S. A. (2008), *Al-ʕarabiyyatu al-ʔurdiyyatu al-haǧīn fī al-ʔimārāti al-ʕarabiyyati al-muttaḥidati. Maṭalun min madīnati Dubayy,* Unpublished MA thesis, Amman, University of Jordan

ANDREI A. AVRAM • Is Professor of English Linguistics at the University of Bucharest, Romania. He holds a PhD in linguistics from the "Iorgu Iordan – Al. Rosetti" Institute of Linguistics of the Romanian Academy, Bucharest (2000), and a PhD in linguistics from Lancaster University (2004). His research areas are pidgins and creoles with various lexifiers, language contacts, and phonology. Publications include *Pidginurile şi creolele cu bază engleză şi franceză ca tip particular de contact lingvistic* (2000), *On the Syllable Structure of English Pidgins and Creoles* (2005), *Fonologia limbii japoneze contemporane* (2005), book chapters on pidgins and creoles as well as Maltese in edited volumes, and articles in *English World-Wide, English Today, Journal of Language Contact, Cochlear Implants International, Lengua y migración, Études Créoles, Linguistics in the Netherlands, Mediterranean Language Review, Linguistik Online, Papia: Revista Brasileira de Estudos do Contato Linguístico, Acta Linguistica Asiatica, Linguistica Atlantica, Romano-Arabica, Revue roumaine de linguistique*.

E-MAIL • andrei.avram@lls.unibuc.ro

AGREEMENT WITH PLURAL CONTROLLERS IN OMANI ARABIC

Preliminary remarks

Simone Bettega

ABSTRACT • In the course of this paper, I examine the factors responsible for variation between singular and plural agreement with plural controllers in Omani Arabic. Particular focus is put on non-human plural controllers, since this is where the most variation is found. Both controller-related and target-related factors are taken into account. Animacy, specificity and morphological structure of the controller are found to be the factors having the greatest impact on the kind of agreement which obtains, along with target type (consistently with the predictions of Corbett's Agreement Hierarchy) and distance between target and controller.

KEYWORDS • Oman, Agreement, Plural, Arabic dialectology

1. Introduction

Agreement in spoken Arabic has repeatedly proven to be a complex subject for linguists to describe. A wealth of literature exists on the subject (to name but a few relevant works: Belnap 1999 on Cairene Arabic; Hallman 2000 on Lebanese Arabic; Holes 2016: 326-53 on Bahraini Arabic, Prochazka & Gabsi 2016 on Northern Urban Tunisian Arabic), highlighting the presence of great inter- and intra-dialectal variation. While the former may arise as a consequence of the structural differences existing between the various Arabic dialects[1], the latter appears to be the consequence of several overlapping factors, including word order, animacy of the controller, and others (Belnap, 1991; Brustad, 2000).

[1] In Southern Bedouin Tunisian, for instance, gender distinction in the plural is preserved, while it has collapsed into a set of generic plural forms in the urban North (Ritt-Benmimoun, 2016). Even dialects that display the same set of morphological markers of gender and number,

In the course of this article, I will focus specifically on agreement patterns with plural controllers in modern Omani Arabic (OA). OA is a dialect (or, better yet, a bundle of dialects) which up to this day has received limited attention from scholars. This is regrettable because OA shows a number of features which, from a dialectological perspective, appear to be rare, if not unique[2]. In particular – and of specific interest to us here – all Omani varieties have preserved gender distinction in the plural forms of the adjective, verb and pronoun (Holes 1989: 448-9). This is an uncommon trait as Arabic dialects go: in particular, the only existing study dealing with agreement patterns in a dialect which has preserved this distinction is Ritt-Benmimoun's recent work on Bedouin Tunisian Arabic (see Ritt-Benmimoun, 2016). A second, partial exception is represented by Owens & Bani-Yasin (1987), which investigates agreement variation in a rural Jordanian dialect: Owens' and Bani-Yasin's research, however, is sociolinguistic in nature, dealing with the influence that MSA agreement rules may have on the type of agreement we observe in the colloquial varieties. I have extensively commented on Owens' and Bani-Yasin's work elsewhere – highlighting the insights it provides as well as its shortcomings – and I will not repeat myself here. The interested reader is referred to Bettega (in press): in that article, which is intended as complementary to the present one, I offer a more in-depth analysis of the question of agreement in OA from the perspective of socio- and historical linguistics. In the course of this paper, on the contrary, I will address the topic from a purely synchronic perspective, without taking into account the potential effects of contact-induced variation and change over time. In particular, what I aim at investigating here is the cause of variation between singular and plural agreement with plural controllers[3].

In the next paragraph, I will briefly describe the database I have used for the present research, along with the methodological tools I have employed for the analysis of the material; in the third paragraph, I will present the results of the study.

---

however, may differ wildly in terms of agreement behavior: it is the case, for instance, of Cairene and Bahraini Arabic.

[2] For an overview of the typological characteristics of OA, see Holes (1989).

[3] When non-human controllers are involved, variation between singular and plural agreement is actually variation between feminine singular and feminine plural agreement (that is, targets depending on controllers which refer to non-human entities tend to show feminine agreement). Admittedly, gender variation can sometimes occur with non-human controllers as well, so that we have masculine plural and masculine singular agreement with inanimate referents. In the first case, this is probably the result of the influence of more prestigious varieties (in which masculine agreement is the norm for all plural nouns, see Bettega, in press). The (statistically uncommon) occurrence of masculine singular agreement with plural controllers has probably to do with word order and the process of grammaticalization that certain targets are currently undergoing in OA. Some remarks on this last point can be found in paragraph 3.6, but in general this article is not concerned with the issue of gender variation.

2. Data and Methodology

The present research is based on the analysis of five hours of audio material drawn from a popular Omani TV-show called *Yōm u-yōm* (يوم ويوم). *Yōm u-yōm* is a computer-animated sit-com revolving around the daily lives of a number of characters living in Muscat, who face farcical and sometimes surreal adventures. In particular, the whole second season of the show was examined for the purpose of this article: the season was aired during the month of Ramadan 2012, and it consists of 25 episodes lasting between 10 and 15 minutes each.

The analysis of this material yielded a corpus consisting of 180 plural controllers[4], with 269 corresponding targets (for the terms "target" and "controller" see Corbett, 2006). This corpus was analyzed by adapting the methodology suggested by Belnap (1991, 1993, 1999) and Brustad (2000). In particular, a number of morpho-syntactic, semantic and pragmatic factors were taken into account, which have been suggested to have an impact on the kind of agreement which obtains. Factors related to the controller will be discussed in paragraph 2.1, while factors related to the target will be discussed in paragraph 2.2.

2.1. Controller-dependent factors

The following factors have been considered when analyzing the controllers which appear in my data:

i) *Head type:* all plural nouns in Arabic can be classified as either sound or broken, depending on their morphological status. So-called "sound" plurals are those which are derived from the singular by simple suffixation, while "broken" plurals result from the alteration of the internal vocalic structure of the word. This morphological subdivision can be combined with the semantic category of animacy (that is, whether the controller is human or non-human). The combination of these two features, one morphological and one semantic, gives rise to an overarching category I will refer to as "head type" (following Belnap), according to which it is possible to classify plural nouns as human sound, human broken, non-human sound or non-human broken.

[4] This figure includes 5 dual controllers and 11 chains of conjoined nouns. As Blanc (1970: 43, 49) notes, "the dialectal dual is not a grammatical concord category [...]. Dual nouns take plural concord". As far as chains of conjoined nouns are concerned, these were counted as a single token each. On chains of conjoined controllers, see paragraph 3.3. It has to be noted that only explicit controllers have been included in the corpus: in other words, targets referring to an implicit, unmentioned third person plural controller were left out of the present analysis. This is because it would have been impossible to determine what was the exact nature of the controlling element the speaker had in mind. For the same reason, I decided to leave controllers consisting in a single demonstrative out of my corpus as well.

ii) *Definiteness:* whether the controlling noun is definite or not, either by the presence of an explicit element such as the definite article, or because it occupies a specific syntactic slot that makes it inherently definite (e.g. first element within a synthetic possessive construction).
iii) *Qualification:* whether the controlling noun is qualified or not (for instance by an adjective, a relative clause, or a possessive element).
iv) *Concreteness*: whether the controller refers to an abstract or concrete element[5].
v) *Specificity*: unlike the other factors, specificity cannot be associated with any particular formal marker, nor it is easily described in terms of a binary opposition. Brustad (2000: 24) defines specificity as "the extent to which the speaker has a specific entity in mind". I followed Belnap's (1991: 76) lead in coding each head of the corpus as either specific, non-specific or unclear. My labelling must, of course, be regarded as tentative, although in most cases the context of the utterance made it possible to determine with reasonable accuracy whether the speaker was referring to a specific plural entity or not[6].

[5] Concreteness was defined as the property of the controller of referring to tangible, material entities. Admittedly, some ambiguity can sometimes arise with respect to this category: it is the case, for instance, of the word *aflām* "films", which I classified as abstract throughout the corpus, with the exception of one occurrence where it was used to refer to the physical copies of the DVDs one can rent from a store. In general, though, such ambiguous occurrences were few, and analysis of the context made disambiguation possible.

[6] Although for ease of analysis I have decided to treat them as separate categories, it would seem that specificity, concreteness and quantification are to be considered mutually connected, and capable of influencing the speaker's perception of the plural as either scattered or grouped. Belnap (1991: 74-5) builds on Barlow's observation that in certain languages agreement may function as a secondary marker or classifier, which further indicates the perception of referents as a group or as individual entities. Barlow (1988: 306-7) proposes a list of features which are likely to highlight the individuality of members in a group, namely: high agency, animacy, familiarity, and clear definition of individual entities within the group (i.e. boundaries between members are visible). The validity of Barlow's hypothesis for spoken Arabic has been repeatedly confirmed by several authors working on different varieties: see for instance Holes (2016) for Bahraini, Prochazka & Gabsi (2016) and Ritt-Benmioun (2016) for Tunisian, D'Anna (forthcoming) for Fezzani, and Cowell (1964) for Damascus Arabic. Cowell (1964: 423), in particular, observes how "most inanimate plurals, and some animate plurals and collectives, have feminine [singular] agreement in the predicate when collectivity or generality is emphasized rather than heterogeneity or particularity". According to Belnap, the notion of grouped/non-grouped (i.e. scattered) plurals can be related to the pragmatic notion of specificity: in his view "more specific referents would tend to be seen as individuated. Less specific referents are more likely to be perceived as grouped". Belnap notes, also, how certain controllers may be perceived as inherently less specific than others, regardless of their actual context of use. It is the case, for instance, of nouns with concrete referents (as opposed to abstract ones), which according to Belnap are more likely to be perceived as individuated.

vi) *Quantification:* whether the controlling noun is quantified or not, either by a numeral, the dual ending, or another quantifying element such as *kull* "all", *šwayy* "a little", *kṯīr* "a lot", and so on.

2.2. Target-dependent factors

The following factors have been taken into account when analyzing the controllers which appear in my data:

i) *Target type:* whether the target is a demonstrative, an adjective, a verb, an anaphoric pronoun or a preposition expressing possession.
ii) *Distance between target and controller*: expressed in terms of phonological words, from a minimum of 1.
iii) *Word Order*: whether the target precedes or follows its controller.

3. Analysis

I will now present the results of the analysis relative to each one of the factors introduced in the previous paragraph. Paragraph 3.1 deals with the category of head type, while paragraph 3.2 with the remaining factors related to the controller. In Paragraph 3.3 I will briefly address the question of chains of conjoined controllers as a particular type of controller. Paragraphs 3.4, 3.5 and 3.6 will deal with target-dependent factors (target type, distance between target and controller and word order respectively).

3.1. Controller-dependent factors: head type

In his study on Cairene colloquial, Belnap (1993: 102) found the type of the head to be the factor having the greatest effect on the kind of agreement which obtains. My data shows that this prediction is generally valid for OA as well. Table 1 illustrates the kind of agreement triggered by five different kinds of controllers, namely: inanimate broken, inanimate sound, human broken, human sound and the collective *nās*, 'people' (I follow Belnap in analyzing *nās* separately from other controllers; concerning *nās*, see in particular paragraph 3.2.1 below)[7]. Out of 180 controllers constituting my corpus, 23

[7] Two animate non-human controllers appear in my data, with four corresponding targets. Since these figures are too small to be of any statistical relevance, I decided not to distinguish between animate and inanimate non-human controllers. Belnap (1991: 62), working on a corpus more than three times bigger than mine, lamented such paucity as well, finding only nine controllers denoting animals among his data. He noted that these controllers "patterned between human and inanimate, suggesting an animacy hierarchy". Although the natural low frequency of this

were not classifiable according to the broken/sound parameter[8], and therefore do not appear in the table. The remaining 157 were associated with 233 targets, subdivided as follows:

	Type of agreement		
	Singular	Plural	Total
Targets depending on non-human broken controllers	59 (69,4%)	26 (30,6%)	85
Targets depending on non-human sound controllers	46 (68,7%)	21 (31,3%)	67
Targets depending on the collective *nās*	5 (31,3%)	11 (68,7%)	16
Targets depending on human broken controllers	1 (2%)	48 (98%)	49
Targets depending on human sound controllers	0 (0 %)	16 (100%)	16

Table 1: Head Type

As can be seen, human sound controllers categorically require plural agreement. Plural agreement is also nearly categorical with human broken controllers, with only one exception appearing in the corpus, shown in example (1): here a bus driver is complaining about his job, explaining how he always has to wait for passengers who get on the bus at the very last moment. Thus, the word *awādəm*, 'people, persons', in this sentence has a very generic reference, since the driver has no definite person or group of persons in mind:

(1) *əl-awādəm* *illi* *məttaxxar-a* *ʕalā* *d-dwām*
ART=person.PL REL be.late-SG.F PREP ART=work
'The people who are late for work'.

My findings in this respect are consistent with what Belnap (1993: 100-1) reports about Cairene Arabic, the only difference consisting in his percentage of human broken controllers triggering singular agreement being higher than mine (11%). It is possible, however, that the analysis of a larger corpus of spontaneously occurring speech would

kind of controllers in the speech of urban populations should come as no surprise, Belnap's remark is interesting and surely worth of further investigation: interviews with native speakers explicitly aimed at soliciting the occurrence of such controllers (concerning topics such as pastoral and agricultural life) might provide more data to shed light on the subject.

[8] Because they were either dual, quantified by a numeral higher than ten, or consisting in a chain of conjoined nouns (see paragraph 3.3). Note that in OA, as in other varieties of Arabic, nouns quantified by a numeral higher than ten appear in their singular form.

yield more varied results for OA as well[9]. Human sound controllers, as already mentioned, systematically trigger strict agreement in both mine and in Belnap's corpus: examples (2) and (3) are representative of this tendency:

(2) *əl-mumarriḍ-āt* *fēn-hən*
ART=nurse-PL.F where=PRON.3PL.F
'The nurses, where are they?'.

(3) *əl-masūl-īn* *mā* *yi-xabbar-u* *ḥadd*
ART=supervisor-PL.M NEG PRES.3-inform-PL.M anyone
'The supervisors do not inform anyone'.

Non-human controllers, on the contrary, appear to favor singular agreement, though not as strongly as human ones favor plural agreement. The collective *nās* seems to pattern in between, although with a marked preference for plural agreement (this will receive separate attention in paragraph 3.2.1). Here as well my results appear to match those of Belnap, although with one important difference: inanimate controllers in Belnap's corpus of Cairene speech show a much more pronounced preference for singular agreement (91% of the targets depending on inanimate sound controllers and 92% of the targets depending on inanimate broken controllers) than those in my corpus do. As pointed out in note 9, this might be due to the inherently different nature of the agreement systems in Cairene and OA. I have examined this hypothesis in more detail in Bettega (in press), and in the present context I will limit myself to note how, although non-human controllers tend to favor singular agreement, a significant portion attracts plural agreement instead (specifically, 47 targets out of a total 152, which is to say 30,9% of all targets depending on non-human controllers, both broken and sound). The question obviously arises as to what the reasons behind such variation are: in the course of the next paragraphs I will focus specifically on the kind of agreement triggered by non-human controllers and the factors governing it.

[9] Although it is possible that the somehow more formal level of my texts determined the absence of human controllers attracting deflected agreement, another explanation might reside in the fact that, while in Cairene Arabic the agreement system is binary in nature (that is, any plural head may either trigger feminine singular or generic plural agreement) the Omani one is tripartite (a plural head being able to trigger feminine singular, masculine plural or feminine plural agreement). The balance of two such systems is obviously different, and this might have a consequence in affecting the kind of agreement which can or cannot occur in a given context. More research on this point is obviously needed.

3.2. Controller-dependent factors: definiteness, qualification, concreteness, specificity and quantification

Since, as shown in table 1, the distinction between sound and broken non-human controllers does not appear relevant in determining the kind of agreement they attract, let us now consider the corpus of data in terms of the sole distinction human/inanimate[10]. This allows us to reintroduce in the analysis those controllers which did not fit a division based on the sound/broken criterion. Of the 178 heads present in my corpus, 127 (71,3%) refer to inanimate entities: as can be seen in Table 2, 172 (64,9%) targets out of 265 are dependent on those heads.

Targets depending on human controllers		Targets depending on inanimate controllers	
Type of agreement	Occurrences	Type of agreement	Occurrences
Singular	6 (6,5%)	Singular	111 (64,5%)
Plural	87 (93,5%)	Plural	61 (35,5%)
Total	93	Total	172

Table 2: Agreement with human and inanimate controllers

As already discussed, plural human controllers in OA almost categorically require plural agreement (the only exceptions to this rule are 5 occurrences of singular agreement triggered by the collective *nās*, 'people', and the one case of singular agreement with a broken human controller discussed in the previous paragraph).

As far as inanimate controllers are concerned, as we have seen, the situation is not as clear-cut as it is for human ones. Although there exists a tendency for non-human controllers to attract singular agreement, around one third (35,5%) of their total targets carry a morphological (mostly feminine, as pointed out in note 3) plural marker instead. To determine the reasons behind this variation, let us now consider the factors from *ii* to *vi* listed in paragraph 2.1[11].

[10] The two controllers denoting animate, non-human referents referred to in note 7 have not been included in the following analysis.

[11] Note that in tables 3 and 4 the percentages are calculated on the number of controllers triggering singular or plural agreement in at least one of the targets with which they are associated, rather than on the number of targets bearing singular or plural agreement. This was done in order to prevent controllers associated with a high number of targets to have a greater statistical weight than controllers associated with only one target. In addition, the number of total controllers under analysis in table 4 is slightly lower, because 4 heads were qualified as unclear with respect to the category of specificity and therefore excluded from the total.

	Out of 82 inanimate controllers attracting singular agreement	Out of 45 inanimate controllers attracting plural agreement
Definite controllers	59 (72%)	36 (80 %)
Qualified controllers	45 (54,9%)	18 (40%)
Concrete controllers	34 (41,5 %)	32 (71,1 %)
Quantified controllers	16 (19,5%)	16 (35,6%)

Table 3: Quantification, Qualification, Definiteness, Concreteness

	Out of 80 inanimate controllers attracting singular agreement	Out of 43 inanimate controllers attracting plural agreement
Specific controllers	22 (32,5 %)	23 (53,5 %)

Table 4: Specificity

As can be seen from the tables, not all variables seem to affect agreement in the same way: in fact, definiteness appears to have little impact, if any, only slightly favoring plural agreement (8 percentage points). This might seem to go against expectations, but it must be considered that the majority of abstract nouns are usually definite in Arabic: in other words, the category of definiteness could be subordinated to that of concreteness, and is therefore not clear whether it actually plays a role in influencing agreement.

Qualification, on the other hand, seems to favor singular agreement, since formally qualified controllers have almost 15% more chances of attracting singular agreement. Again, this might seem counterintuitive, but – as we will see in paragraphs 3.4 and 3.5 – among all target types attributive adjectives are those most likely to trigger deflected agreement. This seems to imply that qualification is subordinate to other factors (namely, target type and distance between target and controller) when it comes to determining the kind of agreement which obtains (this, in turn, would confirm Belnap's prediction that distance between target and controller is one of the most important factors influencing agreement).

The following are examples of definite and/or qualified controllers attracting deflected agreement:

(4) *ayyām-ak* *ṣār-at* *maʃdūd-a*
day.PL=PRON.2SG.M become-PAST.3SG.F numbered-SG.F
'Your days are numbered'.

(5) *a-bā* *a-rakkib* *hāḏēk* *ər-rang-āt* *sbōrt*
PRES.1SG-want PRES.1SG-ride DIST.SG.F ART=range.rover-PL.F sport
'I want to drive those Range Rovers Sport'.

(6) *ḥattā* *ʕyūn-hum* *muġlān-a*
until eye.PL=PRON.3PL.M boiled-SG.F
'Until their eyes are reddened'.

A correlation seems to exist, on the contrary, between the two categories of specificity and concreteness and agreement (though not as strong as that existing between agreement and head type): specific controllers have 21% more chances of triggering plural agreement[12], and concrete controllers have 29,6% more chances than abstract ones of attracting plural agreement. Some examples are provided below of concrete and specific controllers triggering strict agreement. Compare in particular the two occurrences of the word *ašyā*, 'things' in (8) and (9), and the different kind of agreement they attract. The first one refers back to a list of very specific items (a lamp, an air conditioner and a washing machine) which are broken and which the speaker wishes her husband to fix. The second one, on the contrary, refers to the food the speaker is supposed to offer to the people who are gathered in his house for a funeral (here no list is involved, the reference being to the types of food that are customarily presented as an offer to the guests on such an occasion: everybody knows what they are, so there's no need of any specific mention):

(7) *a-šūf-iš* *lābsa* *s-suwār-āt*
PRES.1SG-see=PRON.2SG.F wear.PART-SG.F ART=bracelet-PL.F
illi *štarē-tī-hən* *min-ni* *bas*
REL buy-PAST.2SG.F=PRON.3PL.F PREP=PRON.1SG but
t-ṣaddg-ī *ṭalʕ-at* *yi-habl-an*
PRES.2SG-believe-F turn out-PAST.3SG.F PRES.3-look.good-PL.F
ʕalē-š
PREP=PRON.2SG.F
'I see you're wearing the bracelets you bought from me, but you know what? It turns out they look good on you'.

[12] This is consistent with Belnap's results, which show a predilection for specific heads to attract plural agreement, although not as strong as it appears in my data. Working on Cairene Arabic, Belnap tried to bypass the problems connected with the inherent ambiguity of the category of specificity by administering his informants a psycholinguistic questionnaire. The results of the experiment also seem to endorse the hypothesis of a connection between agreement and specificity. Holes (2016), Prochazka & Gabsi (2016), Ritt-Benmimoun (2016) and D'Anna (forthcoming) all seem to confirm this connection. On specificity see also paragraph 3.2.1 below.

(8) *ašyā* *hāḏēla* *muxtarb-āt*
thing.PL DEM.PL broken-PL.F
'These things [are] broken'.

(9) *min* *wēn* *a-yīb* *lə-flūs* *ḥal* *ha-l-ašyā*
PREP where PRES.1SG-take ART=money PREP DEM=ART=thing.PL
kəll-ha
all=PRON.3SG.F
'From where do I take the money for all those things?'.

The last factor considered in table 3 is quantification. It would seem that quantified controllers tend to slightly favor plural agreement (16,1% more chances of attracting plural agreement than non-quantified controllers), but at a closer inspection the patterns turn out to be more complex. If we set controllers qualified by the element *kəll*, 'all', apart from those quantified by an explicit numeral (including dual nouns), we see that these two groups show opposite tendencies. Nouns quantified by *kəll* appear to show a preference for singular agreement (with 9 controllers out of 15 attracting singular agreement, that is, 60%). Dual nouns and nouns quantified by a specific numeral, on the other hand, show the exact opposite tendency (with 9 controllers out of 15 triggering plural agreement). It has to be noted that, of the 6 nouns quantified by a numeral which trigger singular agreement, two are quantified by a numeral higher than ten[13], one is accompanied by the very generic quantifier *darzān* "a dozen", and one is an occurrence of the word *ayyām*, 'days', which in my data systematically triggered singular agreement in all contexts (probably because it is inherently abstract and non-specific; Owens & Bani-Yasin, 1987, report the same for the rural Jordanian dialect they investigate; see also paragraph 3.2.1 below).

In conclusion, though the number of quantified heads in the sample is too small to identify any strong statistical trend, it would seem that different types of quantifiers tend to attract different types of agreement. To further substantiate this hypothesis, I present here the transcription of two WhatsApp voice notes in which two friends are discussing some work-related issues[14]. The targets and controllers appearing in these texts are obviously not part of the corpus of data that I have analyzed for the present research, since this is limited to material drawn from the *Yōm u-yōm* tv show: however, I decided to include them here because of their relevance to the point in question. The texts are rather long, so I have not inserted interlinear glosses: the relevant plural controllers and their targets are highlighted in boldface:

[13] Such numerals in the literature have been associated with non-specific reference and low individuation, probably because it is more difficult for the brain to keep track of large numbers of individual entities, being easier to perceive them as an indistinct group. See for instance Brustad (2000: 23). Belnap as well has documented a strong correlation between numerical quantification and plural agreement in Cairene Arabic (see Belnap, 1991, and Belnap, 1993).

[14] I am indebted to my friend and teacher Ahmad Al-Suleimaini for sharing these texts with me, offering precious comments and helping in innumerable other ways with my work.

əl-yōm saggalna **sitt nuṣūṣ** taqrīban fa-**hadēla s-sitta** əlli **hənna mumtāzāt** yaʕnī fa-bi-t-tālī ana kənt afakkər yaʕnī kān bāqī **arbaʕa** bə-nsaggil-**hən** b-ṣōt-ī u-arsil-l-ak iyya-**hən** əl-gimʕa bas ənta tqūl mā aʕrif matā bāġi-n-**hən** əl-yōm willa bākir fa-bi-t-tālī axallaṣ-l-ak iyya-**hən** əl-lēle inšāḷḷah wa-arsil-l-ak iyya-**hum**

'Today we recorded **six texts** more or less, and **these [pl.] six**, **they [pl.] are really good [pl.]**, I mean, and so, I was thinking, I mean, that there were **four** left, and that we could record **them [pl.]** with my voice and send **them [pl.]** to you on Friday, but [now] you tell me you want **them [pl.]**, I don't know, today or tomorrow, and so, I will finish **them [pl.]** for you tonight if God wills, and I will send **them [pl.]** to you'

As can be seen, the controllers here are numbered by a numeral lower than eleven (six in the first case, then four), and have highly specific reference (the two friends are discussing the dubbing of certain excerpts of a documentary they are working on together). All agreement targets are plural. The conversation then continues with the two speakers discussing the number and date of delivery of the texts: in the end, the same speaker from the example above formulates a new proposal. The referent (the recordings) has remained the same, but here they are being referred to as a plurality rather than a set of separate entities (as demonstrated by the use of the quantifiers *noṣṣ*, 'half' and *kəll*, 'all'): the agreement, accordingly, shifts from plural to singular:

nsaggil **noṣṣ fi-l-āxar min ən-nuṣūṣ** u-b-narsil-l-ak iyya-**hā** inšāḷḷah baʕd ḏālik **ət-tasgīlāt** kəll-**ha** mā **tāxud** waqt inšāḷḷah bas **tasgīlāt xafīfa**

'We will record **half of the remaining texts**, and we will send **them [sg.]** to you if God wills, after that the recordings, all of **them [sg.]** will not **take [sg.]** time if God wills, [they're] just **short [sg.] recordings**'

3.2.1. Controller-dependent factors: some remarks on lexical specificity

As far as the category of specificity is concerned, it has to be noted that certain lexical items are perceived as inherently less specific than others, regardless of their actual context of use. It is the case, as we have already seen, of nouns with concrete reference (as opposed to abstract ones, which are perceived as having more generic reference), but it is also the case with hypernyms. For instance, Belnap (1991: 77) notes how 'women' is inherently less specific than 'grandmothers': "a given token of 'women' or 'grandmothers' may be used specifically or generically, but the core meaning of 'women' still remains more generic than that of 'grandmothers'".

If, as we have seen, specificity is connected with agreement, it follows that inherently less specific nouns should attract singular agreement more often than specific ones. As Belnap (1993: 104) notes, in Arabic "there is no more generic term used to refer to human beings than *nās*, 'people'". As we have already seen, the collective *nās* is the only human controller in our corpus to occasionally attract singular agreement (along with one occurrence of *awādəm*, 'persons', see example (1) above): out of 10 occurrences of *nās*, 5 attract singular agreement (for a total of 5 targets), and 5 attract plural agreement (for a total of 11 targets). It would seem that this variation, again, has to do

with the level of specificity denoting the referent in the mind of the speaker (i.e., even a generic term such as *nās* can be used more specifically or more generically). Compare, for instance, examples (9), (10) and (11) below: in the former, the dialogue takes places between two friends who are both attending a funeral, and the reference is to the people who are presently gathered in the house of one of the two (therefore, to a specific group of persons, presently visible to both the speaker and the hearer). In (10), on the contrary, the speaker is referring to people in the most generic way possible, since he is offering population growth as an explanation for the exceptional levels of traffic his friend Mḥād is complaining about. In (11) the reference is again generic, denoting random Facebook users who will flock to see the picture in question:

(10) *ən-nās* *b-yi-rūḥ-ūn* *w-enta* *min*
ART=people CONT=PRES.3M-go-PL CONJ=PRON.2S.M PREP
əṣ-ṣubḥ *ḥāṭṭ* *ən-nās* *ṣaḥan* *māl* *tamar*
ART=morning put.AP ART=people dish GEN date.COLL
'The people are going away, and since this morning you have given the people [only] a dishful of dates?'.

(11) *ṣadaq-t* *yā* *mḥād* *u-trā* *n-nās*
be right-PAST.2S.M VOC mḥād CONJ=actually ART=people
t-zīd
PRES.3SG.F-increase
'You're right Mḥād, and there's actually more people!' (lit. 'actually people increase').

(12) *t-ḥaṭṭ* *əṣ-ṣūr-a* *u-šūf*
PRES.2SG.M-put ART=picture-SG.F CONJ=see.IMP.2SG.M
ən-nās *kēf* *bā-t-gī*
ART=people how FUT=3SG.F-come
'You put the picture, and see the people how they will come!'.

As far as non-human controllers are concerned, it is also worth noting that items which recur with high frequency in the texts, and whose referents are inherently generic, systematically attract deflected agreement throughout the corpus: it is the case for instance of *ayyām,* 'days', and *umūr*, 'matters, affairs'. In example (13), we see *ayyām* triggering singular agreement even though it is quantified by a numeral lower than ten. In example (14) an occurrence of *umūr* attracts singular agreement:

(13) *a-ʃtaqid* *yōm-ēn* *u-mā*
PRES.1SG-reckon day-DUAL CONJ=NEG
b-yi-zīd *ʃan* *ṯlāṯat* *ayyām* *il-yāy-a*
FUT=PRES.3SG.M-increase PREP three day.PL ART=come.PART-SG.F
'Two or three days, I reckon' (lit. 'I reckon two days, and it's not going to be more than the three coming days').

(14) *u-fī-h* *umūr* *ḥatta* *āna* *mā*
CONJ=PREP=PRON.3SG.M affair.PL even PRON.1SG NEG
a-ʃraf-ha
PRES.1SG-know=PRON.3SG.F
'And there are things in it even I don't know (them)'.

3.3. Controller-dependent factors: chains of conjoined controllers

Before moving on to the analysis of the factors influencing agreement which are related to the nature of the target, some remarks are due concerning one last, statistically uncommon type of controller. Up to this moment, I have treated chains of conjoined controllers in the same way as other kinds of controllers[15]. Their composite nature, though, would seem to call for more detailed inspection. Commenting on coordinated heads, Belnap (1991: 81) writes in his thesis about agreement in Cairene Arabic that "heads containing one or more singular count noun categorically occur with plural agreement [...]. However, heads which consist only of plural forms show no such tendencies: they behave, instead, like heads consisting of a single plural noun: human heads favor plural agreement, inanimates favor [singular] agreement"[16]. In my corpus, all conjoined controllers denoting human referents contain at least one singular noun, and all of them take plural agreement, as in example (15):

(15) *taw* *sallūm* *u-bāgī* *ṣ-ṣġāyriyya* *kəll*
now sallūm CONJ=rest ART=children all
laʃab-hum *ḥarb* *u-māsk-īn* *ʃaṣā*
game=PRON.3PL.M war CONJ=hold.AP-PL.M stick
u-kann-hum *māsk-īn* *raššāš*
CONJ=PREP=PRON.3PL.M hold.AP-PL.M rifle
'Now, Sallūm and the rest of the kids, their only game is war, they hold a stick and [it is] like they were holding a rifle'.

The patterns of agreement triggered by the six chains composed of nouns with inanimate referents are shown in table 5:

[15] Chains of conjoined controllers were excluded from my analysis of the sound/broken morphological opposition (paragraph 3.1). Since no chain comprising both human and non-human referents was present in my corpus, though, they were all included in the discussion of the category of animacy (paragraph 3.2).

[16] See also Ferguson (1989: 88) about Damascus Arabic: "in a coordinate series of nouns serving as subjects, the agreeing verb or adjective may be feminine singular if ALL the nouns are non-human plural. If, however, even a single instance of a singular noun or a dual appears in the series, the agreement must be plural" (emphasis in the original). Note however that in D'Anna's work on agreement in Fezzani Arabic (forthcoming), some chains containing collective (i.e. morphologically singular) non-human nouns appear which trigger feminine singular agreement.

Type of agreement	Singular	Plural
Chains of singular or plural-and-singular nouns	0 (0 %)	4 (100 %)
Chains of plural nouns	1 (50%)	1 (50%)

Table 5: Chains of conjoined controllers (Non-Human Controllers Only)

As can be seen, chains of plural nouns are the only ones which occasionally attract singular agreement. Although obviously such small figures do not have any statistical relevance, and more research is needed on larger datasets, it is interesting to note that the only chain triggering singular agreement is composed of items which are low in specificity. In (16) a husband is rebuking his wife for no longer taking care of the house as she used to, complaining about all the good habits she has lost. He particularly regrets the fact that now she does not leave his clothes and the perfumes ready for him in the morning, while in the past she used to do that every day. He thus comments:

(16) *u-l-ʕaṭūr* *u-malābs-ī*
CONJ=ART=perfume.PL CONJ=clothe.PL=PRON.1SG
kull-ha *gāhz-a*
all=PRON.3SG.F ready.AP-SG.F
'And my clothes and the perfumes, they [used to be] all ready!'.

Here the husband is clearly not referring to any specific piece of clothing or perfume, since this was a habitual action in the past which recurred every day (and thus involving different clothes and perfumes every time). Compare (16) with example (17), where the reference is, instead, specific. In (17) the character is speaking while looking at the screen of his smartphone, and therefore referring to very specific calls and messages (note also that both controllers are quantified by a numeral lower than eleven):

(17) *sabʕa* *ttṣal-āt* *u-akṯar* *min* *ʕašra* *risāyil*
seven call-PL.F CONJ=more PREP ten letter.PL
wa-lā *fi-hum* *šī* *mumkin* *t-qūl*
CONJ=NEG PREP=PRON.3PL.M thing possibly PRES.2SG.M-say
ʕan-uh *ḍrūrī*
PREP=PRON.3SG.M important
'Seven phone calls and more than ten messages and [there is] not a thing in them you can call important'.

3.4. Target-dependent factors: target type

In his work on Cairene Arabic, Belnap (1991: 87) – building on Corbett's (1983) work on Slavic – proposed a hierarchy of target types based on their respective probability of attracting singular agreement: "Applying Corbett's prediction [...] one would expect the percentage of [singular] agreement to pattern as follows: attributive adjectives > attributive demonstrative pronouns > predicate verbs > predicate adjectives > anaphoric

pronouns". Belnap found this hypothesis to be valid for his data (with the exception of targets depending on the controller *nās*, "people", which patterned differently). Analyzing the kind of agreement triggered by different target types, I found this prediction to be fully confirmed in my corpus as well, as table 6 shows[17]:

Target Type	Probability of attracting singular agreement
Attributive adjective	94,3% (33/35)
Attributive demonstrative	88,9% (16/18)
Predicate verb	64,9% (24/37)
Predicate adjective	55,6% (10/18)
Anaphoric pronoun	40% (24/60)

Table 6: Target Type (Non-Human Controllers Only)

3.5. Target-dependent factors: distance between target and controller

In his survey of agreement in Cairene Arabic, Belnap (1991: 86) found distance of the target from its controller to be the second most important factor influencing agreement (after head type, which we discussed in paragraph 3.1). Corbett (2006: 237) maintains that while distance between target and controller is undoubtedly one of the factors which influence agreement, the Agreement Hierarchy still proves a more powerful explanatory tool. It is probably the case that the two are mutually dependent (since different target types have a statistical tendency to appear in different positions with respect to their controllers)[18]. At any rate, to make my data comparable with those

[17] The sample on which table 6 is based is composed of all targets depending on non-human controllers in my corpus. I excluded targets depending on human controllers from the analysis since, as we have seen, these tend to almost categorically bear plural agreement. As far as the collective *nās* is concerned (which, as we have seen, patterned somehow irregularly in Belnap's data), my corpus did not yield enough occurrences for a statistical analysis to be run. The targets depending on *nās* patterned as follow: 1 attributive adjective (1 showing singular agreement), 2 attributive demonstratives (0 showing singular agreement), 11 predicate verb (3 showing singular agreement), 2 anaphoric pronouns (1 showing singular agreement). Note that of the 11 verbs, 5 bearing plural agreement depended on the same occurrence of *nās*. In addition, one target consisting of the possessive particle *ḥaqq* was not considered in table 6. In the whole of my corpus, I found only one such target depending on a plural controller and showing inflection: it was not possible, therefore, to rank possessive particles as to their tendency to bear singular agreement. Belnap (1991) as well does not provide separate data for the Cairene equivalent of *ḥaqq* (*bitāʕ*).

[18] Belnap (1991: 86) thus comments on this point: "it would appear there may be a functional

of Belnap, I provide below an analysis of how distance between target and controller (in terms of phonological words) can affect agreement. In table 7, I provide data relative to all controller types. An objection could however be raised concerning the lumping together of controllers with high agentivity and low agentivity (since the latter have lesser chances to be associated with certain target types, such as verbs, which tend to occur at a greater distance from their controller than other types, such as adjectives). Therefore, in table 8 I provide data relative to non-human controllers only. As can be seen, both tables show a clear tendency: the further away a target is from its controller, the more likely plural agreement is to occur. Obviously enough, the overall percentage of singular agreement is higher in table 8, since this is far more common with non-human controllers: however, the two tables show the same kind of monotonic decrease[19].

Distance from controller	Total targets	Targets showing singular agreement
1	67	52 (76,1%)
2	48	18 (37,5%)
3 or 4	38	13 (34,2%)
5, 6 or 7	32	5 (15,6%)
Between 8 and 31	27	3 (11,1%)

Table 7: Distance between target and controller (all controller types)

Distance from controller	Total targets	Targets showing singular agreement
1	59	52 (88,1%)
2	32	17 (53,1%)
3 or 4	25	13 (52%)
5, 6 or 7	18	5 (27,8%)
Between 8 and 31	15	3 (20%)

Table 8: Distance between target and controller (non-human controllers only)

basis to the distance factor discussed here, as concerns the notion of recoverability of information. The nearer an agreement locus is to its head, the more immediate is the association between the two: [singular] agreement [is] far less likely to interfere with the interlocutor's perceiving the grammatical relashionship between the head and locus".

[19] In the corpus, a total of 13 targets appear which refer back to a controller previously mentioned by another speaker. In these cases, calculating the distance between target and controller was impossible, and these loci have been excluded from tables 7 and 8. Only two (15,4%) of these

As an example of the effect that distance can have on agreement, see (18) and (19) below, in which the same (highly non-specific) controller, *aġrāḍ*, 'things, stuff', triggers singular and plural agreement respectively. The occurrence of plural agreement in the second example can be explained in terms of distance between target and controller[20]:

(18) *zēn* *inta* *rūḥ* *u-ana*
good PRON.2SG.M go.IMP.2SG.M CONJ=PRON.1SG
b-a-rsil-l-ak *mēsag* *bi-l-aġrāḍ*
FUT=PRES.1SG-send=PREP=PRON.2SG.M message PREP=ART=thing.PL
illi *n-aḥtāg-ha*
REL PRES.1PL-need=PRON.3SG.F
'Good, you go and I will send you a message with the stuff that we need'.

(19) *smaʃ-nī* *ana* *bāġy-a*
listen.IMP.2SG.M=PRON.1SG PRON.1SG want.PART-SG.F
a-naqq *u-a-xtār* *l-aġrāḍ*
PRES.1SG-pick CONJ=PRES.1S-choose ART=thing.PL
bi-nafs-i *yaʃnī* *twaddī-nī* *maʃā-ak*
PREP=same=PRON.1SG that is PRES.2SG.M-bring PREP=PRON.2SG.M
u-a-štrī-hən
CONJ=PRES.1SG-buy-PRON.3PL.F
'Listen, I want to pick and choose the things by myself, I mean, bring me with you and I will buy them'.

As a further example, compare the two targets depending on *mḥawwalāt*, '(telephone) extensions' in example (20), where mixed agreement occurs. Here the mismatch might be a result of the different target types as well as the higher distance of the second target from its controller:

(20) *lə-mḥawwal-āt* *mətġayyar-a* *fōq* *taḥt* *u-mā*
ART=extension change.PART-SG.F above under CONJ=NEG
ʃād *ḥafiḍ-na-hin*
yet memorize-PAST-1PL=PRON.3PL.F
'The extensions changed completely and we haven't memorize them yet' (lit. 'changed up and down').

targets attracted singular agreement. Note also that targets occurring at a distance of three or more words from their controllers were less common than those occurring at a distance of one or two, so I have grouped these targets together in order to create statistically relevant and quantitatively homogeneous sets.

[20] This, in turn, would seem to suggest that distance between target and controller has a stronger effect on agreement than specificity.

3.6. Target-dependent factors: word order

Tables 7 and 8, above, are concerned solely with those targets which follow their controllers. A good number of targets appear in my data which precede their controllers: the type of agreement they attract is shown in table 9.

Distance from controller	Total targets		Targets showing singular agreement	
	All Controllers	Non-Human Controllers	All Controllers	Non-Human Controllers
-3 or -2	11	9	6 (54,5%)	6 (66,7%)
-1	34	28	21 (61,8%)	21 (75%)

Table 9: Distance between target and controller (targets preceding their controllers)

At a general level, we see that the rule which links distance to lower chances of singular agreement remains valid (though targets occurring at a distance of 2 or 3 words have more chances of attracting singular agreement when they appear *before* their controller, while the opposite is true for targets appearing right next to it). It may also be worth noting that 3 out of the 4 targets showing masculine singular agreement which appear in my data (see note 3 above) are to be found in pre-controller position[21]. Compare the two occurrences of the active participle *bāqī*, 'left, remaining' in the examples that follow: in (21) and (22) the participle depends on the same controller (*ayyām*, 'days'), but in the first sentence it precedes its controller (bearing masculine singular agreement), while in the second it follows it (showing feminine singular agreement instead)[22]:

[21] Those are two occurrences of the active participle *bāqī* and one verb. The only occurrence of an adjective following its controller and showing masculine singular agreement is represented by *kṯīr*, 'many, a lot' in the sentence *is-siyyāra hāḏi fi-ha ḏikriyyāt kṯīr*, 'this car has a lot of memories in it' (lit. 'this car, in it [there are] many memories'). In numerous varieties of Arabic *kṯīr* does often appear in the "frozen" masculine singular form, no matter the syntactic environment (see Brustad 2000: 64). This might be due to its high frequency of use, or to the fact that it also occurs as an adverb, with no difference in form, which may contribute to the spread of the non-inflected variant at the expenses of the inflected one. Similar considerations can be made for *bāqī*, which can appear as an adjective, an active participle used with verbal force, but also as a noun meaning "the rest, the remaining one(s)", in which case it is normally used within the boundaries of a synthetic genitive construction along with a second noun, and always in its uninflected form (see for instance example (15) above).

[22] Note that, depending (mainly) on the speaker, the underlying */q/ in *bāqī* can either surface as /q/ or /g/ in OA, as the examples show.

(21) *bāgī* *illa* *bas* *ʕašart* *ayyām*
remain.PART.SG.M but only ten day.PL
'[There are] but ten days left'.

(22) *xallī-na* *n-aʃīš* *əl-ayyām* *illi*
let.IMP.SG.M=PRON.1PL PRES.1PL-live ART=day.PL REL
bāqiy-a
remain.PART-SG.F
'Let us live the days which are left'.

In OA, demonstratives are particularly common in pre-controller position: this might be connected to the high percentage of demonstratives showing singular agreement in my data[23]. My corpus is too small to offer significant statistical evidence in this sense, but it is interesting to note that the only two occurrences of a demonstrative target following its controller (out of 24 total demonstratives) are also the only two occurrences of a demonstrative target showing plural agreement. One such occurrence is shown in example (23):

(23) *bi-iḏn-ēn-ī* *hāḏēna* *samaʃ-t*
PREP=ear-DUAL=PRON.1S DEM.PL hear-PAST.1S
in-nās *ya-qūl-u*
ART=people PRES.3M-say-PL
'With these two ears of mine I've heard the people saying...'

Consider also example (24) below, where the same controller attracts singular agreement in the demonstrative preceding it, and plural agreement in the anaphoric pronouns following it (obviously, target type might be another factor implied here as well):

(24) *hāḏī* *maʃāṣir* *henna* *tbīʃ-hən*
DEM.SG.F pen.PL henna PRES.2SG.M-sell=PRON.3PL.F
kəll-hən
all=PRON.3PL.F
'These pens for henna, sell all of them!'

[23] Prochazka & Gabsi (2016: 252), on the contrary, find plural agreement to be extremely common with demonstratives in Urban Tunisian Arabic. They do, however, remark how singular agreement (especially in verbal targets) is often associated with pre-controller position (Prochazka & Gabsi 2016: 246). Holes (2016: 334) reports the same for Bahrain Arabic. The high percentage of demonstratives showing feminine singular agreement in my data might be connected to the tendency, typical of Bahraini Arabic but possibly of other Gulf dialects as well, to use "the feminine form [of the singular demonstrative] as the 'default' demonstrative of vague reference" (Holes 2008: 199). This, however, remains a hypothesis waiting for further confirmation.

4. Discussion

In the course of this paper, I have shown what the main factors inducing agreement variation in OA are: the type of the controller (i.e. its denoting a human or non-human referent) seems to play a major role, with human controllers almost categorically triggering plural agreement, and non-human ones favoring singular agreement (with, however, a good deal of variation). The morphological status of the controller (i.e. its being a "broken" or "sound" plural), on the other hand, appears to have only a limited influence.

Humanness (or lack thereof) of the controller, however, does not by itself account for all the variation we observe in the data. Non-human controllers and generic human controllers (such as *nās*, 'people') can attract either singular or plural agreement depending on the context. I have shown how a strong correlation exists between agreement and target type, and between agreement and distance between target and controller (the effects of word order on agreement, on the contrary, appear to be minor).

Finally, agreement patterns appear to be influenced by other phenomena as well, such as concreteness and specificity of the controller. More concrete and specific controllers (that is, more individuated and non-grouped) have higher chances of attracting plural agreement. The role of definiteness and qualification, on the contrary, appears to be subordinated to that of other factors (if they actually play any role at all in the process of determining agreement).

These findings are consistent with previous studies on agreement in spoken Arabic. However, certain phenomena that did not receive in-depth treatment here (due to the limited nature of my corpus of data), are surely worth of further investigation. Among these, the role of different types of quantifiers in influencing agreement, the oscillation between masculine and feminine agreement with non-human controllers, and the relation which appears to exist between factors such as animacy, concreteness, quantification and specificity.

BIBLIOGRAPHICAL REFERENCES

Bettega (in press), *Agreement Patterns in Omani Arabic: Sociolinguistic Conditioning and Diachronic Development*, in De Ruiter, J. J., Ziamari, K., *Arabic Between Tradition, Globalization and Superdiversity*, special issue of *Sociolinguistic Studies*

Barlow, M. (1988), *A Situated Theory of Agreement*, Unpublished PhD thesis, Stanford, Stanford University

Belnap, K. (1991), *Grammatical Agreement Variation in Cairene Arabic*, Unpublished PhD thesis, Philadelphia, University of Pennsylvania

Belnap, K. (1993), *The Meaning of Deflected/Strict Agreement in Cairene Arabic*, in M. Eid, C. Holes (eds.), *Perspectives on Arabic Linguistics I*, Amsterdam and Philadelphia, John Benjamins: 97-119

Belnap, K. (1999), *A New Perspective on the History of Arabic Variation in Marking Agreement with Plural Heads*, in "Folia Linguistica", 33/2: 169-185

Blanc, H. (1970), *Dual and Pseudo-Dual in the Arabic Dialects*, in "Language", 46: 42-57

Brustad, K. (2000), *The Syntax of Spoken Arabic. A Comparative Study of Moroccan, Egyptian, Syrian, and Kuwaiti dialects*, Washington, D.C., Georgetown University Press
Corbett, G. (1983), *Hierarchies, Targets and Controllers. Agreement Patterns in Slavic*, London, Croom Helm
Corbett, G. (2006), *Agreement*, Cambridge, Cambridge University Press
Cowell, M. (1964), *A Reference Grammar of Syrian Arabic (Based on the Dialect of Damascus)*, Washington D.C., Georgetown university press
D'Anna, L. (forthcoming), *Agreement with Plural Controllers in Fezzānī Arabic*, in "Folia Orientalia"
Ferguson, C. (1989), *Grammatical Agreement in Classical Arabic and the Modern Dialects. A Response to Versteegh's Pidginization Hypothesis*, in "Al-ʕArabiyya", 22: 5-17
Hallman, P. (2000), *The Structure of Agreement Failure in Lebanese Arabic*, in Billerey, R., Lillehaugen, B., *WCCFL 19. Proceedings of the 19th West Coast Conference on Formal Linguistics*, Somerville, MA, Cascadilla Press: 178-190
Holes, C. (1989), *Towards a dialect geography of Oman*, in "Bulletin of the School of Oriental and African Studies", 52: 446-462
Holes, C. (2008), *The 'mixed' Arabic of the letters of 19th and early 20th century Gulf rulers*, in Lentin J. & Grand'Henry J. (eds.), *Moyen Arabe et Variétés Mixtes de l'Arabe à travers l'Histoire,* Louvain, Peeters: 193-229
Holes, C. (2016), *Dialect, Culture and Society in Eastern Arabia. Volume III: Phonology, Morphology, Syntax, Style*, Leiden, Brill
Owens, J., Bani-Yasin, R. (1987), *The Lexical Basis of Variation in Jordanian Arabic*, in "Linguistics", 25: 705-738
Prochazka, S., Gabsi, I. (2016), *Agreement with Plural Heads in Tunisian Arabic. The Urban North*, in Ritt-Benmimoun (ed.), *Tunisian and Lybian Arabic Dialects. Common Trends, Recent Developments, Diachronic Aspects*, Zaragoza, IEIOP: 239-260
Ritt-Benmimoun, V. (2016), *Agreement with Plural Heads in Tunisian Arabic. The Bedouin South*, in Ritt-Benmimoun (ed.), *Tunisian and Lybian Arabic Dialects. Common Trends, Recent Developments, Diachronic Aspects*, Zaragoza, IEIOP: 261-287

SIMONE BETTEGA • is currently Adjunct Professor of Arabic Dialectology and Postdoctoral Researcher in Linguistics at the University of Torino. His main research interests include the study of syntax, sociolinguistics and linguistic contact, as well as language documentation and linguistic fieldwork. In particular, he has focused on the varieties of Arabic spoken along the coasts of the Persian Gulf and in Oman, where he carried out fieldwork during his PhD.

E-MAIL • simone.bettega@unito.it

REDUPLICATION IN GULF ARABIC

Emma De Murtas

ABSTRACT • Although 'reduplication' is a term widely used in linguistics, to date a truly satisfying definition has yet to be formulated. From a semantic standpoint, there is a relationship between simple and doubled forms; the semantic sphere covered by all forms of reduplication is the one consisting of pluralization, intensity and even causativity, by following the principle of iconicity. Reduplication often offers an extension of the meaning of a word compared to the undoubled base, even though there are examples in which dictionaries do not register any difference in meaning between the base and reduplicated forms. In those cases, in which reduplication adds meaning, the reduplicated nouns are generally divided into clear-cut semantic groups, many of which are common to various languages: for example, diminutives, intensives, those that describe sounds or physical and personal characteristics. Many of these semantic groups are of an iconic nature. Nowadays, even if the arbitrary and conventional character of signs is recognized, a larger attention is brought to the iconic aspects of the languages and, more precisely, to the sound symbolic aspects of languages. Iconicity is the property inherent to different aspects of the language to represent the semantic values (the signifieds) as an image or an icon. In the course of the 20th century linguists have mainly neglected the study of imitative and expressive aspects. Yet the researches conducted on different languages have shown that the sound symbolic phenomena play an important role in the structure of languages
The study of reduplication in Semitic languages has long been neglected and, as far as Arabic is concerned, until recently only marginal attention was paid to it. In Semitics languages, the situation differs from that documented in Indo-European languages, since the (tri)consonantal root structure greatly influences reduplication: in Arabic its principal manifestation is the repetition of a biliteral root base (C1C2). The fact that one speaks of the reduplication of a biliteral root does not necessarily mean that that biliteral root exists. On the other hand, it has been noted that in most cases there is a connection with the known C1C2C2 type of triliteral root.
The purpose of the present article is to demonstrate the iconic role of reduplication in Gulf Arabic, where verbs based on quadrilateral roots are, as Clive Holes (2004) remarks, "among the most expressive, colourful and idiomatic elements in these dialects".

KEYWORDS • reduplication, quadrilateral roots, semantics, iconicity, Gulf Arabic.

1. Introduction

The purpose of this paper is to demonstrate the impact of reduplication on Gulf Arabic as described by Clive Holes in his "Dialect, Culture, and Society in Eastern Arabia" (2000) and to identify the cases in which reduplication may have iconic value. To this

end, I will first introduce the elements of general linguistics necessary in order to define the research field.

2. Sound Symbolism and Arbitrariness

The debate over the nature of linguistic signs dates back to the ancient Greek philosophy of language. In the classic dichotomy between physis (nature) and nómos (law) – the two antithetical hypotheses of the form of the words and their meaning (and therefore of language and the world represented by language) – there may exist either a natural relation, as Plato (Cratilo, trans. 2013) stated, or, on the contrary, as Aristotle (De interpretatione, trans. 1992) believed, a conventional relation shared by society through the use of language.

Since the time of Saussure's studies, modern linguistic theory has concentrated on the principle of arbitrariness (Hinton, Nichols, Ohala, 2006). The relation between the signified and the signifier does not have a direct and natural justification. Instead, this relation is arbitrary and conventional. In linguistics, arbitrariness is one of the characteristics of the linguistic sign. We speak of arbitrariness as opposed to iconicity because the elements of the linguistic sign are not naturally justified, but are rather the consequences of an implicit convention among people belonging to the same linguistic community.

The arbitrary character occurs on both planes: that of expression (signifier) and that of content (signified). De Saussure states that the relation between the signifier and its signified is arbitrary. This arbitrariness does not mean that the signifier derives from a free choice on the part of the speaker, but rather that the signifier is arbitrary with regard to the signified, with which it has no natural relation (De Saussure, 1916).

On the whole, 20th-century linguistics has neglected the study of the imitative and expressive aspects of linguistic sign. Yet research conducted on different languages has shown that sound symbolic phenomena play an important role in the structure of languages (Hinton, Nichols, Ohala, 2006).

Nowadays, even though the arbitrary and conventional character of signs is recognized, greater attention is attached to the iconic aspects of languages and, more precisely, to their sound symbolic aspects. Iconicity is the property inherent in different aspects of language to represent semantic values (the signified) as an image or an icon.

The fundamental instruments of sound symbolism are ideophones. The English word "ideophone" (coined by Doke in 1935 in his classification of Bantu languages) has been widely used to describe languages considered exotic at the time (for example, those of Southern Asia and Oceania). Today it is also used with reference to Western languages. The ideophone, an instrument of iconicity, is a linguistic category with a referential function. Hagège (1985: 307) states that ideophones exist in many languages, but not all of them. They impart an auditory, voiced image of an idea in order to express a state, a sensorial impression, a way of being or moving, an action that does not produce a sound or noise. He defines ideophones as a sort of onomatopoeia that uses rather rare phonic articulations or combinations in order to express a particular mental state associated with certain objects, movements or situations.

Raphael Kaboré brings us back to our specific case when he says that ideophones are not always reduplicated but, among those that are, many are used to describe small objects, small animals and insects, or to name very precise physical and mental qualities, or processes (Kaboré 1998: 359-376).

Some phonological and morphological mechanisms, such as the repetition of syllables or words, play a role in the formation of ideophones, and more generally of all sound symbolic phenomena.

3. Reduplication: Meanings and Areas of Research

Although 'reduplication' is a term widely used in linguistics, to date a truly satisfying definition has yet to be formulated (Prochazka 1995: 39): many different phenomena are attributed to reduplication, even when the only distinguishing feature is simple repetition in the broad sense of the word. Carl Rubino (2005: 11) calls it a systematic repetition of phonological material within a word for semantic or grammatical purposes and asserts that it is a morphological device used in many languages in the world. However, the common definition of the concept narrows the reduplication phenomenon down to the morphological aspect. For example, Haspelmath (2002: 274) defines reduplication as "a morphological process which repeats the morphological base either entirely or only partially" (italics mine). But what would seem a more efficacious definition is the one formulated by Maas (2005: 395), according to whom it is "a formal linguistic device that can be used at all levels of linguistic structure."

Despite the fact that reduplication strongly influences phonetics and morphology and that in some languages it also has grammatical functions – as in ancient Greek, for example, where the perfect is obtained by doubling the first syllable (γράφω – γέγραφα) – without a doubt its principal purpose lies in semantics.

Among the studies concerning reduplication and its functions, two in particular provide a global vision of this linguistic phenomenon. One of them, conducted by Carl Rubino (2005), identifies all the languages in which the reduplication is productive. The other made by Bernard Hurch (2005) systematically analyses the functions of reduplication using a database. These studies have shown that, up to now, reduplication has been studied as a phenomenon essentially morphologic or phonologic. It is therefore necessary to analyse the impact of reduplication on semantics. In fact, reduplication is one of the means most commonly used to describe non-verbal actions or conditions and is an imitative-analogical form of expression; that is to say, in most cases it is used spontaneously (one need only mention infantile language, with its wealth of reduplicated words). From a semantic standpoint there is a clear-cut relationship between simple and doubled forms; the semantic sphere covered by all forms of reduplication is the one consisting of pluralization, intensity and perhaps even causativity, by following the principle of iconicity according to which an increase in form implies an increase in meaning.

Reduplication often offers an extension of the meaning of a word compared to the undoubled base, even though there are examples in which dictionaries do not register any difference in meaning between the base and reduplicated forms. In those cases, in

which reduplication adds meaning, the reduplicated nouns are generally divided into clear-cut semantic groups, many of which are common to various languages: for example, diminutives, intensives, those that describe sounds or physical and personal characteristics. Many of these semantic groups are of an iconic nature.

4. Reduplication in Semitic Languages: A Historical and Theoretical Outline

Until recently, only rather marginal attention was paid to reduplication in Semitic languages, especially in Arabic. Pott (1862) and Fraenkel (1878) were among the first scholars to study reduplication. Later, Nöldeke (1904) compared a large number of names of animals that were subject to reduplication in various Semitic languages. Brockelmann (1908) was convinced that verbs were also affected by this phenomenon and Boekels (1990) analysed an entire dictionary of classic Arabic and made a list of all the reduplicated roots; however, he limited this study to only one lexeme per root–the verb.

In Arabic the situation differs from that documented in Indo-European languages: indeed, the (tri)consonantal root structure typical of Semitic languages greatly influences reduplication. In Arabic its principal form is the repetition of a biliteral root base (C1C2).

Various attempts have already been made to specify the semantic functions of reduplication: one of the first persons to do so was Pott, who in 1862 anticipated what would be taken up again only a century later by Harold Key (1965) in an article titled "Some Semantic Functions of Reduplication in Various Languages". Key succeeded in placing various functions of reduplication into different categories: he divided verbs into thirteen different classes, nouns into nine, and adjectives into three. Using Key's work as a starting point, Procházka (1995) demonstrated that 53% of reduplicated words are verbs, 35% nouns and 12% adjectives.

As is the case with all languages, in Arabic as well onomatopoeia plays a particular role in reduplication: the situation is complicated by the lack of vocalization, which is one of the distinguishing features of onomatopoeic expression in Indo-European languages, in which, for example, the vowel /i/ is associated with small size and /u/ with large size (Hinton, Nichols, Hoala, 2006). However, this cannot be considered valid in the case of Arabic.

5. The Types of Semantic Reduplication in Neo-Arabic

Procházka recognizes ten semantic types, although in certain respects this division into categories is subjective. Procházka does not place into a separate category those cases in which reduplication would seem to express intensification, but rather leaves them in one of the following ten categories. Consequently, the same lexeme can be found in more than one category.

1. **Acoustic phenomena.** A great many roots fall into this category, many of which might also be considered onomatopoeic. The largest number of words is the one that can be classified with sounds produced by the human voice (for example,

'to stutter', taʿtaʿa), also taking into consideration the fact that in this field we can find sounds such as those produced by coughing and snoring, words that denote speech without meaning, stuttering, names of linguistic errors, names of foreign populations, gossip, and words that express speech in a loud voice. A special subcategory of sounds comprises words that define the sounds made by water and the names of wells and rainwater, although in this latter case it is not clear whether this depends on the volume of the water or rather on the sound and movement it makes.

2. **Movements.** The second semantic type (in terms of quantity) in which reduplication is used quite a lot, in Arabic as well as in the great majority of languages, is the one that defines movements, which can be rhythmical, intense or continuous (for example, tartara, 'to shake').
3. **Optical effects.** Most of the words in this category include lexemes that we find in the category of rhythmical movements, but there are also words that express real acts of visual perception (for example, habhāb, 'Fata Morgana').
4. **Physical or mental qualities.** This category includes those roots whose principal feature is the semantic description of a condition or behaviour: physical or mental states (for example, zarzār, 'to be witty', 'to have a sense of humour'). Most of the words that derive from these roots are connected to human beings, but some refer to animals and, to a lesser degree, to objects.
5. **Names of animals.** The formation of the names of animals by doubling their roots is rather frequent; this phenomenon probably obtains because animals are named after the sound of their cry or call (for example, turtūr, 'turtle-dove')[1].
6. **Names of plants or fruits.** In this case, there is a less plausible explanation than the one given for the reduplication of roots that designate animals. Here we can presume that the roundish shape of various types of fruit may have given rise to this phenomenon. On the other hand, in the case of trees or plants, it may be that the semantic function of reduplication is to express plurality; or again, perhaps the presence of this phenomenon stems from the rustling of leaves (for example, dardār, 'elm').
7. **Rotundity.** This category would seem to be an exception in Semitic languages: in this case the function of reduplication might be that of expressing the reiteration of the movement of spherical objects (for example, kaykīyah, 'egg').
8. **Parts of the body.** The words in this category almost exclusively denote protuberant parts of the body. In this case the semantic functions expressed depend as much on the concept of rotundity as on the concept of something that hangs or leans (for example, ğalğalah, 'cranium').

[1] This is a root that is quite common in ancient Semitic languages and Indoeuropean (Persian and Latin). Due to its widespread nature it is impossible to establish whether it is of endogneous origin or is borrowed. For a discussion of this question, see Criscuolo, Alfredo, (2017), *Lessosemantica degli zoonimi nel semitico Nordoccidentale del I millennio a.C.* Napoli: 135-136.

9. **Geomorphology.** In this sphere as well, the number of words formed through the reduplication of radicals is considerable: although they describe spaces geomorphologically in the broad sense of the term, the largest percentage consists of terms that describe deserts and plains. It must be pointed out that while reduplication is present in terms that describe deserts and valleys, not one has been found that designates a mountain. This leads us to surmise that reduplication is used as a means of linguistic expression in horizontal space (for example, basbas, 'desert'). However, in this semantic class as well there are reduplicated lexemes (daqdāq 'small sand hill') that can be partially related to the concept of rotundity.
10. **Sundries.** This category comprises all those words that have not been included in the preceding ones.

Procházka's precious work is a fundamental starting point for anyone who intends to investigate the semantic value of reduplication. However, the division into classes is an arbitrary choice and the present article proposes a new category, clothing and ornaments, which is not included in the above-mentioned categories. We must also bear in mind that the different basic corpus may generate differences.

Adhering to an appeal made by Stephan Procházka (1995: 68) it has been deemed necessary to make comparisons with the dialectal varieties, which are less subject to the formal rules of Modern Standard Arabic and classical Arabic, in order to demonstrate the theses put forward regarding the iconic function of reduplication.

6. Reduplication in Eastern Arabian Dialects

As Clive Holes (2004) states in his article "Quadriliteral verbs in the Arabic dialects of eastern Arabia": "The EA quadriliteral verb is a particularly interesting subject for lexicology, as its morpho-semantic possibilities seem to have been extended and exploited in ways different from CLA/MSA. The result is that the quadriliteral verbs are among the most expressive, colourful and idiomatic elements in these dialects. They are particularly well represented in descriptions of bodily states, moods and actions, and there is a fairly systematic correlation between various form-types and meaning-types. Perhaps the most striking fact about quadriliteral verbs in the Eastern Arabian dialects is how common they are. Along with the structurally similar Theme II of the triliteral verb, the quadriliteral is the most productive verb type in these dialects (in the sense of 'still producing new verbs')."

The fact that one speaks of the reduplication of a biliteral root does not necessarily mean that that biliteral root exists. On the other hand, it has been noted that in most cases there is a connection with the known C1C2C2 type of triliteral root; for example, xamxam/xamxama ('sweep a lot', 'gobble up food') probably derives from xamm, 'sweep'. However, here we will not deal with the etymology or phono-morphological processes that have led to the reduplicated radical, but will analyse it synchronically as a given element in its original state, both because the object of the present analysis is the influence of reduplication on semantics and because there are numerous studies that analyse the phono-morphological processes that lead to reduplication.

6.1. Semantic Classes

The division into semantic classes has allowed for the division of the roots as follows:

I. Movements[2]

BKBK
bačbač /vi/ whimper, whine, snivel.

CHKCHK
čakčak /vi vn vr/ chunter, grumble (possible onomatopoeic origin).

KHMKHM
xamxam /vi vn *xamxama*/ **1** sweep a lot. **2** gobble food up.

QMQM
kamkam /vt/ nibble on, peak at.

MZMZ
mazmaz /vi/ nibble, snack.

MṢMṢ
maṣmaṣ /vi/ suck on.

NḤNḤ
tinaḥnaḥ /vi/clear one's throat, say 'ahem'.

WRWR
warwar /n pl *warāwir* / pneumatic drill, jack-hammer.

II. Sounds

XḌXḌ
xaḍxaḍ vt vn *xaḍxaḍa* shake, rock.
txaḍxaḍ /vi/ be shaken, rocked, set in motion.

HFHF
hafhaf /vi/fan, waft.

[2] Here we have followed Clive Holes' phonological transcription.

III. Clothing and ornaments

DNDN
dandūn/n pl *danādīn*/ pendant earring.

QBQB
gubgub and gubguba/n pl *gabāgib, gabāǧīb, gubāgīb*/ ornate piece of gold head jewellery worn by women at weddings and *xatma*.

MLML
malmal /n/ muslin.

NFNF
nafnūf / n pl *nafānīf*/ woman's dress, skirt.

IV. Plants

BRBR
barbīr and *birbir* /n/ purslane (*portulaca oleracea*).

KHSHKHSH
xišxāš /coll n/ poppies.

ṬRṬR
ṭarṭūr /n/ **1** bunch (of radishes, onions etc). **2** foolish person.

FLFL
filfil / n pl *falāfil*/ pepper (as a condiment or vegetable).

V. Animals

ǦRǦR
ǧarǧūr and *yaryūr* /n pl *yarāyir*/ shark.

ṢRṢR
ṣarṣūr /n pl *ṣarāṣir*/ cockroach.

ʕWʕW
ʃōʃaw /n/ star-fish.

QBQB
gubgub and gubguba/n pl *gabāgib, gabāǧīb, gubāgīb*/ crab.

NQNQ
naknūk /n/ small fish, tiddler.

The division into classes is a necessarily arbitrary exercise and it is not always easy to place a lexical unit into one semantic field rather than in another. For example, the term *hafhaf* – which means 'fan', 'waft' – may describe both the movement and the sound this movement produces. Here it is considered a movement.

The classes chosen here coincide with some of those that Procházka already found, except for the one that I have decided to call 'clothing and ornaments'. This difference probably also depends on the ethnographic character of Clive Holes' work.

Dealing specifically with the lemmas found, we have both terms such as *falfal / falāfil* ('pepper'), well known also in MSA, and others that pertain to this dialect spoken by fishermen: *gargūr / garāgir* (bee-hive-shaped fish-trap), *šāša* (a type of small primitive canoe made of palm-branches fastened together), *ǧarǧūr* (shark), and so on. This demonstrates that reduplication acts on the lexicon in different ways according to the languages and dialects under consideration and in keeping with the lexicographic sources consulted. Thus, thanks to the ethno-linguistic work carried out by Clive Holes, it is now possible to identify lexemes that represent clothing and ornaments.

7. Conclusions: Some Considerations on Iconicity

Clearly, not all reduplicated roots are iconic, but reduplication often entails a certain degree of iconicity, for example when it describes repeated movements. In order to demonstrate that reduplication itself is a bearer of iconicity I will take the liberty of carrying out a bit of contrastive analysis with some data gathered from one of the principal lexicographic sources of northern Levantine Arabic: Denizaeu's dictionary.

For example, in the case of the lemma *mazmaz*, Holes gives 'nibble' and 'snack' as the definition, while Denizeau's definition, in northern Levantine Arabic, is 'to drink by sucking slowly'. In both cases the lexical units refer to a sound made with the mouth, but in the first case the sound is made while swallowing solids and in the second while swallowing liquids. We can say with a certain degree of accuracy that both lemmas belong to the same semantic field but are not synonyms.

The case of the WRWR root is different: Holes translates *warwar/warāwir* as 'pneumatic drill' and 'jack-hammer', while for *warwar/ywarwer* Denizeau gives the meaning 'to gossip'. In this case the two lemmas are not in the same lexical field. Furthermore, in the first case an object is named for the sound it produces and the second case is an exemple of verbs reproducing the act di speaking (stammer, chatter, murmur etc.); therefore the relationship between signifier and signified is not arbitrary but rather iconic.

This first analysis clearly reveals the role that reduplication plays in semantics. While up to the present the morphological value of this phenomenon has been analyzed, it now seems necessary to engage in an in-depth study of the relationship between reduplication and the iconic aspects of language. Studies made on other linguistic families (for example, Oceanic languages) and the results of the first data analyzed here are encouraging in this sense. There are two prime considerations to be made at this stage. The first is the fact that this linguistic phenomenon is so widespread among the languages of the world, leaving aside the various morphological rules to which they are

subject. We need only mention that in Semitic languages, whose vocabulary is made up of triconsonantal roots, the reduplication consists of quadri-consonantal roots of the C1C2C1C2 type. On the other hand, we have the correspondence of reduplicated lexemes in various languages (e.g., *glouglouter* fr, *baqbaq / ybaqbiq* ar.pal, *mormorare* it, *murmurer*, fr, *barbar / ybarber* ar.pal.). This could lead one to surmise that reduplication is a linguistic tool of an imitative-analogical type: a prime example of this phenomenon is baby talk. Although it is true that reduplication often carries the semantic value of pluralization due to the principle of iconicity, according to which an augmented form is corresponded by augmented meaning, this does not appear to be sufficient to demonstrate the various functions of reduplication. As we have seen above in the analysis of Clive Holes' glossary, reduplicated roots may describe movement, sound, clothing and ornaments, plants or animals, and this wealth of semantic classes is also valid in the case of other languages, as is demonstrated in the studies made by Key and Procházka. This does not mean that the semantic classes involving reduplication are equal in all languages, but we can easily imagine that they have features that are obviously connected to structure of the language as well as to the culture of those who speak it. For example, in the dialects that Clive Holes has described one can find examples of marine aimals such as *ǧarǧūr* shark, *ʃōʃaw* star-fish, *gubgub* crab, *naknūk* small fish/tiddler because the dialect described is spoken by a community of fishermen.

This paper is merely a synopsis of the results obtained. Much still remains to be done, given the continuous evolution of the dialects (which, as we have seen, entails the ongoing formation of quadriliterals) and given the complexity of this linguistic phenomenon, which is so widespread in the languages of the world.

BIBLIOGRAPHICAL REFERENCES

Aristotle (1992), *Dell'interpretazione*, Milano, Biblioteca Universale Rizzoli

Boekels, K. (1990), *Quadriradikalia in den semitischen Sprachen unter besonderer Berücksichtigung des Arabischen. Dissertation*, Berlin, Freie Universität

Brockelmann, C. (1908-13), *Grundriss der vergleichenden Grammatik der Semitischen Sprachen*, Bde. I-II, Berlin, Reuther und Reichard

Criscuolo, A. (2017), *Lessosemantica degli zoonimi nel semitico Nordoccidentale del I millennio a.C.,* Napoli, Orientalia Parthenopea

Denizeau, C. (1960), *Dictionnaire des parlers arabes de Syrie, Liban et Palestine*, in "Etudes Arabes et Islamiques", Etudes et Documents 3, Paris, Éditions G.P.Maisonneuve

Hagège, C. (1985), *L'homme de paroles*, Paris, Fayard

Haspelmath, M. (2002), *Understanding Morphology*, New York, Oxford University Press

Hinton, L., Nichols, J., Ohala, J. (2006), *Sound Symbolism*, Cambridge, Cambridge University Press

Holes, C. (2000), *Dialect, Culture, and Society in Eastern Arabia, Vol. I. Glossary*, Leiden, Brill

Holes, C. (2004*), Quadriliteral verbs in the Arabic dialects of Eastern Arabia*, in M. Haak et al. (eds.) *Approaches to Arabic Dialects. A collection of articles presented to Manfred Woidich on the occasion of his sixtieth birthday*, Leiden, Brill: 97-116

Hurch, B. (ed.), (2005), *Studies on reduplication*, Berlin – New York, Mouton de Gruyter

Kaborè, R. (1998), *La réduplication*, in S. Platiel. R. Kabore (eds.), *Les langues d'Afrique subsaharienne*, Paris, Ophrys : 359-76
Key, H. (1965*), Some semantic functions of reduplication in various languages,* in "Anthropological Linguistics", 7 (3): 88-102
Maas, U. (2005), Syntactic reduplication in Arabic, in B. Hurch, Studies on reduplication, Berlin – New York, Mouton de Gruyter: 395-429
Nöldeke, T. (1904), *Tiernamen mit Reduplikation*, in T. Nöldeke (ed.), *Beiträge zur semitischen Sprachwissenschaft*, Strassburg, Trübner: 107-23
Plato (2013), *Cratilo*, Bari, Laterza
Pott, A. F. (1862), *Doppelung (Reduplikation, Gemination) als eines der wichtigsten Bildungsmittel der Sprache, beleuchtet aus Sprachen aller Welttheile*, Lemgo – Detmold, Meyer
Procházka, S. (1995), *Semantische Funktionen der reduplizierten Wurzeln im Arabischen*, in "ArOr", 63: 39-70
Rubino, C. (2005), *Reduplication*, in M. Haspelmath et al. (eds.), *The World Atlas of Language Structures*, Oxford, Oxford University Press: 114-17
Rubino, C., (2005b), *Reduplication: form, function and distribution*, in B. Hurch, *Studies on reduplication*, Berlin – New York, Mouton de Gruyter: 11-29

EMMA DE MURTAS • is a current PhD student in Human and Social Sciences at Université Sorbonne Paris Cité – Université Paris Descartes. She received her undergraduate degree from University "L'Orientale" of Naples with a thesis in Semitic Philology, written under the supervision of Professor Riccardo Contini. Her undergraduate thesis dealt with the contribution of Tawfīq Kanʿān to the study of Palestinian Arabic dialects, and in particular with the lexicon registered by Kanʿān in his numerous articles. Her current research is focused on the semantic functions of reduplication in North Levantine Arabic Dialects. Emma De Murtas is also interested in semiotics, typological linguistics, contact linguistics and dialectology.

E-MAIL • emmademurtas@hotmail.it

CARL REINHARDT'S LEXICAL DATA (1894): DISTINCTIVE FEATURES AND BORROWINGS

Roberta Morano

ABSTRACT • The linguistic studies carried out so far in Oman are very few and located in specific areas of the country; moreover, most of the data provided for Omani dialectology date back to the end of the nineteenth and the beginning of the twentieth century.
Carl Reinhardt's work – Ein arabischer Dialekt gesprochen in 'Oman und Zanzibar, dated 1894 – pays attention to the grammar, particularly to the phonology and morphology, of the Banū Kharūṣ vernacular, spoken in the areas of Nizwa and Ristāq (northern Oman), but also among the élite of Zanzibar island. The main purpose of his work was to provide a valuable linguistic guide to the German soldiers quartered on the island and in the Tanganyka region, which were for a short time an imperial German colony. The material supplied by Reinhardt still plays an essential role for neo-Arabic linguistic and dialectology, since it constitutes the richest available description of Omani Arabic, although deprived of a final glossary.
Reinhardt's lexical data, as outlined by Vollers (1895: 512), is extremely rich and characterized by some specific traits that make this vernacular different from any other Southern Arabic dialect.
In this paper, I'm going to present the main features of Reinhardt's lexicon, analyzing the huge amount of quadriconsonantal roots, some affinities with Yemeni and Gulf Arabic, the occurrence of some variations between the original meaning of a root and its use in the Banū Kharūṣ vernacular and, finally, borrowings from foreign languages.
This description is preliminary to my PhD fieldwork - which will be held by the end of this year - in the al-'Awābi district, where I intend to check if Reinhardt's dialect is still spoken and carry out a three-way linguistic comparison involving the dialect described by Reinhardt, the dialect actually spoken by present-day natives of the Wādī Banū Kharūṣ and the dialect of Ristāq.

KEYWORDS • Arabic dialectology, Omani Arabic, Oman, Carl Reinhardt, neo-Arabic linguistics.

1. Introduction: Reinhardt's lexical data

The Omani dialect described by Reinhardt[1] belongs to the Banū Kharūṣ tribe, and is spoken between Ristāq and Nizwa in the northern part of the country. This vernacular is different from both the one spoken in Musqaṭ as described by Jayakar[**](1889), and the one spoken on the coastal area between the capital city and Ṣūr.

The main purpose of Reinhartd's work was to provide a linguistic guide to German soldiers quartered on the island of Zanzibar and in the Tanganyka region, which were at that time – and for a short time – an imperial German colony. Despite the importance of Jayakar's lexical data, Reinhardt's material, according to Noeldeke[2], offers a clear overview of the Arabic spoken in Oman. Reinhardt's informants came from the city of Ristāq, the Banū Kharūṣ[3] tribe and the al-'Awābi tribe[4].

Karl Vollers[5] (1895: 487), in his study on Reinhardt's work, reports on an indigenous classification of the territory in four main districts: (a) *Ǧaʿalān*: infertile area stretching from the mountains of Ǧabal Akhḍar to Ras al-Ḥadd, mainly inhabited by nomads; (b) *'Umān*: Musqaṭ, the capital city, and surroundings; (c) *al-Bāṭina* (الباطنة): literally meaning "interior", an area densely populated and extremely fertile; and (d) *al-Ẓahra* (الظاهرة): literally meaning "inland", an area stretching from the western side of Ǧabal Akhḍar to the interior.

According to this classification, the Banū Kharūṣ vernacular is the one spoken in the *'Umān* area by the sedentary rural population. However, Reinhardt states that his Omani variety is Bedouin and not *ḥaḍari*, which is opposite to Vollers' (1895: 491) idea of this vernacular being an isolated and conservative Neo-Arabic dialect of Southern Arabia. Furthermore, he analyses some features, reported by Reinhardt's informants, that seem to make it part of the *ḥaḍari* class. For example:

- the realization of phoneme /q/ as voiced velar stop /g/;
- the realization of phoneme /ǧ/ as voiced velar stop /g/;
- the realization of phoneme /k/ as voiceless velar stop.

Regarding morphology and syntax, Reinhardt's analysis is fairly accurate and detailed, though the lack of a lexical repertoire somehow decreases its clarity.

Reinhardt's lexical data is extremely rich and characterized throughout by some specific traits that make this vernacular different from any other Southern Arabic dialect, mainly due to the particular geographic position of the country and to its linguistic contacts with different populations during its history.

[1] Reinhardt (1894), consulted alongside the critical review of Noeldeke (1895) and Vollers (1895).
[2] Noeldeke (1895a: 7).
[3] Descendant of the Yaḥmad, a branch of al-Azd, this tribe has played an important role in the history of Ibaḍiyya in Oman.
[4] Branch of al-'Abri tribe.
[5] German orientalist (1857-1909).

2. Specific traits in Reinhardt's lexical data

Some specific traits are found in the names of cardinal points, seasons and meteorological entities, such as: (1) *kōš* is "southern wind"; (2) *ezyeb* is "Northern wind", but the same word is, in the Neo-Arabic dialect of the Red Sea, the warm "South-Western wind"; in a similar way, the Ethiopic *azêb* is "Eastern wind", as in 'Adanī Arabic (Yemen)[6].

The Omani lexicon is particularly rich in variations between the original meaning of a root and its use in the Banū Kharūṣ vernacular. The root 'YŠ, "food, nourishment", means "rice" in Oman, whilst in Egypt and Yemen indicates "barley bread", in Sudan "millet" and a typical kind of "puree" in the Sahara Desert.

Another characteristic is the name for the "olive": whereas normally the words used to mean "oil" are *ḥall* or *salīṭ*, the Arabic noun *zaytūn* (Standard Arabic for "olive"), in Oman is used to indicate the "guava fruit", typical to tropical countries. The same happens also with the word *wa'l*, "gazelle" which is used in different parts of Arabic-speaking world to mean "ibex".

There are also some specific traits of Southern Arabic[7], such as the root RTᶜ, "to remain, stay"[8] or nouns like *ʿarraš* "to marry", *sabarah* "morning chill", *ġabšah* "early morning, dusk", etc.; some archaic features in the semantics, such as *rām* "to be able to"[9] from the Semitic root RYM, which means "to be high"[10].

Noeldeke (1895a: 21) also reported the richness of Reinhardt's lexical data, in which he found ancient and uncommon words: for example, the Yemeni noun *suḥḥ* "dates" is realized in Oman as *soḥḥ*; there are several nouns already known in other dialects but presented in this variety with unusual meanings (e.g. *wāḥa* "to reach, catch"; ĞMᶜ "to gather" stands here for "to sweep" and *ǧummaᶜ* for "rubbish, trash") and, finally, words with a different meaning from Classical Arabic, but found in other Neo-Arabic dialects such as the root ŠWF "to see", ḤDM "to work, to serve" and adjectives like *šēn* "ugly" or *zēn* "beautiful", respectively derived from nouns *šayn* and *zayn*.

3. Affinities with Yemeni and Gulf Arabic

Affinities with Yemeni and Gulf Arabic are many, in terms of both roots and meanings. Talking about Yemeni Arabic in general is not appropriate, since in Yemen there are a huge number of dialects – many of which not mutually intelligible. However, it is possible to take it into account because of its large amount of archaisms that allow, in some cases, to determine a guideline in the development of lexical varieties in the Southern part of the Arabian Peninsula.

[6] Vollers (1895: 507).
[7] Vollers (1895: 509).
[8] Cfr. RTᶜ in Copeland Biella (1982: 497).
[9] Ibidem.
[10] Cfr. RYM in Copeland Biella (1982: 487).

Sometimes nouns undergo changes in their internal structure but retain the same meaning, whereas take on multiple meanings despite keeping the same root stem. This is the case of Yemeni root ʿNB (*ʿanb al aḍām* "mango"; *ʿinab* "grapes"; *ʿanbarūd* "pear") which in Omani *embe* has the sense of "mango".

The word *yebs*, which means "narrow" in Omani Arabic, acquires the meaning of "drought" in the form of *ybūsīye*, corresponding to the Yemeni and CA word *yibbīs* or *yābis*, which means "dry".

Same behavior for the word *ḫarṭūm* ("face"), which is *ḫarṭūm* ("proboscis; water hose") in Gulf Arabic and *ḫurṭūm* ("nose") in Yemeni Arabic. Particular is also the word *tzelzel* ("to shake") which in the Yemeni form *zilzilat ʿarḍ* means "earthquake". Finally, the Omani word for "wife", *čhil*, is very similar to the Yemeni word *kahlah* / *čahlah* ("old woman") and to the Arabic root KHL (e.g. CA *kahl* "middle aged man").

Related to the vocabulary of Gulf Arabic are, among others, the Omani terms: *bkem* (< *bakam* in Gulf Arabic) "dumb, voiceless"; *zerbūl* ("tights") which turns into *zarbūl* ("woolen sock").

There are also some specific lexical affinities, although to a considerably lesser extent than others, with the Najdi Arabic and Classical Arabic, such as the terms: *raġġāğe* "belling" (< CA *rġa* "protest gurgling of the camel"); *ḥedeb* "hunchbacked" (< Najdi Arabic and CA *aḥdab* "curved"); *tdelhem* "to be cloudy" (< Najdi Arabic and CA *idlahamm* "to get dark").

4. Loanwords

Loanwords are prominent in Reinhardt's lexical data. The interference between two languages involves mainly the lexicon (Fusco 2008: 39): in this field, in fact, it is easy to recognize the magnitude and the number of loans in any given language. The loan is not a sporadic mention of foreign words, but a terminology effectively used and institutionalized in the linguistic system of a language model. In order to recognize and study a loanword, several factors must be taken into account: firstly, the demonstration of a relationship of dependency between the foreign word and the chosen one in the language model; secondly, even the global linguistic system which the loanword is in touch with plays a fundamental role, since it affects and contains the use of the term or its yield; and finally, it is also important to reconstruct the historical processes that led to the contact and the possible interference between populations.

Omani Arabic has multiple loans from Persian, Hindi, English and, to a lesser extent, from Swahili and Portuguese. This occurrence is due not only to incursion of foreign populations into the country during the centuries, but also to the flourishing naval business that Oman has enjoyed for most of its history.

Loanwords are, for example, abundant also in the toponymy of Oman, with mainly Persian influences: the origin of the place name Sīb is in the Persian word for "apple"; Ristāq in Persian means "village" and the term Khōr Fakkān indicates an "estuary". A possible etymology of the choronym "Umān" is, according to the British Arabist and missionary George Percy Badger, the local term *ʿamān* used to mean "house, permanent settlement" (Vollers 1895: 486).

Loanwords from Persian are, along with Hindi ones, the most conspicuous group in Omani lexicon. The term *bidār* ("worker") comes from the Persian equivalent that means "awake, active"; *tefaq* ("rifle") comes from the Persian *tofak* ("blowgun"); *ġaliyūn* ("tobacco" in Omani Arabic) means "pipe" in Yemeni Arabic and originates from the Persian *ġaliyān* ("hookah") (Reinhardt 1894: 126).

Noeldeke (1895a: 22), also a famous Iranist, reports several other loans, such as: *ḫumra* "dates" < Pers. *ḫurma*, with metathesis; *zenǧel* "thicket" < Pers. *ǧanǧal*; or *hest* "big, much" < Pers. *hast* "it exists"; *nemūne, nemne* "model" < Pers. *nmwnh*; *bitk* "blacksmith hammer" < Pers. *putk* "hammer"; *ṣardit lḥawe* "cold is arrived" < Pers. *sard* "cold".

Loanwords from English are, for example: *fīf* ("pipe"); *anǧer* ("anchor"); *inǧenīr* ("engineer").

Hindi loanwords, instead, are related to nautical or import/export field: *benǧri* ("bracelet"); *hōri* ("boat"); *ǧūdi* ("dry dock"); *kittara* ("saber").

Contrary to expectations, Swahili loanwords are extremely low in the Omani lexicon notwithstanding the length of time an Omani sultanate has existed on Zanzibar Island and the long association that existed between the two countries throughout the seventeenth century after the expulsion of the Portuguese. The Arabic dialect spoken in Zanzibar has many similarities with the Omani one. It is a sedentary vernacular and has no ties with any of the Modern South Arabian languages.

5. Quadriconsonantal Roots

What seems clear from the analysis of Omani vocabulary is how rich it is in quadriconsonantal roots. This peculiarity lies not only in the extremely high occurrence of this particular kind of roots, or in their developing morphological and semantic processes different both from MSA[11] and from Classical Arabic, but also in the fact that they are the most productive verbal model. This feature is also shared in part with other dialects of Eastern Arabia, particularly some dialects of Bahrain, as demonstrated by Clive Holes (2004) in his specific study. On the origin of quadriconsonantal roots, there are several hypotheses including a dialectal derivation from words that are completely unknown today.

Almost all verbs in quadruple root possess, in the meaning, an intensive value. This also applies, in part, to the second deriving form of dialectal verbs in Classical Arabic. In the Gulf and Najdi dialects, they mainly have an intensive/repetitive value in addition to the causative one.

This abundance can be explained, on one hand, by the large occurrence of foreign loanwords (e.g., from the root HNQR comes the word *hanqri* "rich", originated from Hindi; or the verb *kansal* "to cancel" from English), and, on the other hand, by a

[11] Modern Standars Arabic.

denominative development, also from Classical Arabic (e.g. *tǧarham* "to attack, to assail ferociously" < CA *ǧirham* "epithet of a lion"; *ṣandaǧ* "to cover over, roof over" < *ṣandūǧ* "box, basket"; Holes 2004: 98). Furthermore, the occurrence of quadriconsonantal roots can be explained by a phonological extension of the basic form, as a consequence of the insertion of some phonetic elements or through the repetition/reduplication of radicals (e.g. ĠRĠR *ġarġar* is "to gurgle"; KZKZ *kezkez* is "to shiver, to chatter from the cold"), both in verbs and nouns (e.g. TRTR *terter* is "sand"; FLFL *filfil* is "pepper"; BLBL *bulbul* is "nightingale"; Reinhardt 1894: 55).

The reduplicative group is, in fact, the most substantial among the quadriconsonantal roots. According to Holes' study, they originated from doubled roots with intensive or repetitive value as compared to the original verb. For example: *lamlam* "to collect together, go around collecting from here and there" < *lamm* "to collect"; *dağdağ* "to knock, beat" < *dagg* "to knock" (cf. San'ani Arabic دقدق); *naṭnaṭ* "to jump, to flee" < *naṭṭ* "to jump"; *ṭamṭam* "to completely submerge, fill to the brim" < *ṭamm* "to cover over" (Holes 2004: 100).

Sometimes the deriving forms of these reduplicate roots differ completely from the meaning of the original verb (e.g. *tġaṣġaṣ* "to be stuck, obstruct" < *ġaṣṣ* [noun] "block" and not < *ġaṣṣ* [verb] "to choke").

Another hypothesis is the derivational origin from triliteral verbs or nouns with a weak vowel, even if, according to Holes, this can be true only for few cases: *ḍaḥḍaḥ* "to be late in the morning" < *ḍiḥa* "late morning"; *salsal* "to drain" < *sāl* "to flow" (Oman, Bāṭina coast).

Other reduplicative verbs include onomatopoeic verbs, a category well represented in most of Arabic dialects and not just in the Eastern ones (e.g. *ḫağḫağ* "to stutter"). Many of these verbs are also formed through the insertion of a liquid consonant, which was probably the original second radical then lost: *barbağ* "to blow, to make bubbles in water" (cf. CA *baqbaqa* "to make a gurgling sound" or CA *baqqa* "to pour abundant rain"); *ḫarḫaš* "to jingle, rustle" (cf. CA *ḫašša* "to clink [jewels]"); *tbalbas* "to be speechless (with fear)" (cf. CA *balasa* "to be desperate, speechless, stupified"; Holes 2004: 104-105).

In Omani Arabic, there are a number of verbs with a nasal consonant inserted in the second position, such as *tsansaḥ* "to slither, slide down" from CA *saḥḥa*, *tasaḥḥa* and *tasaḥsaḥa*, used to describe water flowing.

In some Bahrain dialects, the meaning of reduplicative forms can be linked to an older form of Arabic, rather than to their dialectal original (e.g. *šabšab* "to work hard, run hither and thither" < CA *šabba* "to prance"; dialectal *šabb* is "to set fire to"; Holes 2004: 102).

Onomatopoeic verbs or verbs describing rapid and sudden actions use a pattern with the insertion of a semivowel in the basic theme, in order to modify the meaning: *lōlaš* "to ululate"; *tsāsar* "to whisper to one another" (Holes 2004: 106)[12].

[12] Cfr. Ingham, 1994: 82, for Najdi *sāsar* "to whisper to", the only verb in this category to present a /ā/ instead of a /ō/ or a /ē/.

Eastern Arabic dialects, in order to denote colors or defects and physical characteristics, insert, after the first radical, the semivowel /w/, alongside the use of the ninth deriving form (e.g. *ḥōmar* "to be, become reddish"). The likely effect of a contamination is the variation, in these verbs, of /ā/ and /ō/. On this pattern are also formed the denominative forms: *tsōlaf* "to have a chat, exchange stories" < *sulāfa* "matter, affair". A good example of this process is the verb *ṭōrab* (Holes 2004: 108) "to dance the *trēmbo*" (a typical dance of some Shiite communities, performed by four women around the bride during the preparation and also during the wedding festivity) < CA *ṭarab* "delight, rapture" (referring particularly to music).

With the insertion of the semivowel /y/, Eastern dialects use this to denote a mental and physical state, generally with a pejorative value: *tbēḥal* "to become stingy"; *thēbal* "to act like a stupid". The same pattern is also found in San'ani Arabic for diminutive verbs (Watson 2006: 191).

One of the most difficult things in the searching of the origin of quadriconsonantal verbs is the different meanings of an identical root in the Arabic-speaking areas. For the Peninsular area – in Bahrain and in the Gulf area – the verb *karfas/čarfas* means "to hit the ground", but in Najdi Arabic it means "to bent".

There are also verbs with the semivowel /w/ in third position. They are numerically low and their origin is still unsure. An example is the verb *haǧwal* ("to get rid of something"), used in some Shiite villages of Eastern Arabia: one of the most credible hypotheses is the derivation from the CA Imperative form *haǧ* < *hiǧǧ* "leave!" + *wall* "go away!" (Holes 2004: 113). This is an example of *blending*, a not particularly common procedure in neo-Semitic or among the European languages (e.g. *smog* < *smoke* + *fog*; Holes 2004: 114.

Finally, there is evidence of quadriconsonantal roots as result of shorthand symbols, which classical Arabic is rich of (e.g. *basmala* or *haǧwal*). The verb *saḥlab* "to drag" appears to be the combination of the two roots SḤL and SḤB: one of the meanings of SḤL, in classical Arabic, is "to scrape, peel", whereas in the Baḥrayn Arabic it is "to become smooth". SḤB, in almost all its varieties, means "to pull". Therefore, the verb *saḥlab* will gather in its global meaning all the semantic combination of these elements.

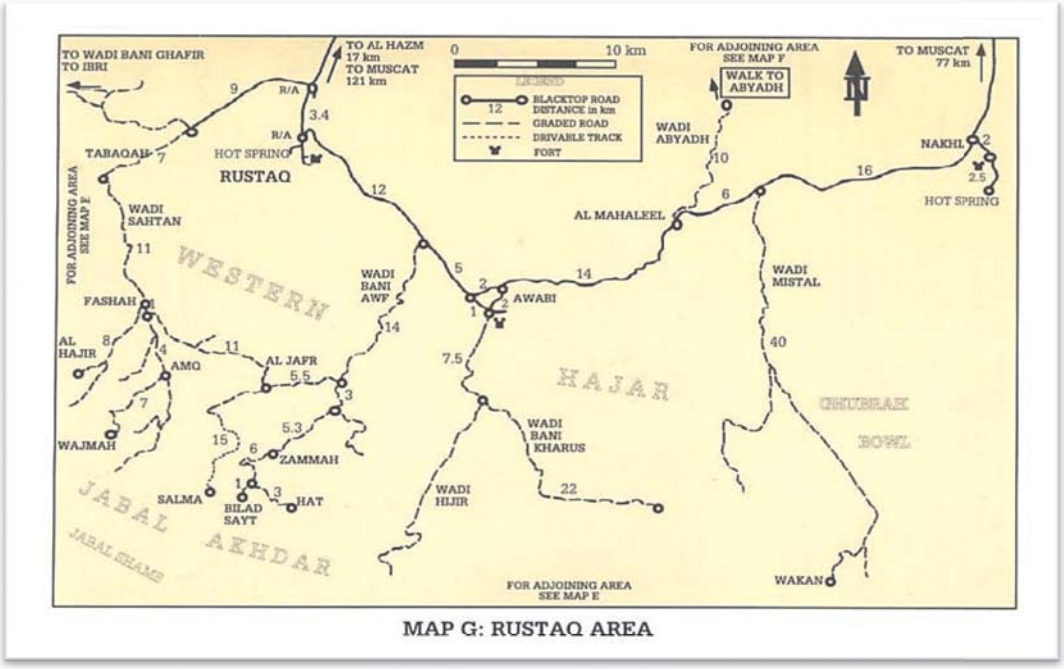
6. Conclusions

The main purpose of this paper was to generically frame the lexical variety and the specific characteristics of Omani Arabic. The specific features presented here – especially the large amount of loanwords and quadriconsonantal roots – make the Arabic dialects of Oman highly distinctive in comparison to other dialects of the Arabian Peninsula.

So far, all the studies carried out on dialectal diversification in Omani Arabic focused on phonological and morphological isoglosses, according to current dialectological practice. With the exception of some specific areas of the country, a systematic study of the lexicon which allow us to have a clear lexical and semantic classification of this variety, still does not exist. However, lexical diversification also plays an important role, despite the fact it is less studied because of the absence of a reliable and comprehensive repertoire for Omani Arabic.

Overlooking the Arabian Sea and not far from Iran and India, Oman has been influenced by quite a numerous cultures and languages over the centuries, and that is what makes its variety of Arabic extremely interesting from a lexical perspective.

A new and deeper interest in lexical research for Omani Arabic would be beneficial in order to develop a comprehensive lexical repertoire for this variety.



Map 1: Wadi Banū Kharuṣ and surroundings

Map 2: Position of al-ʻAwabi district in Northern Oman

BIBLIOGRAPHICAL REFERENCES

Copeland Biella, J. (1982), *Dictionary of Old South Arabic. Sabaean Dialect*, Chico (California): Scholar Press

Fusco, F. (2008), *Che cos'è l'interlinguistica*, Roma: Carocci

Holes, C. (2004), *Quadriliteral verbs in the Arabic dialects of Eastern Arabia*, in M. Haak et al. (eds.), *Approaches to Arabic dialects. A collection of articles presented to Manfred Woidich on the occasion of his sixtieth birthday*, Leiden, Brill: 97-116

Noeldeke, T. (1895a), *Über einen arabischen Dialekt*, in "WZKM", 9: 1-25

Noeldeke, T. (1895b), *Nachträge zu dem Aufsatz "Über einen arabischen Dialekt",* in "WZKM", 9: 177-179

Piamenta, M. (1990-91), *Dictionary of post classical Yemeni Arabic*, I-II, Leiden, Brill

Praetorius, F. (1880), *Über den arabischen Dialekt von Zanzibar*, in "ZDMG", 34: 217-231

Qafisheh, H. A. (1997), *The NTC's Gulf Arabic – English dictionary. A compact dictionary of the contemporary Arabic of the Middle East*, Lincolnwood (Chicago), IL, NTC Publishing Group

Reinhardt, C. (1894), *Ein arabischer Dialekt gesprochen in 'Oman und Zanzibar*, Stuttgart – Berlin, W. Spemann

Socin, A. (1896), *Rev. of Reinhardt 1894*, in "Göttingische Gelehrte Anzeigen": 125-129

Stumme, H. (1895), *Rev. of Reinhardt 1894*, in "Literarisches Zentralblatt": 18-19

Vollers, K. (1895), *Rev. of Reinhardt 1894*, in "ZDMG", 49: 484-515

Watson, J. C. E. (2006), *Arabic morphology: Diminutive verbs and diminutive nouns in San'ani Arabic*, in Morphology, 16 (2): 189-204

ROBERTA MORANO • is a current PhD student in Linguistics & Phonetics at University of Leeds (UK). Her main research interests are on Arabic dialectal varieties spoken in Oman. Her research project aim is to analyze, morphologically and phonologically, the Arabic dialect spoken in the al-'Awabi districts (Northern Oman), by comparing it with the variety described by Carl Reinhardt in 1894. She graduated from University of Naples "L'Orientale" in 2011 with a BA thesis specifically focused on Carl Reinhardt's work, under the supervision of Professor Riccardo Contini. Her BA thesis analysed all the morphological and lexical features presented by the German linguist over a century ago and it also had a lexical glossary at the end. She then obtained a MA degree at the same University in 2014, working on a MA thesis entitled Glossary of Omani Arabic (in Italian). In this work, studies of other scholars have been analysed, focusing on different areas of Oman and comparing them with the main version of Carl Reinhardt (1894). Also in this thesis, she compiled a lexical glossary, including many varieties of Omani Arabic.

E-MAIL • mlrmo@leeds.ac.uk

«QuadRi»
Quaderni di RiCOGNIZIONI
ISSN 2420-7969

è una collana di

RiCOGNIZIONI
Rivista di lingue, letterature e culture moderne
ISSN: 2384-8987

http://www.ojs.unito.it/index.php/ricognizioni/index
ricognizioni.lingue@unito.it

Diparimento di Lingue e Letterature Straniere e Culture Moderne
Università di Torino
http://www.dipartimentolingue.unito.it/