

Sociolinguistic variation in the *Yāl Saʿad* dialect in northern Oman

Sara Bakhit Al Sheyadi

A thesis submitted for the degree of Doctor of Philosophy

Department of Language and Linguistics

University of Essex

April 2021

@ Sara Al Sheyadi 2021

To Bader Ali

Contents

List of Tables.....	vi
List of Figures.....	vii
List of Maps.....	viii
Abstract.....	ix
Acknowledgements	xi
<i>Introduction.....</i>	<i>1</i>
<i>Chapter One: Sociolinguistic profile.....</i>	<i>4</i>
1.1 Oman and the Bāṭina Coast: A brief background.....	4
1.1.1 Location and topo-geography	4
1.1.2 Population	9
1.1.3 Economy: before and after the discovery and production of oil.....	9
1.2 Al-Suwaiq: Geographic, demographic, and socio-economic structure.....	11
1.2.1 Location	11
1.2.2 Demography.....	14
1.2.3 Socio-economy	16
1.3 History and tribal structure.....	19
1.3.1 Ancient Oman	19
1.3.2 The Omani Empire and maritime influence.....	20
1.3.3 Tribal structure and the Arabicisation of Oman	20
1.4 Linguistic profile and dialect topography of Oman	23
1.4.1 Ethno-linguistic minorities in northern Oman	24
1.5 Literature on Omani Arabic	27
1.5.1 An overview of selected previous works	27
1.5.2 Omani Arabic as a separate dialect group.....	30
1.5.3 Bedouin vs. Ḥaḍari.....	35
1.5.4 Foreign influence on Omani Arabic	38
1.5.5 Omani Arabic as spoken in al-Suwaiq.....	40
<i>Chapter Two: Notes on the phonology, morphology, and syntax.....</i>	<i>42</i>
2.1 Preliminaries	42
2.2 Phonology	42
2.2.1 Consonants.....	42
2.2.2 Vowels	48
2.2.3 Ghawa syndrome and Najdi re-syllabification	51
2.3 Morphology/ morphophonology	53
2.3.1 Suffixation.....	53
2.3.2 Assimilation of /h/ in the 3 rd person suffix pronouns	54

2.3.3	The definite article	55
2.4	The apophonic ‘internal’ passive (AP).....	55
2.4.1	Apophonic passive in Oman	59
2.4.2	Apophonic passive in the study area.....	63
2.5	Negation in the dialect	66
2.5.1	Negating with <i>mā</i>	67
2.5.2	Negating with <i>ab</i>	69
2.5.3	Negating with <i>am</i>	77
2.5.4	Negating with <i>lā ~ la</i>	81
Chapter Three: Methodology.....		83
3.1	Preliminaries	83
3.2	The sample.....	84
3.2.1	Defining the sample	84
3.2.2	The sampling method.....	85
3.2.3	Sample stratification	86
3.3	The researcher.....	91
3.4	The pilot study.....	91
3.5	The fieldwork	93
3.6	The interview procedure	94
3.6.1	Distribution of the informants.....	97
3.7	Methodological considerations	98
3.8	The external predictors in this study	99
3.9	The dependent variables in this study.....	100
3.9.1	The (dʒ) variable	100
3.9.2	The DEF variable	101
Chapter Four: Literature on the (dʒ) variable.....		102
4.1	Diachronic development(s).....	102
4.2	Current geographical distribution of /dʒ/ and key issues.....	104
4.3	/dʒ/ in Oman.....	113
4.4	(dʒ) as a sociolinguistic variable: Trends and issues.....	117
4.5	(dʒ) as a sociolinguistic variable: Trajectories of change	128
4.5.1	Lenition and fortition: Internal constraints	129
Chapter Five: Data analysis and findings for the (dʒ) variable.....		138
5.1	The dependent variable	138
5.2	Notes on the lexical distribution of the variants.....	141

5.3	Envelope of variation and token extraction.....	142
5.4	The coding procedures for (dʒ) and the modelling process.....	145
5.4.1	Internal predictors	145
5.4.2	Modelling.....	149
5.4.3	Results and discussion	153
5.5	Summary and conclusions.....	179
Chapter Six: Literature on the definite article.....		184
6.1	Definiteness in Arabic.....	184
6.2	The Arabic definite article: socio-historical background.....	187
6.2.1	The origin of the definite article	187
6.2.2	The form of the definite article	189
6.2.3	The syntax of the definite article	191
6.3	Definiteness and the definite article in the dialect of <i>Yāl Saʿad</i>.....	192
6.4	Null definite in the dialect of <i>Yāl Saʿad</i>.....	194
6.5	Other related varieties with a null definite.....	195
6.6	Diagnostic tests for the NULL category	200
6.7	The linguistic distribution of the NULL category in the dialect.....	204
6.8	Sociolinguistic studies on the definite article.....	208
6.8.1	Tagliamonte and Roeder on Yorkshire English.....	209
6.8.2	Alqahtani on Tihāmi Qaḥṭāni Arabic.....	210
Chapter Seven: Data analysis for the DEF variable.....		212
7.1	Quantitative analysis	212
7.1.1.	Levels of analysis.....	212
7.1.2.	Preparing the data for analysis	213
7.2	Data analysis: external predictors.....	217
7.2.1	Distributional statistics.....	217
7.2.2	Multivariate analysis	223
7.3	Data analysis: Internal constraints	230
7.3.1	Coding procedure for the linguistic constraints.....	231
7.3.2	Modelling process	238
7.3.3	Results and discussion	241
7.4	Summary.....	260
Conclusion.....		263
References		268
Appendices.....		285

List of Tables

Table 2.1 Consonant inventory in this dialect.	43
Table 2.2 Combinatorial possibilities for /q/, /dʒ/, and /k/ reflexes in B-type varieties in Oman (Holes, 1989: 451).	46
Table 2.3 Bedouin dialectal types in Omani Arabic (Holes, 1989: 449-454).....	47
Table 2.4 Apophonic passive tokens distributed across different verb types.....	64
Table 2.5 Apophonic passive tokens distributed across morphological categories.....	64
Table 2.6 Negation of the active participle with <i>am</i>	77
Table 2.7 Negation of the passive participle with <i>am</i>	78
Table 3.1 Distribution of the informants across age, gender, and locality.	97
Table 5.1 Example of (dʒ) coding.	148
Table 5.2 Internal and external predictors included in the final model.	152
Table 5.3 Results for the use of [gʲ] in the dataset.	154
Table 5.4 Distribution of [gʲ] per age group.	159
Table 5.5 A crosstab showing the proportion of the use of [gʲ] per age and gender.....	161
Table 5.6 A crosstab for the use of [gʲ] per age, gender, and locality.....	164
Table 5.7 Results for the use of [gʲ] in the dataset; J26MTH excluded.	172
Table 7.1 Distribution of the dependent variable per age and gender.	221
Table 7.2 Distribution of the dependent variable per age, gender, and locality.	222
Table 7.3 Rbrul results of the DEF variable: external predictors.....	228
Table 7.4 Distribution of the dependent variants for Salim and Khamis.....	231
Table 7.5 Coding for the type of noun factor group.	233
Table 7.6 Coding for the syntactic function factor group.....	236
Table 7.7 Sample of the token coding.	237
Table 7.8 The final model for the multivariate analysis.	241
Table 7.9 Results of the DEF variable: linguistic constraints.	242
Table 7.10 Distribution of the dependent variants per the morphological number.	243
Table 7.11 Distribution of the dependent variants per the morphological form.....	244
Table 7.12 Results of the DEF variable: gender included.....	245
Table 7.13 The distribution of the DEF variable per the type of bare noun category: ‘common’ vs. ‘proper’.....	246
Table 7.14 The distribution of the DEF variable per the type of bare noun category: ‘abstract’ vs. ‘concrete’.....	247
Table 7.15 Distribution of the DEF variable per syntactic function.....	248
Table 7.16 Distribution of NULL for the whole dataset per age group and type of noun.	251

List of Figures

Figure 3.1 The distribution of informants' villages across three localities.....	94
Figure 3.2 The external predictors in this study.	99
Figure 4.1 Diagrammatical pattern of the phonetic changes of Old Arabic /dʒ/ (adopted from Al-Nassir, 1993: 44).....	103
Figure 4.2 The development of the /q/ reflex of /dʒ/ in Urban Egyptian Arabic (based on Hary, 1996).	106
Figure 4.3 A reconstruction of /dʒ/ (based on Blanc, 1969).	109
Figure 4.4 A summary of the developmental stages of Arabic /dʒ/ and the contemporary distribution of the main variants in the Arabic-speaking world.	113
Figure 5.1 Screenshot of E-LAN as used in the analysis for this variable.	143
Figure 5.2 Distribution of (dʒ) variants per age group (a).	160
Figure 5.3 Distribution of (dʒ) variants per age group (b).	160
Figure 5.4 Use of [gʲ] per age and gender (a).	162
Figure 5.5 Use of [gʲ] per age and gender (b).	162
Figure 5.6 Use of [gʲ] per age, gender, and locality (N=2,510).	165
Figure 5.7 Use of [gʲ] per speaker; the speaker codes from bottom to top: the speakers' initials, exact age, gender and locality (N=2,510).	166
Figure 5.8 Use of [gʲ] per speaker in the three age groups (a); N=2,510.....	170
Figure 5.9 Use of [gʲ] per speaker in the three age groups (b); N=2,510; J26MTH excluded.	172
Figure 6.1 The subgrouping of the Semitic language family (Rubin, 2008: 62).	188
Figure 7.1 Stages of the quantitative analysis of the DEF variable.....	213
Figure 7.2 Distribution of the dependent variants in the dataset.	218
Figure 7.3 Distribution of the dependent variable per locality.	219
Figure 7.4 Distribution of the dependent variable by age.....	220
Figure 7.5 Distribution of the dependent variable per gender.	221
Figure 7.6 Distribution of the dependent variable per speaker.....	224
Figure 7.7 Linguistic constraints on the DEF variable in this study.	232

List of Maps

Map 1 Topography of Oman (Och, P., 2000).....	6
Map 2 Political map of Oman (d-maps.com).....	7
Map 3 The Bāṭina Region showing the topography of the Bāṭina plain (Atlas of Oman)	8
Map 4 Tribal map of Oman (Wilkinson, 1987: 391).....	22
Map 5 Present day linguistic situation in Arabia (Holes, 2006b: 30).....	34
Map 6 Linguistic situation in Arabia at around 1000 A.D. (Holes, 2006 b: 28)	39
Map 7 Dialectal map of Oman; al-Suwaiq is marked as No. 5 in the map.....	41
Map 8 Location of al-Batha locality (Google Maps, 2021).....	89
Map 9 Location of al-Tharmad locality (Google Maps, 2021).....	90
Map 10 Location of the three localities in this study (Google Maps, 2021)	93
Map 11 Distribution of the [j] variant of /dʒ/ (Johnstone, 1965: 235).	111
Map 12 The distribution of the [j] variant of /dʒ/ in Oman (reproduced from Holes, 1989: 450).	115

Abstract

This work presents a variationist sociolinguistic investigation of the dialect spoken by the Bedouin *Yāl Sa‘ad* tribe living in the neighbouring towns of al-Suwaiq and al-Miṣin‘a along the Bāṭina coast in northern Oman. Holes (1989) classified al-Suwaiq as a ‘mixed’ dialect area where both Bedouin (B) and *Ḥaḍari* ‘sedentary’ (H) dialect types are used. In such ‘transitional’ or ‘border’ areas, a high degree of variation occurs where an H/B fusion seems to be the norm (ibid: 447). The thesis aims to further explore the sociolinguistic situation of this area. The focus of the thesis are two sociolinguistic variables: one is phonological, namely the (dʒ) variable, and the other is morphosyntactic, namely the definite article (DEF). These variables are analysed quantitatively using descriptive and multivariate statistics for a sample of forty men and women distributed across three age groups and three localities within the study area.

The multivariate results on (dʒ) show that the use of the traditional variant [j] is quite salient with an overall proportion of 71.9% in the (dʒ) dataset. The overall proportion of the incoming variant [gʲ] is 28.1%. Generally speaking, the middle age group leads both of the other age groups in the use of [gʲ]; men use it more than women, while older women are the most linguistically conservative group. Locality is not selected as a statistically significant predictor. In terms of the linguistic constraints, the use of [gʲ] is mostly favoured with a preceding coronal and in polysyllabic words; it is mostly disfavoured in monosyllabic words, and when preceded by a palatal sound.

On the other hand, the use of the traditional variant of the definite article, NULL, is very infrequent with an overall occurrence of 3% in the whole dataset, compared to the overt article *l-* (97%). Older women and al-Tharmad locality are the most linguistically conservative in the use of NULL, whereas the middle age group are the least users of this

variant. In terms of the linguistic nature of this variation, the results so far have emphasised that the cultural aspect of certain noun types influences the variation at hand, in that, generally speaking, tokens with the NULL variant in the speech of older speakers are core dialectal items that correlate with the Bedouin culture and traditional lifestyle.

The saliency of the first variable and the survival of the second in this transitional area are explained in the light of the prestige the former entertains in the *Yāl Sa‘ad*'s tribal territory which happens to be reinforced by that of the neighbouring Gulf states; on the other hand, the survival of the ‘marked’ variant NULL is attributed to the relative homogeneity in the demographics of certain localities which promotes a resilience to maintain the ‘Bedouinness’ of the dialect that is ‘indexical’ of the speakers’ Bedouin identity and heritage (Eckert, 2008; Eckert and Labov, 2017). Differences in the behaviour of men and women and the three age groups is explained in the light of contact these sub-groups are exposed to through social and geographical mobility, but generally speaking, men and the middle age group show the most variation in the case of the first variable and the least variation in the case of the second.

Acknowledgements

ماشي شراة العلم في هذي الحياة
ويوصل الإنسان مع يبغى ويريد
وياللي تعلم وارتفع في مستواه
يكتب على التاريخ والأمة سعيد
الشاعر الفقييد/ بحيت بن عبيد الشياي

There is nothing in life quite like learning
It takes people to where they aspire
Those who learn and enrich themselves
Are destined to be blissful amongst their people,
and throughout time. *Bakhit Obaid Al Sheyadi*

I am grateful to a number of people for the completion of this project, but firstly, I would like to sincerely thank my supervisor Professor Enam Al-Wer for her everlasting support, valuable expertise, and loving personality; thank you for believing in me. I am also indebted to the work of Professor Clive Holes on Oman and eastern Arabia which indeed is a valuable resource to all of us following his lead.

At Essex, I would also like to thank my supervisory panel members: Professor Peter Patrick and Dr. Ella Jeffries. Two colleagues have been a great asset in the making of this thesis: Dr. Uri Horesh and Dr. Maris Camilleri, who provided valuable feedback and offered a great insight to different parts of the thesis and to my conference papers. Uri also co-supervised the thesis for a while and was an advisor in my supervisory panel for most of the first part. In addition, I would like to say thank you to Professor Peter Patrick and Dr. Vineeta Chand for the enriching theoretical and practical modules on variationist sociolinguistics, and Dr. Wyn Johnson for the phonology modules throughout my first year, which have set up the

foundation for my adventure in the world of variationist sociolinguistics at the University of Essex. Special thanks also goes to the Arabic Sociolinguistics research group (2015-2018), and the Lexical Functional Grammar research group (2018-2019), both of which have been a source of inspiration, lively discussions, and valuable feedback. In particular, I would like to say a huge thank you to Dr. Hind Aloadini, Dr. Siham Rouabah, Dr. Mohammed Al-Rohili, and Dr. Wisam Alshawi, for their support throughout the different analysis stages. Hind has been constantly supporting me and pushing me forward since the first months of my arrival at Essex up until now; a thank you is never going to be enough. I would also like to thank the ladies at the departmental office (2015-2021) for their help with the administrative bit of my PhD program and my Colchester friends, both local and international for making my family's and my stay a memorable one. Finally, I would like to thank Professor Clive Holes (Oxford) and Dr. Faith Chiu (Essex) who examined this thesis for their valuable feedback and comments.

In Oman, I would like to express my gratitude to Sultan Qaboos University for funding this PhD project, and to my informants from al-Suwaiq and al-Miṣin'a for giving me the opportunity to conduct interviews with them and for being so welcoming and helpful, with special thanks to Rawan Al Rashidi, and to Alyaa and Nassir and their lovely children for sparking my love for their Bedouin dialect. I would also like to thank all of those who helped with arranging and/or conducting some of the interviews, with special thanks to Abdullah Al Sheyadi, Ruqayya Al Sheyadi, Hajar Al Sa'di, Tawadod Al Sheyadi, Idris Al Sheyadi, and Jum'a Al Sheyadi. I would like to also thank my friend and colleague Dr. Azza Al-Kendi for her unconditional friendship and support. My dearest Ghaida, my family, thank you all for your everlasting love, support, and prayers. Finally, my deepest gratitude goes to my husband, Bader, my children Albara, Alyazan, and Nawara, and my sister Sheikha and her family for standing by my side throughout the ups and downs. I can never thank you enough!

Some IPA symbols and their equivalents in the *Encyclopedia of Arabic Language and Linguistics* (EALL) as used in the thesis

IPA	EALL
ʔ	ʔ
θ	t̤
dʒ	j
ʒ	ʒ
ħ	ħ
ð	d̤
ʃ	ʃ
tʃ	č
s ^ʕ	s̤
d ^ʕ	d̤
t ^ʕ	t̤
ð ^ʕ	d̤
ʕ	ʕ
ʁ	ġ
g ^j	g ^y
k ^j	k ^y
ɬ	ɬ
j	y
ɪ	i
ʊ	u
a:	ā
u:	ū
i:	ī
e:	ē
o:	ō

The glossing of the Arabic examples in this thesis generally follows the Leipzig Glossing Rules (2008). Two notations not included in the former list, but which are used in thesis are: PART (particle), and IN (infix).

<blank page>

Introduction

Despite the rich linguistic past and present of Oman, there is a relative dearth in descriptive and linguistic literature on Omani Arabic. However, it seems that there is a growing interest by Omani and non-Omani linguists to explore this area in terms of different levels of linguistic enquiry, e.g., Holes's (1989-2016) survey and descriptions of Omani Arabic along with descriptive accounts of some dialectal features and varieties, Webster's (1991) remarks on the Āl Wahība dialect, Eades's (2009a), (2009b) and (2011) papers on the retention of the internal passive in a northern Bedouin variety, on the Šawāwi group of al-Jabal al-'Axdar, and on a transitional variety in the interior, Al-Aghbari's (2004) investigation of an aspect of the morphology of the dialect of Muscat, Davey's (2013) detailed sketch of the variety spoken in Ḍufār, Bettega's (2016) study on some aspects of the syntax of Omani Arabic, Al-Balushi's (2016) relatively recent dialectal survey of Omani Arabic, Morano's (2019) comparative account of the Banī Xarūš dialect in al-'Awābi, and Ambu Saidi's (2019) variationist investigation of the dialect of Nizwa immigrants living in the capital Muscat; this, in addition to relatively recent studies conducted and written in Arabic, e.g., Hamid *et al.*'s (2008) phonological study of the dialect of Nizwa, Al-Abri's (2002) study of the phonology of al-Ḥamra dialect, and Al-Salhi's (2015) description of the phonology and morphology of al-Ḥōqēn dialect. This thesis aims to contribute to this literature. It investigates sociolinguistic variation in the speech of the *Yāl Sa'ad* tribe living in the towns of al-Suwaīq (as-Suwēq) and al-Mišin'a along the Bāṭina coast in the northern part of Oman and thus, it also aims to fill in the mosaic of Arabic variationist sociolinguistic works for this part of the Arabic-speaking world. Two sociolinguistic variables are analysed in this study namely, the (dʒ) variable and the definite article variable (DEF), in light of age, gender, and locality as

external predictors and a set of internal constraints that are hypothesised to govern the variation at hand.

Research questions

Since this study is the first multivariate quantitative examination of the variation in this area, the general research objective is exploratory in nature in that it aims to shed the light on the sociolinguistic situation in this area and to situate the Bedouin dialect spoken in the area within the general descriptive literature available on Bedouin Omani varieties, but also within the Arabic variationist sociolinguistic scene. The focus of the analysis presented in this thesis is to understand the linguistic variation observed in the speech community under study in the light of the relevant intra- and extra linguistic constraints. Following are the main research questions that this study investigates:

1. What are the social parameters that drive sociolinguistic variation in the dialect, and to what extent can they help us understand the variation at hand?
2. What is the intra-linguistic nature of the variation in the use of the dependent variables chosen for analysis?
3. What is the effect of the transitional nature of the study area on variation in this speech community?

Outline of the thesis

The thesis is outlined as follows:

Chapter One focuses on the sociolinguistic profile of Oman and the study area. It provides an overview of Oman and the Bāṭina region in terms of topo-geography, history, demography, and socio-economy; it then introduces the study area. The chapter also includes an overview of the linguistic situation in Oman and provides an overview of the descriptive literature on Omani Arabic.

Chapter Two presents a description of selected phonological and morpho-phonological features in the dialect of *Yāl Sa‘ad*; the second part of the chapter presents a preliminary analysis of the internal passive and the negation system in the Bedouin dialect spoken in the study area.

Chapter Three presents the methods used in this research; it discusses the sample, the fieldwork, and the interview procedure, then lists the external and dependent variables in this study.

Chapter Four and Chapter Five provide an analysis of the first sociolinguistic variable, (dʒ). Chapter Four provides an overview of the descriptive and sociolinguistic literature on (dʒ). Chapter Five presents a quantitative analysis of the variation in (dʒ) in the *Yāl Sa‘ad* dialect. It 1) describes the coding procedure, the modelling process, and the results of the multivariate analysis of this variable, 2) presents the results of descriptive statistics of the external variables, and 3) provides a discussion of the results. The chapter ends with a summary and some conclusions on the status of (dʒ) in this speech community.

Chapter Six and Chapter Seven present the second variable, namely the definite article (DEF). Chapter Six presents an overview of the literature on definiteness and the definite article in Arabic; it also provides an overview of definiteness and a description of the null definite article in the dialect under study. Chapter Seven presents descriptive and multivariate analyses of the DEF variable in this dialect. The chapter ends with a discussion of the results and some conclusions.

The thesis ends with a summary and some reflections on the findings. It also highlights some of the setbacks and challenges and puts forth some recommendations for further research.

Chapter One: Sociolinguistic profile

This chapter presents a brief background on the topo-geographic, demographic, socio-economic profiles of Oman and the Bāṭina region in which al-Suwaiq and al-Miṣin‘a are administrative districts. Then, it provides a background of the geographic, demographic, and socio-economic profiles of al-Suwaiq and introduces the speech community under study. The second part focuses on the ethno-linguistic diversity in Oman. It provides a brief description of the history, tribal structure, and ethnolinguistic minorities in Oman, and then presents some previous literature on Omani Arabic, with special focus on the description and classification of Omani Arabic as a distinct dialect group, along with the main sub-classifications of this variety.

1.1 Oman and the Bāṭina Coast: A brief background

1.1.1 Location and topo-geography

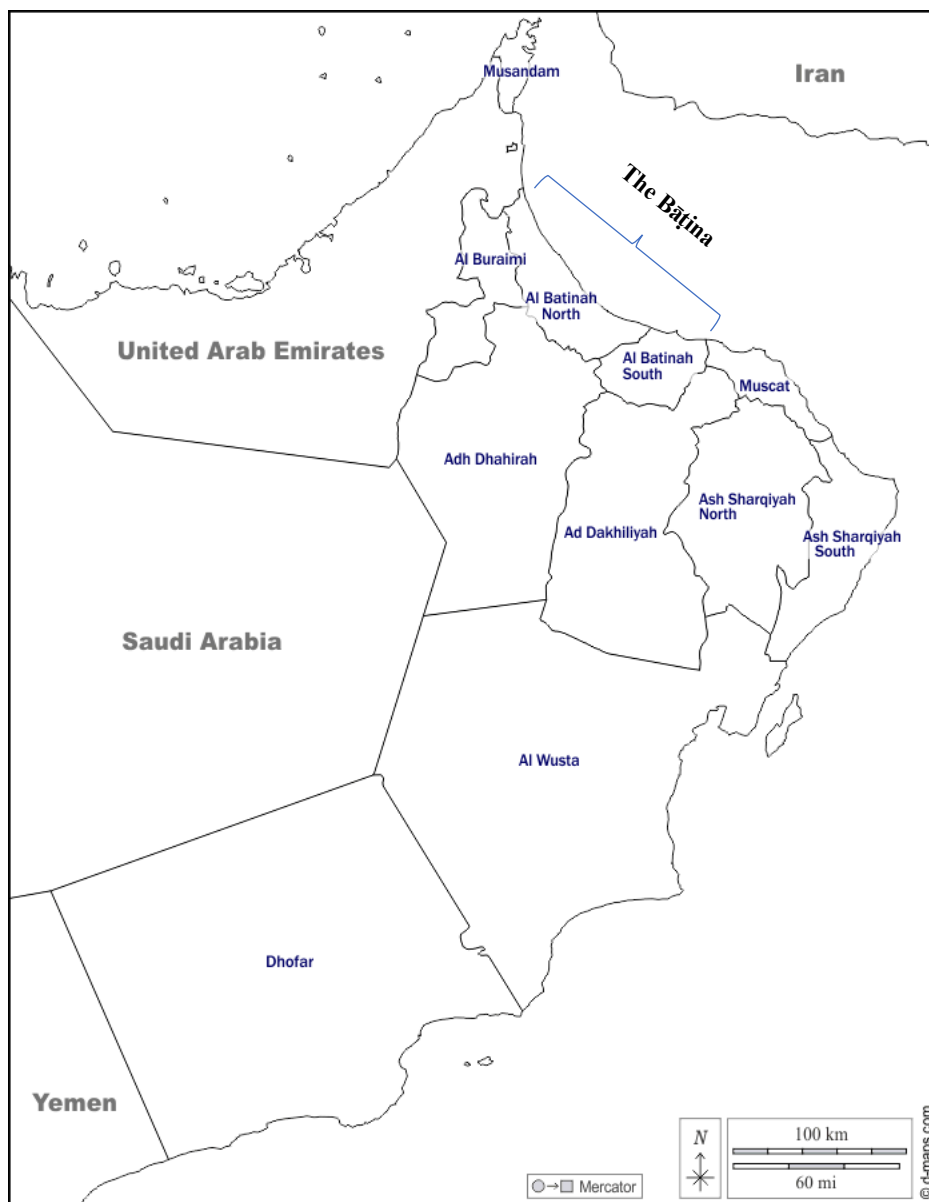
Oman is located on the southeast corner of the Arabian plate and overlooks the Arabian Gulf, Gulf of Oman, and the Arabian Sea. Covering a total land area of 309,500 Km², Oman is the third largest country in the Arabian Peninsula (Ministry of Information, 2010: 42). The Sultanate of Oman shares borders with Yemen from the south and the south-west, Saudi Arabia from the west, and the United Arab Emirates (UAE) from the north and north-west. The location of Oman is strategic because it is situated at the edge of the Arabian Peninsula with a 3,165-kilometre coastline facing the Indian Ocean, and the Arabian Sea, as well as political and commercial control over the opening to the Arabian Gulf at the Hurmuz Strait in Musandam Peninsula (ibid).

The topo-geography of the Sultanate is quite diverse, including deserts, mountain ranges, coastlines, and fertile plains. These terrains are a result of over 800 million years of geological and climate changes (Ministry of Information, 2006: 12). The area can be geographically divided into four terrains (Map 1): the Ḥajar mountain range and mountain foothills, the Bāṭina coastal plain, the southern region of Ḍufār, and the arid desert plains dividing the south and the north of Oman (Ministry of Information, 2010: 42-49). The Ḥajar mountain chain extends from the northwest towards the southeast of Oman. It is 650 kilometres long and about 130 kilometres wide, with an altitude of 3,000 metres at its highest point in Jabal Shams in the hinterland. The Ḥajar mountains are the backbone of northern Oman; resembling a human backbone, this natural barrier separates the Bāṭina coastal plain on the inside of the range from that of the desert in the outside (Wilkinson, 1987: 21; Ministry of Information, 2010: 43). This mountain chain is divided by a natural gap into two main parts, the eastern Ḥajar and the western Ḥajar and ends in the northmost part of the Sultanate, namely the Musandam Peninsula which is located at the entrance of the Arabian Gulf, and from which, Iran to the north is only a few kilometres away (see Map 1). Musandam's coast is the only coast that the Sultanate shares with the Arabian Gulf. Furthermore, the desert part of Oman is an extension of the deserts in central Arabia. It features diverse landscapes ranging from salt flats, oases, and sand dunes, most famous of which are the Wahība Sands. At the southern end of the Sultanate, Ḍufār is considered as a separate topographical area with its southern mountain range which includes Jabal Samḥān, Jabal al-Qamar, and al-Qara mountains (Ministry of Information, 2010: 42-49).



Map 1 Topography of Oman (Och, P., 2000).

Politically, the Sultanate of Oman has four governorates, namely Muscat, *Ḍufār*, Musandam, and al-Buraymi (local: li-brēmi), and five administrative regions: the Bāṭina, *Ḍāhira*, *Dāxiliyya*, *Šarqiyya*, and *Wustā* regions (see Map 2). The Bāṭina and the *Šarqiyya* regions are administratively further divided into Bāṭina North, Bāṭina South, *Šarqiyya* South, and *Šarqiyya* North.



Map 2 Political map of Oman (d-maps.com).¹

The Bāṭina region is where the study area is located. This region extends from Barka in the southeast to Šināš towards the northwest. It is situated on the Bāṭina coastal plain, an open area which is 237 kilometres long and 25 kilometres wide, and which extends from Rās al-Ḥamra in Muscat in the south to Xaṭmat Milāḥa on the borders with the UAE towards the North. The Bāṭina plain is a fertile plain since it has been created by soil brushed away by

¹ Available at: https://d-maps.com/carte.php?num_car=27733&lang=en

erosion and deposition of the eastern Ḥajar mountain flanks through *wādīs* (creeks and inlets) created by rainwater,² which made the area suitable for farming activities and agricultural growth. The area's physical geography is described as a “low-laying plain, sandy towards the sea, clayey in the interior, and stony as the hills are approached” (Lorimer, 2003: 6), as can be seen in Map 3.



Map 3 The Bāṭina Region showing the topography of the Bāṭina plain (Atlas of Oman)

² (Och, P., 2000: 20).

1.1.2 Population

Most of the population in the northern part of Oman is concentrated around the Ḥajar mountain chain, and in the south in the city of Ṣalālah in Ḍufār; the rest of Oman is relatively sparsely inhabited. In the Bāṭina region, the human settlements were established around the water lines, palm farms, and wells, whereas the coastal villages merged together (Wilkinson, 1987: 28; Lorimer, 2003: 7). It is noteworthy that most of the population in Oman is settled, yet there are different groups of nomadic, and semi-nomadic communities, especially in the desert parts, and some transhumant groups³ in the mountains in both northern and southern Oman (Peterson, 2004: 259). The latest population statistics show that the current population of Oman is 4,507,323 (Omani nationals, N=2,751,842; expatriates, N= 1,755,481) (NCIS, 2021: 3),⁴ compared to 2,773,479 in 2010 (Census Administration, 2010: 5-7). The Bāṭina region (north and south) is second in terms of population density (N=1,259,877) after the capital Muscat which is the most populated area in Oman (NCIS, 2021: 3). In the early 1900s, Lorimer estimated the settled population of the Bāṭina region at 105,500 (2003: 7)⁵⁶.

1.1.3 Economy: before and after the discovery and production of oil

In the days before the discovery of oil, Oman's economy depended on agriculture, maritime trade, fishing, and animal husbandry as the main economic activities. Agriculture and cultivation of crops centred around water resources, like creeks, springs, and water canals.

³ Eades (2009b) on the dialect of these groups.

⁴ The National Centre for Statistics and Information. Available at: https://www.ncsi.gov.om/Elibrary/LibraryContentDoc/bar_22%20April%202021_2027226f-88f3-4304-ac41-8b025c8dc585.pdf, last accessed [25-04-2021].

⁵ Cf. Miles (1966: 380) who stated that the population of the Bāṭina in the early 1900s was about 700,000 inhabitants which is probably overstated.

⁶ The Bāṭina plain had not always been heavily populated. The area was subject to invasions via the coast [mainly Persian invasions], or to the rainstorms and hurricanes that killed a lot of people (Al-Sa'di, 2015: 14-18). Historical references show that in 865 A.D. and during the reign of ṣ-Ṣalt bin Mālik, there was a heavy rainfall and a lot of the people died; the date harvest and farms were destroyed; the Bāṭina plain was not populated for years afterwards (ibid: 15).

Whereas Omani dates had been exported to different places, other crops like mangoes, bananas, watermelons, and vegetables were just enough to sustain the areas that grew them. Pomegranates, figs, cereals, sugar canes, olives, walnuts, and other types of produce were also grown in mountain areas (Al-‘Āni, 1999: 46). Due to its rich, fertile soil, the Bāṭina region has been one of the top producers of dates and vegetables which are sold in Omani markets and exported outside the country (Ministry of Information, 1995: 12). The Bāṭina coastal strip has also played an important role in the region’s economy, since fishing and maritime trade have long been the main pillars of the economies of the coastal areas and ports, including those along the Bāṭina plain. In the past, small fish, like ‘*ūma* ‘sardines’ [and *qāṣi* ‘anchovies’], were taken to the interior as fertilisers and as food for animals, while other types of dried fish were exported (ibid). In addition to fishing, diving for pearls was another activity that Omanis practiced in the past (ibid; also in Wilkinson, 1987: 21), particularly in Sūr and Muscat, which exported pearls to other places in Arabia (Al-‘Āni, 1999: 46-47).

Another important aspect of the economic make-up of the Bāṭina region is the fact that it has always been a vital mining site with a history of mining that goes back to 5,000 years ago when products were sold in the markets of Mesopotamia (Ministry of information, 1995a: 13). Copper and other minerals have been long produced due the abundance of such natural resources in the area, particularly in Ṣuḥār (ibid). In addition, given its strategic location as the coastal gate of the peninsula, Oman facilitated trade of foreign goods from India and China with the rest of the peninsula and Iraq (Ministry of Information, 1995b: 20). The exportation of frankincense, camels, and horses also contributed to ancient international trade with Oman. Omani frankincense ‘*lubān*’, which is grown in the mountains of Ḍufār, was exported to other areas in the Arabian Peninsula and Europe (Al-‘Āni, 1999: 52). The Bāṭina region has always facilitated local trade by bridging the capital and the rest of Oman with other Arabian Gulf countries (Ministry of Information, 1995a: 12).

Furthermore, after the discovery of oil and natural gas reserves, the economy and industry in Oman boomed. Since the first production of Omani oil in 1967, Oman started massive developmental, industrial, and urbanisation projects. This discovery has contributed to transforming the infrastructure in Oman and the lifestyle of the population in general (Ministry of Information, 2010: 240-244). As far as the Bāṭina region is concerned, one of the developmental cornerstones is the Ṣuḥār industrial estate, a large project that has facilitated international trade with countries of the Indian Ocean, Iran, China, and the rest of the world (Ministry of Information, 1995a: 12). Many major plants for petro-chemical industries have started their production since the 2000s. Another main project is the Marine Sport City that was inaugurated in 2010 in al-Miṣin‘a, and which regularly holds international sport events and championships. In addition, the Bāṭina coastal road, and the express way are two major infrastructural developments in the region. While the former is largely still under construction, the latter is almost completed and has been another primary route linking Muscat to the Bāṭina region.

1.2 Al-Suwaiq: Geographic, demographic, and socio-economic structure

This section gives a brief background on the sociolinguistic situation of al-Suwaiq where most of the informants interviewed for this study come from.

1.2.1 Location

Al-Suwaiq is a coastal town in the middle of the Bāṭina region, located on the coast of the Gulf of Oman with an area of 3,840 km². The town’s administrative boundaries extend to eighty kilometres long and over forty-five kilometres at its widest point (see Map 3). It is locally known as *s-suwēq* ~ *li-swēg^y* ~ *swēg^y*⁷ which is a diminutive form of the word *sūq*

⁷ The first is a Ḥaḍari (sedentary) and the other two are Bedouin local pronunciations of the toponym.

‘market’, since the town has always been a commercial site known for its local and maritime trade. People from neighbouring towns and mountainous areas of the inland have for so long been in contact with the town as their main market where they buy goods, dates, and fish and sell mountain produce and herbs. In addition, al-Suwaiq have always been associated with camel husbandry, and so, this name may also mean a place where camels are raised, derived from the verb /ja-su:q/ ‘3-ride.IPFV.SGM’ (Ministry of Information, 1998: 11; Al-Shibli, 2010: 25-28). The town is bordered by al-Miṣin ‘a town toward the south and southeast, al-Xābūra town toward the north, and al-Rustāq town and al-Ḥōqēn locality toward the west. There are a hundred and one villages within the town’s administrative boundaries, starting with Ġurfat ‘Āl Ḥamad in the south to al-Ḥujayra (local: li-Ḥyēra) in the north.

Furthermore, there is a highway linking al-Suwaiq directly with the capital Muscat which is about 135 kilometres to the southeast and the city of Ṣuḥār, the international industrial port and historical city in the Bāṭina, which is about 100 kilometres up north. This double highway is part of the main and oldest highway that links southern to northern Oman and then the UAE in the north. With regards to the newer major routes in the region, as far as the study area is concerned, while the new coastal road is still largely under construction, the express way is operating. The latter is meant to be a ‘non-stop’ route to and from Muscat and mostly used by people commuting for work.

The topo-geographic nature of al-Suwaiq is diverse, reflecting that of the region. In addition to the 45-kilometres long coastline, there is an agricultural strip along its plain area immediately after the coastline, then a sandy area (locally known as *sēḥ* or *syūḥ.PL*), and a mountainous area at the foothills of the Ḥajar mountain chain (Ministry of Information, 1998:

11-12).⁸ The plain is the most populated area and is up to ten to fifteen kilometres wide. The land around this area is fertile due to the abundance of underground aquifer water, which is extracted by wells, one of which is locally known as *tawi* or *bidi*. This led to the expansion of agriculture along the plain. Many types of crops including mangoes, tomatoes, potatoes, watermelons, and a verity of vegetables, in addition to date palms are grown in this area. On the other hand, the desert part is a ten to twenty kilometres wide flat land with very little vegetation and few inhabitants; however, it has been important for some of the Bedouin element of the town's population who depend on camel and cattle since the area becomes quite pastoral after rainfall, and during winter (ibid: 15).

The mountainous part towards the south of the town is an extension of the larger Ḥajar mountain chain; it is locally known as *ḥgūr* ~ *ḥg'yūr* ~ *ḥyūr*. It is about ten to eighteen kilometres wide. It encloses a chain of *wādis* 'inlets', water springs, and 'aflāj (*falaj*.SG) 'water canals'. Around these water resources, the density of the mountain population increases, as the case with the Bāṭīna region in general. In addition, lots of dates are grown in this area along with some mountain crops like pomegranates, lemons, and grapes (Ibid: 15). Another important geographic characteristic of al-Suwaiq is the major *wadīs* flowing from the western Ḥajar mountains and from Rustāq, which meet the sea in al-Suwaiq; when rain comes some of these inlets overflow (Z'ētar, 2014: 12). In addition, the coastline which stretches to forty-five kilometres long and ten to fifteen kilometres wide is home to the main port and the main market in the town.

⁸ Al-Shibli divides the town into three main areas, namely the coastal area (including the agricultural strip), a plain (Bedouin) area, and a mountainous area (2010: 135-40). The current study is situated within the first area.

1.2.2 Demography

Al-Suwaiq is one of the most populated towns in the region, and according to the National Centre of Information and Statistics, it was the second top town in the number of live births in 2015.⁹ In the 1993 census, the town's population was 74,000 people; it increased to 80,000 in 1998 (Ministry of information, 1998: 16). There was a considerable increase of about 30,000 in the population by 2010. According to the 2010 census statistics, al-Suwaiq was the most populated town in the Bāṭina region with a total of 111,711 out of 772,590 residing in the region, with 85.4% Omani nationals, and 14.6% expatriates (Census Administration, 2010: 16). Omani men and women were almost equally distributed with 50.4% and 49.6% respectively (ibid).

Lorimer's (2003: 6) description of the town is quite significant because a similar image of the demographic (and linguistic) diversity is characteristic of the town today.¹⁰ Whereas the coastal villages closer to the coast tend to be generally ethnically and socially diverse along the Bāṭina coast,¹¹ communities who live farther inland of the area seem to be less diverse. Al-Suwaiq's maritime history, strategic location, and natural resources attracted many people from farther inland in the adjacent mountain foothills and from neighbouring towns who moved closer to the town's centre in search for better lives (Ministry of

⁹ According to the 2015 statistical report on births and deaths in the Sultanate, al-Suwaiq came second in the Sultanate (after al-Sīb in Muscat) (National Centre of Information and Statistics, 2016: 30).

¹⁰ When describing the town at the beginning of the last century, Lorimer wrote that al-Suwaiq is a city and a port on the Bāṭina coast. It is fourteen miles far from al-Miṣin 'a which borders it from the south and the south-east, and 22 miles away from al-Xābūra bordering it from the west and north-west. It is a relatively populated area with about six hundred houses and huts. The people inhabiting the area are *Ṣawāliḥ*, and *Baluch*, and they live on fishing and farming. There are also six families of Khojas and seven [merchant] families of Hindus. al-Suwaiq is one of the ports in which Wādi Banī Gāfir ends. It has ten *baqarāt* and five *badanāt* 'large commercial ships' that sail across the coast from Muscat to other Gulf ports, in addition to twenty fishing boats. There are about eight thousand and four palm groves which are watered through wells that use aquifer water. About a hundred camels, ninety donkeys, a hundred and fifty cattle, and four hundred sheep and goats are owned by the people of al-Suwaiq. It seems that Lorimer's description probably applies to the main market area or the centre of the town.

¹¹ Brockett (1985: 3).

Information, 1998: 11). Therefore, it is expected to find a varied tribal structure.¹² However, *Yāl Sa‘ad* (also known as *Āl Sa‘ad*; singular: *Sa‘di*) is the major tribal element in the town (See Map 4). *Yāl Sa‘ad* is one of the Nizarite tribes in Oman, and along with the Hawāsina, in the neighbouring town of al-Khabora (l-Xābūra), they represent the two main and largest tribes in the Bāṭina coast, *Yāl Sa‘ad* being the largest in al-Suwaiq and the bordering town of al-Miṣin‘a (Lorimer, 2003: 6; Al-Siyābi, 1964: 11). *Yāl Sa‘ad* trace their ancestry back to Sa‘ad bin Baker bin Hawazān bin Muḍar bin Nizār bin Ma‘ad bin ‘Adnān. They are powerful and rich and, in Oman, mainly concentrated in the Bāṭina region, and the sheikhs of this tribe are Āl Ḥamad bin Hilāl, who have been dwelling in al-Suwaiq town for many decades (Al-Siyābi, 1964: 11).¹³ Some historical references indicate that this tribe must have been in the area for over three centuries (Al-Sa‘di, 2015: 30).¹⁴ A quick survey of Al-Sa‘di’s (2015) first volume shows that this tribe has been actively involved in the politics of the study area in particular and northern Oman in general. There are a number of sub-groups or families that belong to this tribe distributed over different locations in the area. The tribe’s traditional territory extends from *Wādi G‘āsim* in al-Miṣin‘a to *li-Ḥyēra* to the northwest of al-Suwaiq, an area which is locally known as *Miṣhāb Yāl Sa‘ad*. One of my informants said that the name *miṣhāb* here means a tribal confederation.¹⁵ In addition to *Yāl Sa‘ad*, some of the main

¹² Some tribes that live in the area along with *Yāl Sa‘ad* are: ‘*Āl Bū Ršēd*, ‘*Āl Bū Grēn*, and the *Ṣawāliḥ*, *Jahāwir*, *Kiyūmiyyīn*, *Xarūṣiyyīn*, ‘*Awāmir*, *Mšāfra*, *Ṣrūḥ*, *Sawābi‘*, ‘*Gawārib*, *Qanānba*, *Šikēliyyīn*, *Jawābir*, *Masā‘id*, *Hanādīs*, *Šiyādiyyīn*, *Ma‘āwil*, *Na‘āmīn*, *Nawāfil*, *Maxāmra*, *Mana‘*, *Manāwra*, *Hinā‘iyyīn*, *Ġasāsna*, *Baluch*, and *Zidjāl*, just to name few (Z‘ētar, 2014: 20-21).

¹³ *Bin* means ‘son of’.

¹⁴ In these sources, *Yāl Sa‘ad* are described as ‘*Arāb Sāḥil l-Bāṭina*, i.e., the Arab of the Bāṭina coast (ibid: 30-31). These Arabs=Bedouins have date farms and fields on which they grow wheat, sugar cane, cotton, and cloves (ibid: 70-71).

¹⁵ Local Bedouin pronunciation of the toponyms here. Many of my informants contend that these are the borders of the tribe in the study area at the present.

tribes in *Miṣḥāb Yāl Sa‘ad* include ‘*Āl Bū Ršēd*, ‘*Āl G‘arād*, ‘*Āl Brēk*, ‘*Āl Hilāl*, and the *Hadādba* (Al-Sa‘di, 2015: 31).¹⁶¹⁷

1.2.3 Socio-economy

Traditionally, the people of al-Suwaiq were and many still are fishermen along the coast, farmers along the fertile plain, sheep herders and farmers in the mountainous area, or cattle and camel breeders in the sandy inland. The town has always been one of the biggest markets in the Bāṭina region. There are three main markets in the town, namely *it-Tarmad*, *li-Bdāya*, at the two edges and *Sūq s-Suwēq* ‘al-Suwaiq market’ at the centre of the town. People from neighbouring towns come frequently to these markets (Z‘ētar, 2014: 16), and in addition to Ṣuḥār and Ṣaḥam, and Barka along the Bāṭina coast, al-Suwaiq is reported as a main supplier of fruit and dates which other towns are dependent on (Al-Siyābi, 1964, via Al-Shibli, 2010: 29; Phillips, 1967: 35). Small farms watered through water aquifers still abound and the town is one of the main suppliers of dates in Oman.¹⁸ According to a Bedouin informant, in addition to farming, the Bedouin community breed camels and cattle for a living.

In addition, as a historical port, though smaller in comparison to other ports in Oman, the town traditionally played an important role in maritime trade between Muscat, Ṣuḥār and the Persian Gulf (Lorimer, 2003: 8). It maintained central control over exported goods (rice,

¹⁶ Ditto.

¹⁷ Lorimer (2003: 455-56) described the tribe as a large tribe of about 13,000 members who belong to six different branches. They occupy a 25-mile-long piece of land towards the eastern part of the Bāṭina; some of the coastal areas within their territory are also inhabited by other tribal elements in the region. Their territory’s traditional border is a low lying inlet, part of the larger Wādi Bani Xarūs, which divides their territory and thus is called Wādi Qāsim [locally: Wādi G‘āsim]. Their western borders are in the Xaḍra area. The villages that belong to this tribe are Maṣan‘a, Ṭuwi š-Šāwi, Ṭirēf, Miladda, Ṣūr al-Garaṭ [Ṣūr al-Qarat], Tarmad, Raggāṣ, Ġrēfa, Baḥat Swēg [al-Baḥa], Xaḍra and Ṣūr Ḥayyān [Ṣūr Hayyān], the first and the last three are coastal villages. *Yāl Sa‘ad* belong to the Ibadhite sect; they are well-off since they own hundreds of thousands of palm groves, a large amount of farming land, and many commercial and fishing boats; *some of them* are Badu [Bedouin; emphasis is mine].

¹⁸ E.g., the farms in the plain and the mountainous area in the town produced 12.3 thousand tonnes of dates in 2015 (Ministry of Agriculture and Fisheries, 2015: 17).

spices, coffee, sugar, salt, ... etc.) and local produce of fish in the immediate area through a number of landings (locally known as *manāyil*) (Z'ētar, 2014: 43-46). Fishing and seafaring have traditionally been an important aspect of the socio-economy of town (§ 1.2.2). In addition to providing fish for local consumption, sardines, dried and salted fish have been used by the Bedouin communities in the area and by people living in the mountain areas as food for animals at periods when there was a shortage of 'alaf' 'fodder' (ibid). Until recently, one scene of the beaches of the villages near the fish market in the centre of the town was the square-shaped plots of dried sardines and anchovies.

Furthermore, before the 1970s, there was barely any infrastructure in the town. There were no schools; the only form of education was through local Quran circles. Children were taught Quran, and the basics of Modern Standard Arabic (MSA) by a *m'allim ~ m'allam* 'a learned person' who mostly would not have been officially trained or educated.¹⁹ However, the socio-economic scene of the study area has been influenced by the major infrastructural and socioeconomic developments that have taken place since the 1970s, as part of the development to the whole country after Sultan Qaboos succeeded to the throne of Oman. Main roads, schools, government and private establishments, healthcare centres were established in the area, and have been expanding since then. The main markets have also been expanding since then. Today, the main market in the town has become an even more important part in the town's economy, with larger hypermarkets and big brands opening branches there. This is in addition to the expansion and restructuring of the main port and fish market nearby the main market, which has revived the fishing sector and maritime trade in the area with commercial ships, mainly from Iran, landing in this port.

¹⁹ Based on an interview I conducted in 2015 with a Ḥaḍari (sedentary) informant from the area.

These developments also brought along an improvement in the socioeconomic situation of residents; their lifestyle and work patterns have been generally influenced by these developments. It can be argued that the demographics of al-Suwaiq and the larger Bāṭina region have been also influenced by the developmental projects in the area. For example, the newly Bāṭina coastal road is constructed along the beaches in the region, and in the process of building this road, many houses in the immediate adjacency of the coast have been demolished and the residents have moved to live in other areas in the town. Some moved to privately owned houses close to where they originally lived or chose to live in government housing units that were established in their immediate villages; however, a considerable number of families relocated farther inland. For instance, a lot of families in the town centre relocated to the villages on the westside of the highway which cuts the town centre in half. This area has also been a place where many younger families settle. In addition, the recently established industrial area in the western part of al-Suwaiq is becoming an important aspect of the socio-economy of the town; a lot of Omani families settled in this part of the town.

In terms of education, there were two main schools that started operating in the academic year 1971-1972, namely Halima Al-Sa‘diyya for girls and Al-Warith bin Ka‘ab for boys. There were 583 students in twelve classrooms, with twelve teachers only in both schools. By the academic year 1997-1998, there were twenty-nine schools, fifteen for boys and fourteen for girls. The number of students considerably increased with a total of 27,398 students distributed across 744 classrooms and taught by 1,089 teachers for the same academic year (Ministry of Information, 1998: 51). However, by the academic year 2015-2016, the number of schools in the town increased to forty schools, and the number of

teachers doubled (2,482), with a student-teacher ratio of 11, according to the National Centre for Statistics & Information (2016: 22).²⁰

In addition, adult literacy and education programmes in Suwaiq started in the academic year 1973-1974, in Al-Warith bin Ka‘ab School, with 129 students (88 men, and 41 women) in three classrooms. However, this number increased to 601 students across eleven centres by the year 1997-1998 (Ministry of Information, 1998: 51). In addition to government schools, there are a number of local private schools, offering kindergarten and primary education; the first private school started in the academic year 1989-1990 (ibid: 51). This number has increased, and two international schools which offer bilingual primary education have been established since then.

1.3 History and tribal structure

1.3.1 Ancient Oman

Archaeological investigations in different areas have concluded that some ancient civilizations and settlements have long existed in Oman, as far as the sixth millennium B.C. (Ministry of Information, 1995b: 20). At one point in Oman’s history, the Phoenicians came and settled in Şūr in eastern Oman, and after they left, the Assyrians came. The latter were replaced by the Akkadians during whose reign Oman’s economy and trade had significantly prospered (Al-‘Azīzi, 2014: 71). In addition, scientific and archaeological reports have stated that ancient settlements with organised socio-economic and political activity were present in the Bāṭina region; these societies developed trade relations with neighbouring communities (Ministry of Information, 1995b: 20). Other discoveries concluded that ancient Oman was

²⁰ *General Education Statistics Report: 2015-2016*. Available at: https://www.ncsi.gov.om/Elibrary/LibraryContentDoc/bar_School%20Education%202015%20-%202016%20_aea9e37d-d67e-4d15-9f4f-c0d029b1a4d1.pdf, last accessed [30-12-2016].

inhabited by the Sumerians who first started the copper mining in northern Oman (Al-'Azīzi, 2014: 70). It is noteworthy that the ancient Majān, 'the land of copper' civilization, expanded sometime around the first millennium B.C. to include the southern parts of the peninsula from the Strait of Bāb al-Mandab in Yemen to Hurmuz Strait in Oman. In fact, remains of copper mining sites in the mountainous part of the city of Sohar provide evidence that copper mining in the area started as early as the third millennium B.C. (Ministry of Information, 1995b: 20-23).

1.3.2 The Omani Empire and maritime influence

Oman has long been an important centre for trade and maritime influence. Oman's vessels had sailed between the Indian Ocean and the ports along the Arabian Gulf coast for thousands of years. Omani sailors are well-known for their skills and knowledge of these routes, and of one the best Arab sailors, the Omani Ahmed bin Majid, helped the Portuguese explorer Vasco De Gama to navigate across the Indian Ocean to the Indian coasts (Al-Ṭā'ī, 2008: 8). In addition, Oman occupied the east coast along Zanzibar, Mombasa, Kenya, and then expanded inside and reached up to Congo and Rwanda (Ibid, 2008: 10). Many Omani minorities still live in these African countries, and still speak Arabic and practice Islamic and Omani traditions.

1.3.3 Tribal structure and the Arabisation of Oman

It appears that the first 'Qaḥṭāni' Arab existence in Oman dates to the time when the Yemenite 'Umān bin Qaḥṭān was appointed a governor on Oman (Al-'Azīzi, 2014: 71). Al-Ṭā'ī also wrote that some of the ancient Arab tribal settlements in Oman were Banū Tasam in the Jaww area in the Ḍāhira region, Banū Judays in Tuwām (Buraymi), and Māzin of 'Azd in the coast of the Gulf of Oman and along the coast of Trucial Oman (UAE). The inland was inhabited by Banū Riyām, in al-Jabal al-'Axḍar area, in addition to Banū Ḥadīd, Banū

Quḏā‘a, and Banū Tamīm tribes.²¹ On the other hand, Ja‘lān, in the eastern region, was inhabited by ‘Abdu l-Qays tribe. The tribe of Muḏar settled in Dabā, at the northernmost end of Oman, and later they founded the city of Musandam (2008: 6-7).

Nevertheless, historical references on the existing Arab tribes of Oman seem to agree that the Arab element of the population in the area largely originated from two primary tribal origins, namely the Yemenites or ‘Azd ‘Umān, and the Nizarites or Nizār²² (e.g., Ibn Rāziq, 2010: 3; Wilkinson, 1987: 73-78; ‘Āšūr, 1980: 26). These tribes are also the two main tribal origins of eastern Arabia (Holes, 2016: 6-7). The Qaḥṭāni ‘Azd is one of the largest Arab tribes and are larger in number than the Nizār (‘Āšūr, 1980: 18). They established earlier tribal settlements in Oman, which date back to as early as 100 A.D. when the first wave of immigrations from south-west Arabia across the southern end of the Empty Quarter desert into Ja‘lān and from Bahrain through the coast of Trucial Oman, into Buraymi started.²³ These tribes brought along their tribal alliances with whom they developed new ties during immigration (Wilkinson, 1987: 73).²⁴ In addition, it is believed that the main migration and later settlement of the ‘Azd of ‘Umān in the area is dated back to the sixth century A.D., when most of the Arab tribes in Saba’ (Sheba) in Yemen had to migrate in the aftermath of the collapse of the Ma’rib dam in search for better lives. Those led by Mālik bin Faham were the first ‘Azd group to settle in Oman, through the route along southern Oman (Wilkinson, 1987: 75). They were part of ‘Azd Al-Sarāh who after migrating to Oman came to be known as ‘Azd ‘Umān,²⁵ ‘Umān being the name of their homeland back in Yemen (Al-Sālmī, 1347:

²¹ These are apparently non-‘Azd tribes, rather some allies.

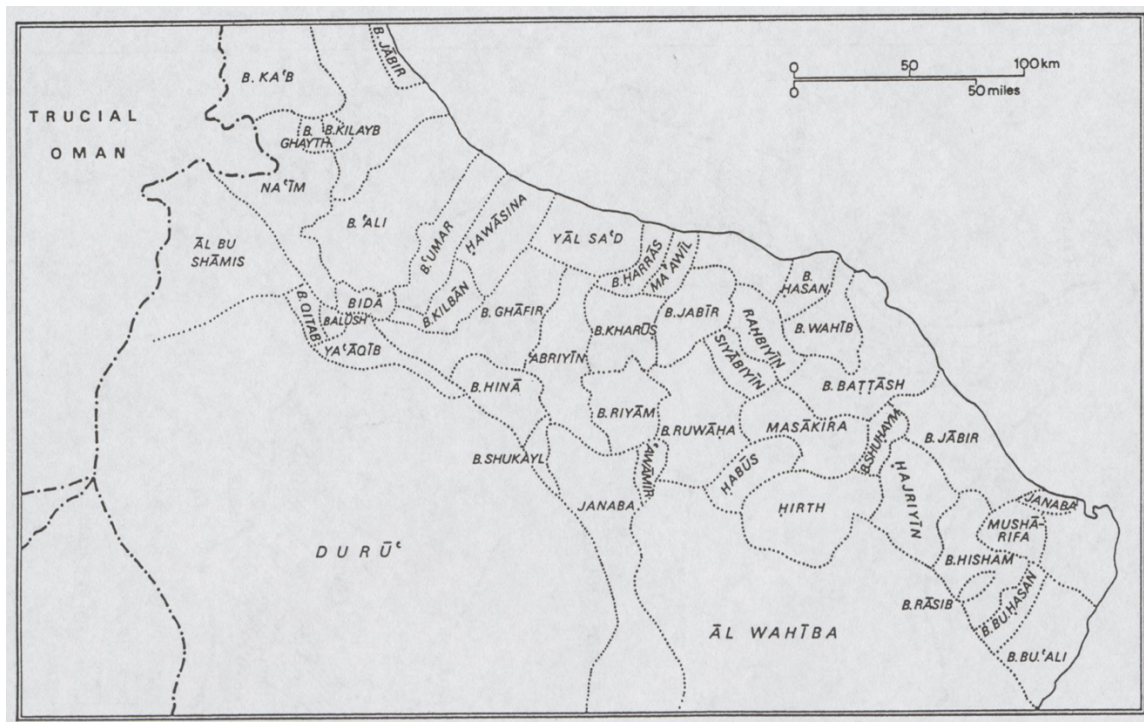
²² Banū Naḏar.

²³ In 120 A.D., after the first rupture of the Ma’rib dam, according to Badger’s preface to Ibn Rāziq (2010: VI).

²⁴ See Wilkinson for a detailed account written in English of the tribal composition of Oman’s population, their settlement patterns, their religious affiliation, and social structure (1987: 72-145).

²⁵ Further details on the exact events of why Malik bin Fahm left his people can be found in Al-Salmi (1347: 10) and ‘Āšūr (1980: 15-16).

19; ‘Āšūr, 1980: 15-16).²⁶ Mālik bin Faham is an important personality in the history of the area since he defeated and expelled the Persians who occupied the coastal line in northern Oman at that time (Al-Sālmī, 1347: 14-20; ‘Āšūr, 1980: 16-25; Al-Ṭā’ī, 2008: 7-8). On the other hand, the Nizarites are said to have originated from the ‘Abdu l-Qays which is part of the bigger Banū Rabī‘a tribe, the descendants of the ‘Adnāni Naḍar bin Ma‘ad, and who originally came from Ḥijāz through Najd. ‘Abdu l-Qays migrated to Bahrain; some migrated to Oman (Ibn Rāziq, 2010: 3). Map 4 presents a sketch of rather recent main tribal territorial concentrations in Oman, which are also listed in Lorimer’s *Gazetteer of the Persian Gulf* (1908, 1915).²⁷



Map 4 Tribal map of Oman (Wilkinson, 1987: 391)

Keeping this history in mind, it is no wonder that the demographic/ethnic and linguistic composition of the Omani society as we see it today is diverse.

²⁶ At the time of Persian rule, Oman used to be known as ‘Mazūn’, i.e., the land of rain (‘Āšūr, 1980: 25; Al-Sālmī, 1347: 19). More on the etymology and records of the word ‘Umān can be found in Al-‘Azīzi (2014: 71-72).

²⁷ Lorimer (2003).

1.4 Linguistic profile and dialect topography of Oman

Commenting on the linguistic situation in the Gulf area, Holes stated that, generally-speaking, modern Gulf dialects have witnessed similar linguistic influences (2006a: 210). Persian is a contributor to the vocabulary in the dialects, in addition to South Asian languages, mainly Hindi ~ Urdu. Many immigrants from Iran passed on Persian to their children and grandchildren who may have no knowledge or some functional knowledge of Persian due to a long-term assimilation into Arabic (ibid). The same is the case with the Baluchi community in the Gulf. Baluchi is mainly spoken at home and in-group conversations (ibid: 211). Also, in the Omani context, Omanis (or Zanzibari Omanis) who returned from East Africa and other Arab and Gulf states before and after Sultan Qaboos came to power in Oman came back mostly well-educated, with competent knowledge of Swahili; others also spoke English or French along with Swahili (ibid: 211). This ethno-linguistic diversity has always been an element of northern Oman, especially along the coastline. Wendell Phillips, an American archaeologist who was on expedition in Oman in the early 1950s, stated that Arabs constituted only half of the population of Muscat, the rest being minorities who were free to practice their religions and customs, including minorities from India, Senegal, and Iran (1986: 13-14). He also mentioned that the Hindi minority controlled the trade, and the Indian currency was the one used in the markets of Muscat (ibid). The Arabic spoken at the time was mostly influenced by Swahili due to the Omani-East African relations, was full of English and Hindi borrowings, and sounds quite like Persian and Gulf Arabic (ibid). Following is some further information on the ethno-linguistic minorities of Oman.

1.4.1 Ethno-linguistic minorities in northern Oman

When talking about the richness and wealth of ethnicities and languages in Oman, it is important to mention that Oman as a region could be geographically as well as politically divided into two main parts, northern Oman, and southern Oman. This is because the northern part of Oman has witnessed a common history that is in different ways distinct from that of Ḍufār, the southern region (Peterson, 2004: 257).

In the northern part, Oman has a considerably big Swahili ~ Zanzibari community. The Swahili-Omanis are descendants of Omanis who moved to East Africa (Zanzibar, Tanzania, Kenya) and Central Africa (Rwanda, eastern Congo, and Burundi). Most of the Arab Omanis who long ago settled in Africa started to return to Oman after the decline of the 'Āl Bū Sa'īd reign in Zanzibar in the mid-1960s. The remaining Swahili-Omanis started to move back to Oman in the second half of the twentieth century, when Sultan Qaboos invited Omanis who live outside Oman to return to contribute to the building of the new country (Valeri, 2009: 19-21). Again, this group spoke Swahili and English fluently, yet those who came from Central Africa spoke French (ibid: 20-21). It is also noteworthy that most of the Omanis who immigrated to Africa in the first place came from tribes that resided mainly in the interior part of northern Oman, namely the Ḥārtī, Ḥabsi, Mazrū'i, Xarūši, Kindi, Rawāḥi, Riyāmi, Maskari, and Maḥrūqi tribes (ibid: 20). The Swahili community nowadays lives in different places in northern Oman but is mostly centred in the Dāxiliyya (interior), Šarqiyya (eastern) regions, and the capital city area, but also scattered in other towns in the Bāṭina Region. Some Swahili-Omanis keep ties with East Africa, and they travel back and forth to Zanzibar because they either have family or personal connections, possess land, or run business back there. Since most members of this community speak fluent English and have

been professionally trained in other Arab and Gulf countries before they returned to Oman, they enjoy economic and social privileges.²⁸

Another important ethno-linguistic minority in Oman is the Baluchi community. Baluchi-Omanis are descendants of the Baluchis who originally came to Oman to join the Omani rulers' armed forces up until recently (Valeri, 2009: 21). Previously, the Baluchi community settled in Muṭraḥ, Muscat, and Qurayyāt towns of Muscat, along the different towns of the Bāṭina coast, and near 'Ibri, a town in the Ḍāhira Region. Nowadays, the Baluchi community mainly lives in the capital area and along the Bāṭina coast. Most Baluchis speak the Baluchi language fluently besides Arabic, yet some have a very limited knowledge of their heritage language. Some Baluchis still hold very close connections with their place of origin; some still visit relatives or possess land (ibid: 22).

In addition, one of the non-Arab communities in Oman is of Indian origin. The Omani Indian relations date back to the 15th century, but Indian and Baluchi merchants have lived in Oman long before. These merchants have contributed to the prosperity of Oman for many centuries (Valeri, 2009: 22). One main sub-group of Indian descendant Omanis is the Lawātya group who originally came from the Sind. They are concentrated in the capital area, mainly in Sūr l-Lawātya in Muṭraḥ, and Muscat towns. The second Indian sub-group is the Bānyān (plural of Bānyāni), which originally means a Hindu trader, and in the Gulf, the Bānyān communities live in different Gulf ports (Holes, 2016: 6).²⁹ Wilkinson mentioned the Hindu Bānyān among other foreign communities of merchants including Persians, Khojas, and Baluch, etc. who were allowed to live and to set up businesses on the coast of Muscat (1987: 69). Earlier on, they also were granted permission to live in Ṣuḥār

²⁸ Some of the Swahili community members belong to the ruling class. PDO, the main oil company in Oman is known to mainly recruit Swahili-Omanis.

²⁹ Also (locally) means a non-Muslim merchant (ibid).

(ibid). These communities were, and still are, free to practice their religions and traditions (ibid). The Bānyān managed to eventually settle in other major ports along the Bāṭina coast, but not in the inland (1987: 69). There are some Bānyān families still living in al-Suwaiq; they are mainly merchants of textiles, and groceries; some own supermarkets and spice mills; they mainly speak Hindi.

The ‘Ajam are a Šī‘i group whose ancestors came from Southern Iran. Nowadays, they mainly live in the Bāṭina coast in Ṣuḥār and Xābūra according to Valeri (2009: 2004), but also in the Afrāḍ and Fāw villages in al-Suwaiq. They speak Arabic as their first language and very few of them may speak Persian as well. In addition, the Baḥrānīs who originally came from different areas along the coast of the Arabian Gulf have been living in Oman for about a hundred years. Like the former group, they are Šī‘i and mainly speak Arabic. Nowadays, in Oman, they live mainly in the capital city area and enjoy economic and political privileges (Valeri, 2009: 24).

The Southern part of Oman is a mosaic of languages with different groups of inhabitants who speak Modern South Arabian Languages (MSALs). For Ḍufār, the native inhabitants are distinguished based on their territory, culture, occupation (i.e., lifestyle), and most importantly, language (Peterson, 2004: 259). Four main MSALs are spoken in southern Oman, namely, Ḥarsūsi, Mehri, Šeḥri (Jabbāli), and Soqāṭri. Simone-Senelle gave estimates of the speakers of MSALs in Oman: around 100,000 speakers of Mehri, 50,000 Soqāṭri, 5,000 Šeḥri, and about 600 or less Ḥarsūsi speakers. Baṭḥari was virtually moribund, and Hobyōt had as few as a hundred speakers living along the borders with Yemen (1990, via Davey, 2013: 28). The Mahra people, as can be seen from these estimates, are the largest Modern South Arabian community in the area. Historically, the Mahra people lived in the Mahra Kingdom which stretched from the mountains of Ḍufār to Ḥaḍramawt in Yemen. Ḍufār’s part of Mahra in the Šiḥr (mountain) area is believed to be originated by the Mahra

bin Heidan tribe, which is part of the larger 'Azdites who immigrated to Oman with Mālik bin Faham (Al-Ṭā'ī, 2008: 7).

1.5 Literature on Omani Arabic

1.5.1 An overview of selected previous works

One of the earliest works on Omani Arabic was conducted by Jayakar (1889) who presented a relatively detailed description of the phonology and morpho-phonology of what he called the dialect of “Ahal-ul Hadar of Oman”, i.e., the dialect of the settled people of Oman, particularly, the Muscat area and the coastal areas near Muscat. He also compiled an English-Omani Arabic glossary of about six hundred words and the first collection of three hundred and twenty Omani proverbs.³⁰ Another earlier study exploring Omani Arabic is Reinhardt's (1894) detailed study of the Banī Xarūṣ dialect, an inland sedentary dialect spoken in Wādi Banī Xarūṣ in the Bāṭina Region;³¹ he based his description on the speech of members of this tribe he interviewed in East Africa. Along with his description, he provided a hundred pages of transcriptions of different texts and two hundred proverbs.³² Also, Rhodokanakis (1908, 1911) carried out an extensive description of Ḍufāri Arabic, a lexicon, and Arabic-German collection of poetry and prose.³³

Other studies were conducted during the second half of the twentieth century, including Galloway's (1977) survey of structural features of Omani Arabic in general, glossary, and selected texts,³⁴ and Brockett's (1985) monograph on Omani Arabic as spoken in

³⁰ (via Brockett, 1985: 1).

³¹ Wādi Banī Xarūṣ is in the southern part of the Bāṭina region. It is about one hour and seven minutes' drive from the centre of al-Suwaiq.

³² (via Brockett, 1985: 1).

³³ (via Davey, 2013: 29).

³⁴ (via Holes, 1989: 446).

the Xābūra town³⁵ which provides a detailed Arabic to English glossary, mainly of agricultural terminology, drawn from thirty hours of tape recordings. The monograph also comments on some important aspects of the grammar of Xābūra Arabic and twenty-six proverbs. Two doctoral theses focused on the speech of ‘coastal’ Omani Arabic. Shaaban’s *Phonology of Omani Arabic* (1977) investigates the changes in coastal Omani Arabic as a result of the social and political changes that the area has witnessed, especially the spread in Modern Standard Arabic through education and media. He focused on major phonetic and phonological processes operating in Omani Arabic and its verbal morphology, tracing a century’s changes in the dialect since Reinhardt’s (1894) study of Omani Arabic. He based his description on four young Omani students at the University of Texas. The other study is by Glover (1988) on the phonology and morphology of Omani Arabic as spoken in the Muscat area, specifically in the old town of Muscat, based on the speech of four women native to the area and their families. She also discussed the effect of exposure to Modern Standard Arabic on the local dialects because of the development in the educational sector in Oman (1988: 17). In addition, other community specific studies include Webster’s (1991) paper on the speech of the Bedouin community of the Wahība Sands, and Eades’s (2009a) analysis of the internal passive in a northern Bedouin variety, his (2009b) description of the transhumant groups in the hinterland of al-Jabal al-’Aḡḡar, and his (2011) paper on the transitional dialect of the interior towns of al-Dārīz and al-Mintirīb.

A more recent detailed study by Davey (2013) documents the coastal Ḍufāri Arabic as spoken by the community living along the area’s coastal plain. His study is the second and only study written in English on Ḍufāri Arabic beside Rhodokanakis (1908, 1911). Davey’s study provides a sketch of the grammar and an overview of the prominent linguistic characteristics of the dialect. He collected data from different sources, including recorded

³⁵ Another coastal town in the Bāṭīna region; it borders al-Suwaiq from the northwest.

free speech, to present a detailed description of the phonology, morpho-syntax, as well as analysing the historical development and synchronic use of function words (Davey, 2013: 17). One main finding is the fact that this dialect shares some features with the northern Omani Arabic dialects, as well as neighbouring dialects of Saudi Arabia and Yemen with which the region shares strong historical connection (ibid: 17). He draws the reader's attention to the alarming pace of change, linguistic and cultural, that this dialect and other communities in the Arabian Peninsula are undergoing, due to the rapid transformation in lifestyle driven by globalisation and modernity (ibid: 21). In addition, Bettega's (2016) study focuses on certain aspects of the syntax of Omani Arabic, namely the agreement patterns, the analytic genitive, the active participle, and the verb. He based his analysis on data collected from the comedy TV show *Yōm b-Yōm* 'day by day' along with naturalistic interview data from native Omani Arabic speakers.

Furthermore, two very recent studies on varieties of Omani Arabic were conducted by Morano (2019) and Ambu Saidi (2019). Morano's study contributes to the dialectologist literature on Omani Arabic. It focuses on the phonology, morphology, syntax, and lexicon of the dialect of Banī Xarūṣ. It also compares Reinhardt's (1894) lexicon and description of the grammar to her findings on the variety as currently spoken in al-ʿAwābi, a district in the inland of the Bāṭina region. In addition to focusing on the comparative aspect of this work, the study also provides a detailed analysis of the lexicon of the dialect, a comprehensive glossary, and documents some of the popular folklore. The second study by Ambu Saidi (2019) is a variationist quantitative investigation of the sociolinguistic variation and change in the dialect of the migrants from the city of Nizwa in the Omani hinterland to Muscat. She looked at five phonological and morphological features of the dialect, namely the labialisation of /i/, vowel syncope in CV.C word-onsets, the use of the traditional yes-no question clitic [-ə], the use of the traditional future prefix [ʔa-], and the affrication of the

2SGF suffix, in the speech of 38 informants stratified by age, sex, age of arrival to Muscat, and length of stay in this area. She found that there is a change in progress in the case of the first four variables, where older speaker group (25-50 years old) are the ones leading this change. Ambu Saidi interpreted her findings in the light of the linguistic awareness of this group with regard to the stereotypical nature of the traditional Nizwa variants in the context of the capital city, and thus the convergence to the latter's norms and expectations of the linguistic market. She also argues that the maintenance of the traditional variants or divergence from the capital's norm in the case when the speakers monitor their speech is on the other hand driven by the speaker's desire to assert of their local identity and group membership.

In addition, Holes's work on Omani Arabic and the Gulf varieties provides significant insight on the linguistic situation in Oman and situates Omani Arabic in relation to the wider peninsular and Arabic context; § 1.5.2 presents an overview of some key contributions of Holes since the 1980s pertinent to the descriptive literature of this variety.

1.5.2 Omani Arabic as a separate dialect group

Omani Arabic is spoken mainly in Oman but can also be found along the east coast of Africa, mainly in Kenya (15,000 L1 speakers in 1995)³⁶ and Tanzania (in Zanzibar and the surrounding area, it is dormant; there are no known L1 speakers).^{37,38} The dialects of Oman share many features with some but not all of the dialects of the neighbouring UAE. In his study of eastern Arabian dialects, Johnstone (1967) classified Omani Arabic as a separate

³⁶ According to Ethnologue. Accessed [December, 2016], at: <https://www.ethnologue.com/language/acx>.
³⁷ (ibid).

³⁸ § 1.5.2.1 and § 1.5.3 provide more detail on what Omani Arabic is in terms of typology, classification, and in relation to the neighbouring areas and dialect history in the Peninsula.

Arabic dialect group. He proposed a dialect map of the Arabian Peninsula which shows four main dialect groups:

1. The north Arabian dialect group which includes Syro-Mesopotamian (the Syrian-Jordanian deserts), *Šammari* and *ʿAnazi* (central Arabia), and east Arabian dialects (the Arabian Gulf).
2. The Hijazi dialect group covering the dialects of the western coast of the peninsula bordering the Red Sea.
3. Omani Arabic which is spoken in the eastern most part of the peninsula. However, Johnstone classified the Buraymi dialect of Oman as an east Arabian dialect along with the UAE one.
4. The *Ḍufāri* dialect of Oman is classified as part of the southwestern Arabian dialect group, along with the dialects of Yemen, *Ḥaḍramawt* and *ʿAden* (ibid: 1).

Two decades after Johnstone's survey, Holes presented a geographical survey of Omani Arabic based on tape-recorded conversations with forty speakers gathered between 1985-1987 in more than thirty locations in Oman (1989: 446-448). In his survey, he provided a classification and a sub-classification of the two main dialect types in Oman, the Bedouin type (B-type), which is mainly found in the deserts west, south, and southeast of the mountains, and the *Ḥaḍari* type (sedentary; S-type), mainly of the mountainous inland. It is noteworthy that he excluded two areas in his survey, namely Muscat, due to the huge influx of non-*Mascati* Omanis and expatriates- Arab and non-Arab- to the area, affecting its social and linguistic structure, as well as *Ḍufār*, due to shortage of data on this area (ibid: 446). Also, it is noticed that the Musandam peninsula is not included in the description.

This survey further highlights the distinctive 'bundle' of phonological and morphological features which Omani dialects have in common in order to situate Omani

Arabic within the wider peninsular dialect topography. According to Holes, some of the following features do exist in other dialects in the neighbouring areas, but they only exist as a ‘bundle’ in Oman where all of them also exist in both dialect types, the B-type and the S-type (ibid: 448-449).³⁹ Here are six of the listed features, but this is not an exhaustive list:

1. The 2SGF possessive/object suffix is a southern [-iʃ] like the one found in Yemen, and the Baḥārna of Bahrain,⁴⁰ not [-iʃ] which is characteristic of the dialects of the Gulf, including Buraymi.⁴¹
2. The active participle infix, –/in(n)/–, which is inserted between an active participle with a verbal force and a following object pronoun, is obligatory in all Omani dialects, e.g., /ka:tb-inn-ah/ ‘write.ACT.PTCP.3SGM-IN-3SGM.ACC’, cf. /katb-ah/ ‘write.PFV.3SGM-3SGM.ACC’ This feature has also been found in the Baḥārna dialect in Bahrain.⁴²
3. The ‘ghawa syndrome’, which is characteristic of northern, eastern and central Arabian dialects, is largely absent in Omani Arabic. Dialects that have this feature re-syllabify an initial CaC syllable to CCa if the syllable ends in one of these consonants, /ħ, x, ɣ, h, ʕ/.
4. A high frequency of feminine plural verb, adjective, and pronoun forms, a feature which Omani Arabic shares with other central and northern peninsular dialects (ibid: 448-449).⁴³

³⁹ A detailed description of the types of features that Dufāri Coastal Arabic shares with the rest of Omani dialects (northern Omani) can be found in Davey’s study (2013).

⁴⁰ Holes also cited other areas where this feature is found, namely Al-Murra in southern Qatar and the northern part of the Empty Quarter (1989: 448).

⁴¹ Holes stated that Buraymi shares some, not all, of the features with Omani Arabic.

⁴² Holes also stated that this infix is also found in the imperfect verb+ object pronoun construction (1989: 448).

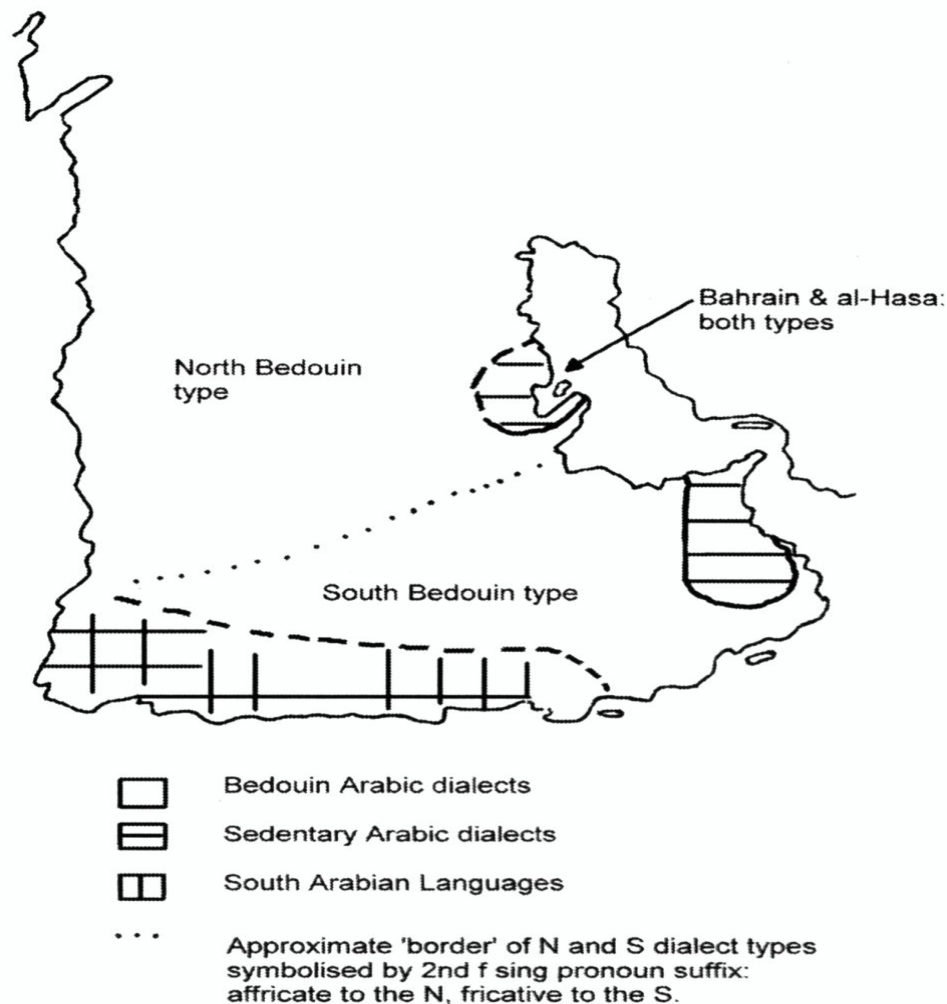
⁴³ This feature separates Omani Arabic, northern, and eastern Arabian dialect groups from those of the Arabian Gulf (eastern Arabian).

5. The heavy occurrence of the internal passive of verb forms I and II, a feature that Omani Arabic shares with other dialects in central Najd, and which distinguishes it from the eastern dialects of the Arabian Gulf littoral.
6. Finally, the interdentalals /θ/, /ð/, and /ðˤ/ are retained (2007: 478).

1.5.2.1 Omani Arabic in the wider peninsular historical linguistic contour: core and periphery

Holes (2006b; 2011; 2012; 2016) argues that the Arabic dialects of the peninsula are very interesting to historical linguists because of their linguistic conservatism and archaism in the sense that they preserve many linguistic features of Classical Arabic (CA) which have become extinct outside the area. He provided an overall outline of the history of the Arabic dialects of Arabia which employs an understanding of the distribution of some of the linguistic features in modern peninsular dialects listed above in the light of peoples' movements and patterns of settlement throughout Arabia to as far back as to pre-Arabic era. Holes arrived at two main conclusions: a) the people along the eastern, south-eastern and southern periphery of the Arabian Peninsula have had common historical and linguistic influences, and this has led to the similarities we see in their dialects today; b) the core in central Arabia is largely different from that of the periphery and that is reflected in the conservative nature and archaism of the dialects spoken there (2006 b: 3-13; 2016: 48; 2012: 242). In addition, according to Holes, the broken 'sedentary' dialect chain formed along the Gulf coast, the coast of Oman, and south-eastern Yemen is ancient and is believed to be 'an older dialect type' in this part of Arabia than the Bedouin one (2012: 240). The Bedouin tribes started migrating to the periphery of the Arabian Peninsula a millennium ago, and over the years, they outnumbered the original populations of the area (ibid: 241). Holes also listed eight S-type features common to the Baḥārna dialect spoken in Bahrain that are also found in

the sedentary dialects of Oman; four of them are also found in the dialects of Ḥaḍramawt and southern Yemen, supporting the above chain link hypothesis (2012: 240). Map 5 sketches the distribution of the S and B dialectal types in the peninsula.



Map 5 Present day linguistic situation in Arabia (Holes, 2006b: 30)

Furthermore, one feature that serves as evidence to the common history of the Arabic speaking communities on the periphery of the Arabian Peninsula (i.e., the Baḥārna in Bahrain, Omanis, and Yemenis) is the aforementioned active participle *-inn-* infix. This infix functions as a “morphological boundary between what is in effect a verb, most often with a perfect tense colouring, with its object pronoun, and never occurs if the object is a full noun” (2011: 81). This infix is argued to be “typologically rare”, “unusual”, “deeply embedded in

the language structure”, and very old, which is the reason why Holes made the assumption that it has a Semitic origin (ibid: 82-86). Although this construction has been attested in Arabic linguistic communities located far from the Arabian Peninsula both in Asia and Africa, it is non-existent in dialects originated in Central Arabia (2012: 240; 2016: 48-49). Holes further explained that the existence of this infix in Arabia originated with the Abd al-Qays and the 'Azd tribes, the indigenous Arab populations of the peninsula's periphery. The Baḥārna of Bahrain claim decent to the Abdul l-Qays and the 'Azd tribes, and since also the sedentary communities in Oman, and south-eastern Yemen are also claimed to be descending from the 'Azd, it is no wonder that their dialects would share linguistic features (ibid: 85-86; 2016: 48). Holes argued that the reason that such construction is also found in linguistic enclaves outside Arabia, like Khurasan, is because 'Abdu l-Qays and the 'Azd were the 'donor tribes' of the Arabic element in these enclaves (ibid: 87-88).

1.5.3 Bedouin vs. Ḥaḍari⁴⁴

The B-type dialects are spoken by some groups who used to follow a nomadic lifestyle in the past but have become urbanised or semi-urbanised, and yet they preserved the “Bedouinness” of their dialect. The S-type is normally characteristic of the long-settled communities of town or village people, who mostly engaged in agriculture. Holes listed three main features or 'markers' that distinguish the Bedouin dialects from the sedentary ones in Oman (Holes, 2012: 237). The first is a classic distinguishing typological marker in the Arabic speaking world, namely Old Arabic /q/, which has a voiceless reflex in the case of the sedentary type dialects, and a voiced reflex in the case of the Bedouin one.⁴⁵ The second feature is the third

⁴⁴ I adapted most of the bracketing and glosses in the sources cited to be in line with this thesis.

⁴⁵ Generally speaking, the voiceless /q/ is found in Oman, parts of Yemen, parts of North Africa, /k/ in the Baḥārna villages of Bahrain and some parts of Palestine, /ʔ/ in Cairo, and most urban centres in the Levant, e.g., Beirut, Damascus, and Jerusalem. The voiced [q] is the most common voiced reflex of *q but there is [ɣ] also which is found in Sudan, and some parts of the eastern region in Oman.

person masculine suffix which is [-ah] in the B-type dialects, and [-oh] in the S-type dialects, as in /kitb-ah/ vs. /kitb-oh/ ‘book.PL-3SGM.GEN’. The third feature that distinguishes the two types is the syllabic structure of the perfect tense of the 3rd person forms. In the B-type dialects, it is a CCvCv(C), as in /rgid-at/ ‘sleep.PFV-3SGF’ and /rgid-aw/ ‘sleep.PFV-3PLM’; however, in the S-type dialects, it is a CvC-Cv(C), as in /raqd-it/ ‘sleep.PFV-3SGF’ and /raqd-o/ ‘sleep.PFV-3PLM’ (ibid: 237).

Furthermore, Holes argues that the B-S distinction as applied in the eastern Arab World do not apply in the same way in Oman, “since B and S as typological labels do not imply a single set of contrasting constituent forms which apply across the whole of the Arab world” (1998: 348). This means that some features, such as the interdental series, are typical of all Omani dialects, whether Bedouin or sedentary in type; in the same way, some features that are regarded as typically Bedouin in other areas, like the feminine plural suffixes of verbal and adjectival forms, are typical of both S-type and B-type Omani dialects. This is not to say that “S” and “B” as labels do not apply in Oman, rather the selection of features that set apart the former from the latter are not the same in the rest of the eastern Arab World, which probably can be understood in the light of the social history of Oman. Holes further maintains that B-type and the S-type dialects are ‘geographically’ rather than ‘socially’ stratified, because of the way these features are distributed across both dialect groups in Oman and the nature of their distribution outside Oman (1989: 448). The S-type is the normal type of dialect, and this is due to its geographical isolation since Oman is an exception to the ‘centre-to-periphery’ migrations of the Bedouin tribes from central Arabia (2012: 241). The socio-political history of the Gulf and Oman explains why Oman, apart from the rest of the states along the Arabian Gulf coast, has retained its distinctive dialect and the supremacy of the S-type over the B-type dialects (ibid).

In addition, as part of his (1989) dialectal survey, Holes provided a sub-classification of these main two divisions of Omani Arabic based on the different phonological combinations of Old Arabic /q/, /k/ and /dʒ/, the syllable structure of Old Arabic CVCVCV(C) forms, and the 2nd person feminine suffix in the different dialects of Omani Arabic. He divided the larger B-type and S-type dialects into the following four sub-types:

1. **Type B1:** in which [g], [k], and [j] are the variants of Old Arabic /q/, /k/, and /dʒ/ consecutively; CCVCV(C) varies with CVC(V)CV(C); *gahwa*-type forms vary with *ghawa*-type forms; /-iʃ/ is the 2SGF suffix. This is the type found with speakers who claimed Bedouin ancestry.
2. **Type B2:** in which the phonological variants of Old Arabic /q/, /k/, and /dʒ/ are [g], [k], and [j]; CCVCV(C) varies with CVC(V)CV(C); there is a *gahwa*-type form only; /-iʃ/ is the 2SGF suffix. This dialect type is found mainly in areas with nomadic Bedouin communities, including recently settled branches of these tribes.
3. **Type H1:** a key S-type dialect which is characteristic of the communities in the interior parts of Oman; here, the Old Arabic /q/, /k/, and /dʒ/ are realised as [q], [k], and [g] ~ [j] ~ [dʒ] consecutively; there are CVC(V)CV(C) and *gahwa*-type forms only, and /-iʃ/ is the variant of the 2SGF suffix.
4. **Type H2:** this type is characteristic of remote, mountainous villages of al-Jabal al-ʿAxḍar villages like Wādi s-Saḥtan and Misfāt l-ʿAbriyyīn.⁴⁶ Here, [k] is the variant of Old Arabic /q/, unconditioned [tʃ] is the variant of Old Arabic /k/, and the affricate [dʒ] is the variant of Old Arabic /dʒ/; there are CVC(V)CV(C) and *kahwa*-type forms only, and /-iʃ/ is the variant of the 2SGF suffix.⁴⁷ (ibid: 449-455)

⁴⁶ These villages are isolated geographically and until recent decades, it was very hard to reach them.

⁴⁷ See Holes (1991) for an account for the fronting of /q/ and the affrication of /k/ in these dialects.

As for the morphological variables, Holes listed four variables whose variants are found to distinguish the B-type from that of the H-type dialects in Omani Arabic; in terms of their distribution, these variants correspond to the above four sub-types within this variety:

- 1) For the 3rd person plural and the 2nd feminine person singular of the imperfective verbs, the B-type dialects have [-u:n], and [-i:n] consecutively, as in /ji-kitb-u:n/ ‘3-write.IPFV-PLM’ and /ti-kitb-i:n/ ‘2-write.IPFV-SGF’ vs. /ji-kitb-u/ and /t-kitb-i/ for the H-type dialects.
- 2) With the 3SGM object/possessive suffix, B-type dialects have [-ah], as in /n-kitb-ah/, whereas the S-type dialects have [-uh], as in /n-kitb-uh/ ‘1PL-write.IPFV-3SGM.ACC’.
- 3) The prefix for the imperfective active verb forms with an initial glottal stop is [ja:-] in the B-type dialects and [jo:] in the s-type, e.g., /ja:-kil/ vs. /jo:-kil/ ‘3-eat.IPFV.SGM’.
- 4) The prefix of the verb forms V and VI is [jti-] in B-type dialects and [jit-] in the S-type ones, e.g., /n-tikallam/ vs. /ni-tkallam/ ‘1PL-talk.IPFV’. (ibid: 449-455).

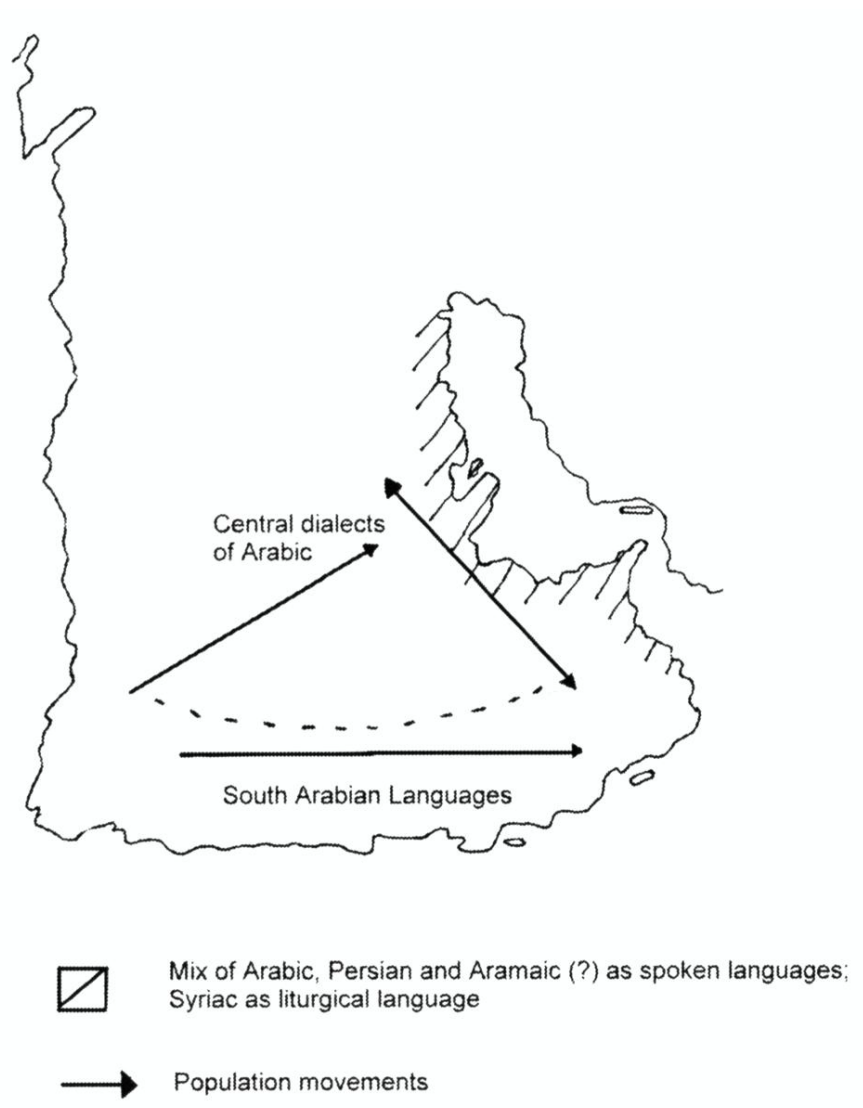
1.5.4 Foreign influence on Omani Arabic

Another descriptive aspect of Omani Arabic is foreign borrowings. Holes (2006 b; 2016) argues that Mesopotamia and South Arabia are the main two sources of ancient linguistic influence on the Arabic dialects in the Gulf.⁴⁸ There is ample evidence of a substrate influence of ancient Semitic, such as Aramaic, Akkadian, and Himyaritic which are indigenous to Arabia, and Greek, which is also another source of foreign influence in the early history of the region (see Map 6).⁴⁹ These varieties were once spoken in the periphery of the peninsula and influenced the lexicon and structure of the *sedentary* Arabic dialects in the area, which is not surprising given the history of the area in which multilingualism has

⁴⁸ Holes’s (2016) introductory chapter presents a detailed account of the sociolinguistic composition of the Gulf and east Arabia.

⁴⁹ Holes (2006b: 9, 16, 19; 2016: 13).

always been a typical state. It is noteworthy that most lexical items fossilised in these dialects are technical words related to the life and culture of the sedentary populations in the Gulf, which is centred on agriculture, seafaring, and fishing; some structural features and functional words that belong to ancient languages also still survive in the Gulf's coastal repertoires (Holes, 2016: 12-32; 2006 b: 14). The Bāṭina Coast, including the study area, is not an exception to the linguistic influence of other non-Arabic tongues on the dialects of the people dwelling there (Phillips, 1967: 62; Glover, 1988: 9-12).

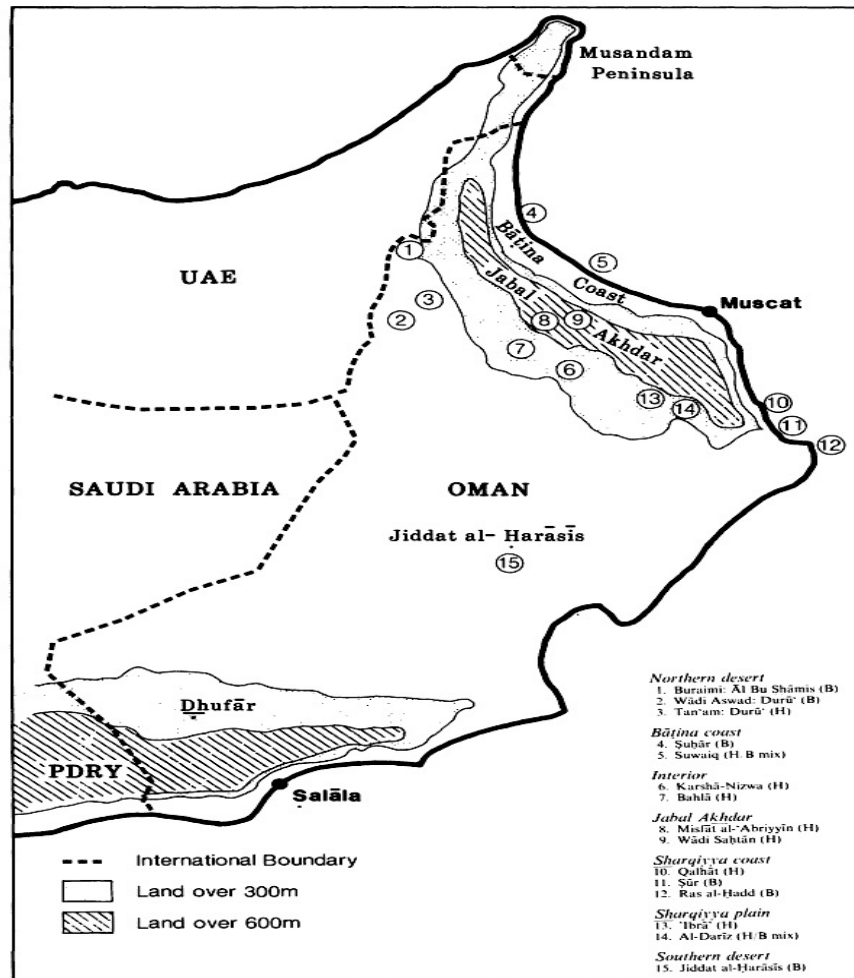


Map 6 Linguistic situation in Arabia at around 1000 A.D. (Holes, 2006 b: 28)

1.5.5 Omani Arabic as spoken in al-Suwaiq

Holes (1989) classified al-Suwaiq as a ‘mixed’ dialect area where both Bedouin and Ḥaḍari dialect types exist. In such ‘transitional’ or ‘border’ areas Bedouin and Ḥaḍari populations live side by side and a high degree of variation occurs where H/B fusion seems to be the norm (ibid: 447). He argues that this is an expected result of long-term contact, social integration, tribal connection, and socio-economical interdependency of the Bedouin and Ḥaḍari populations, facilitated by migration to these coastal towns from the mountain areas and the Gulf (1989: 455).

Holes (1989: 454) argues that the data elicited from the *Suwaiqi* informant is interesting in that his speech does not neatly fit any of the dialect types (explained in §1.5.3 above), because he has Bedouin Type B2 variants of the Old Arabic /q/, /k/, and /j/, but shows Type Ḥ1 *Ḥaḍari* variants of the syllable structure. He also noticed the same for the morphological variables listed in § 1.5.3 above, where in such transitional areas, the distribution of the different variants is not clear cut and characterised by inconsistency and great variation in the choice of variants (1989: 454-5). For the first and fourth variables, the *Suwaiqi* informant used both H and B forms, while for the second variable H-type variants were used and for third one B type variants were used. Map 7 shows Holes’s distribution of the H and B features in Omani Arabic.



Map 7 Dialectal map of Oman; al-Suwaiq is marked as No. 5 in the map

To summarise this chapter, we have seen that the sociolinguistic situation in Oman reflects the long history of the area which extends from the ancient civilisations in the north and the south to the relatively recent settlements of Arab tribes and non-Arab minorities. The chapter also presented some relevant descriptive works on Omani Arabic in general with a focus on the Bedouin-Ḥaḍari dichotomy, and introduced the study area and the speech community under study.

Chapter Two: Notes on the phonology, morphology, and syntax

2.1 Preliminaries

This chapter presents some general remarks on the Bedouin dialect as spoken in al-Suwaiq and al-Miṣin‘a. The first part is a brief description of selected features of the phonology, morphophonology and morphosyntax of the dialect; the second part presents preliminary descriptive analyses of two features, one morphological (the internal passive) and one morphosyntactic (the negation paradigm). The description presented in this chapter is based on interview data collected from native speakers of this variety who come from different villages within the study area, like al-Bārda, al-Manfaš, al-Miladda, and al-Khadhra (l-Xaḍra) locality.

2.2 Phonology

2.2.1 Consonants

Table 2.1 presents the consonantal inventory of the Bedouin variety spoken in the study area. The bracketed sounds are either ‘allophones’ or ‘variants’ of the core phonemes or ‘marginal’ phonemes found in classicised or specialised lexical items. The table is followed by some notes on some of these sounds.

	Bilabial	Labio-dental	Dental	Inter-dental	Alveolar	Post-alveolar	Palatal	Velar	Uvular	Pharyngeal	Laryngeal
	v ⁺	v ⁻	v ⁻ v ⁺	v ⁻ v ⁺	v ⁻ v ⁺	v ⁻ v ⁺	v ⁻	v ⁻ v ⁺	v ⁻ v ⁺	v ⁻ v ⁺	v ⁻ v ⁺
Plosive	b		t d					k (k ⁱ) (k ^h)	g (g ⁱ) (q)		(ʔ)
Emph.			t ^ʕ								
Nasal	m				n						
Fricative		f		θ ð	s z	ʃ			x (χ)	ɣ (ʁ)	ħ ʕ h
Emph				ð ^ʕ	s ^ʕ						
Affricate						(dʒ)					
Tap/trill					r						
Lateral					l (l ^h)						
Glide	w						j				

Table 2.1 Consonant inventory in this dialect.

The uvular stop [q] mainly occurs as a variant of Arabic /q/ in classicised or specialised lexical items. It is noteworthy that this phoneme is a sociolinguistic variable in the study area where [q], which is the same as that of the MSA, is the traditional variant of (q) in the sedentary dialect of the Ḥaḍari population in the study area and [g] is the traditional variant of the Bedouin variety. So, the occurrence of the uvular stop in the Bedouin data may be induced by contact with the sedentary dialect. The items where [q] is used normally have MSA analogues, like /t^ʕari:q/ ‘road’ and /mant^ʕiqa/ ‘region/area’. In addition, the alveolar emphatic stop /d^ʕ/ did not occur in the data. The emphatic inter-dental fricative /ð^ʕ/ is used in

words with etymological /dʕ/, even in the case of classicised lexical items, or in the case of code-switching to the standard variety.

Furthermore, the glottal stop /ʔ/ is also restricted in distribution. In core dialectal items, the glottal stop is usually deleted in different positions in the word, except for some semi-formal or formal words, like /ba-siʔl-iki/ ‘FUT-ask.IPFV.1SG-2SGF’, /l-bi:ʔa/ ‘DEF-environment’, /l-liqa:ʔ-a:t/ ‘DEF-gathering-3PLF’, /l-ʔumu:r/ ‘DEF-matter.PL’, and /qurʔa:n/ ‘Quran’.⁵⁰ While normally, the glottal stop can be simply deleted in word-initial and final positions, word-medially, the deletion of the stop is compensated for by lengthening the vowel of the respective syllable, e.g., /ru:s-ni/ ‘head.PL-1PL.GEN’, /sta:gʲar-ni/ ‘rent.PFV-1PL’, /wara:-hum/ ‘behind-3PLM.GEN’, /ra:j-hum/ ‘opinion-3PLM.GEN’. In addition, the etymological glottal stop can be realised as the glide /j/, e.g., /ʔa:l saʕad/ ~ /ja:l saʕad/ ‘Yāl Saʕad’, and /jinnah/ ~ /innah/ ‘that (complementiser)’.

2.2.1.1 Reflexes of Arabic /k/, /q/, and /dʒ/

In this dialect, the palatal glide occurs as a separate phoneme /j/, e.g., /jo:m/ ‘day’, or the traditional variant of Arabic /dʒ/, e.g., /jidi:d/ ‘new’; the latter is currently in variation with the stop [gʲ] in the study area (see Chapters Four and Five for an elaborate description of Arabic /dʒ/ and a variationist analysis of (dʒ) in this dialect). The voiceless velar stop /k/ is realised as [kʲ] ~ [k̟] ~ [k]. This stop is normally fronted to [kʲ] in the vicinity of front vowels, e.g.,

/a-gu:l-li-kʲ/ ‘1SG-say.IPFV-DAT-2SGF’

/f xidimt-ikʲ/ ‘at service-2SGF; literally: at your service’

⁵⁰ The traditional realisation of Quran in this dialect is ‘gōrān’.

/ħakʕi:-hum/ ‘speech/ dialect-3PLM.GEN’

/mikʕa:n/ ‘place.SGM’

/kʕila:m-hum/ ‘speech/dialect-3PLM.GEN’

In this dialect, the traditional realisation of the 2SGF suffix in the dialect is [-ikʕ] with a fronted velar, whereas the realisation of the 2SGM suffix is [-akʕ] with a back velar. Ingham (1982: 88) describes the masculine suffix, -akʕ in the northern Najdi dialect type as being realised with a very back velar. In addition, [g] and [gʕ] are two allophones of Arabic /q/ in this dialect; the second is a fronted allophone which occurs in vicinity of front vowels, e.g.,

/tila:ħi:gʕ/ ‘traditional youth clothing.PL’

/t-gʕi:m gʕamal/ ‘2-raise.IPFV.SGM camel SGM; i.e., to raise a camel’

/ji-gʕtil-ni/ ‘3-kill.IPFV.SGM-1SG.ACC; i.e., he kills me’

/s-sa:bigʕ/ ‘DEF-previous/ past’

/miɡʕi:l/ ‘staying over (at somebody’s house)’

/rifi:gʕ-ah/ ‘friend-3SGM.GEN’

/ɡʕidda:m-ah/ ‘front-3SGM.GEN’

/swe:gʕ/ ‘al-Suwaiq’

/ɡʕa:bl-i/ ‘watch out.IMP-2SGF’.

So, for the three phonological variables surveyed by Holes (1989), namely, /q/, /k/, and /dʒ/ (§ 1.5.3 in Chapter One), this dialect has [g] ~ [gʕ], [k] ~ [kʕ] ~ [ḳ], [j] ~ [gʕ]. If we focus on the B-type phonological systems in Holes’s (1989) survey, we can see that this variety can be mapped into System 3 (Table 2.2).

	Old Arabic		
	/q/	/k/	/dʒ/
Dialectal system	Omani Arabic		
1	[g] ~ [dʒ]	[k] ~ [ʔ]	[j]
2	[g]	[k] ~ [ʔ]	[j]
3	[g]	[k]	[j]
4	[q]	[k]	[j]

Table 2.2 Combinatorial possibilities for /q/, /dʒ/, and /k/ reflexes in B-type varieties in Oman (Holes, 1989: 451).

As previously mentioned in § 1.5.3, System 3 (and 4) dialects are also spoken by nomadic tribes and their recently settled branches; it is found in the dialect of the Durū‘ tribe in the north-western part of Oman, in the dialect of the Ḥarāsīs tribe in the south, and towards the south of the Wahība Sands, where they are characteristic of the Āl Wahība and Jnaba tribes (ibid: 452). Holes argues that because there is no affrication of the velar stops [g] and [k] in front vowel environments in these dialects, the development of these Old Arabic variables in these Bedouin dialects can be considered ‘more conservative’ (ibid: 452).⁵¹ However, the dialect described here shows another kind of development. While the velar stops in this variety are not affricated in front vowel environment, preliminary analysis of the data shows that these stops are fronted in this environment. Webster (1991) found slight fronting or palatalisation of /k/ in the dialect of the Bedouin tribes in Āl Wahība Sands (System 3, B-2 type). He commented that this process is similar to the velar affrication /k/ > [tʃ] in System 1 and 2 of B1 type varieties, yet he considers it a separate process, and argues that there does not have to be “a historical link or hybrid form intermediate to the B1

⁵¹ Holes (1990) on the affrication of Arabic velar stops.

and B2 types” (ibid: 475); see Table 2.3 which is based on Holes’s (1989) survey listed in § 1.5.3 in Chapter One.

Type	System	Old Arabic			<i>ghawa</i> syndrome	Najdi re- syllabification	2SGF
		/q/	/k/	/q̣/			
		Omani Arabic					
B-1	1	[g] ~ [q̣]	[k] ~ [ṭ]	[j]	<i>ghawa</i> ~ <i>gahwa</i>	CCVCV(C) ~ CVC(V)CV	[-iṭ]
	2	[g]	[k] ~ [ṭ]	[j]			
B-2	3	[g]	[k]	[j]	<i>gahwa</i> -only	CCVCV(C) ~ CVCVCV(C)	[-ij]
	4	[q]	[k]	[j]			

Table 2.3 Bedouin dialectal types in Omani Arabic (Holes, 1989: 449-454).

It is noteworthy that because the slight fronting of /k/ > [kʲ] “may be thought to be a secondary and phonetically natural effect of the preceding /i/” rather than an intermediate stage in the phonologically-conditioned affrication process /k/ > [ṭ] that took place in B-1 type varieties, Webster classified the Wahība dialect as a variant within B-2 type dialects (1991: 476); although the fact that it shows *ghawa* ~ *gahwa* variation, and the palatalised [-ikʲ] instead of [-ij] the 2SGF suffix are ‘suggestive’ of B-1 dialect type (ibid: 475). The dialect under study shows similar features to the former, the only probable difference is that the voiced velar stop [g] in this dialect is also conditionally fronted and thus, this variety could be classified along similar lines; § 2.2.3 elaborates on the two syllabic structural features that distinguish B-1 and B-2 types as they occur in the dialect under study.

2.2.2 Vowels

There are five long vowels, namely /i:, e:, a:, u:, o:/, and three short vowels, namely /i, a, u/, in the phonemic vowel inventory of this dialect, where /i/ and /i:/ are high front closed vowels, /u/ and /u:/ are high back round closed vowels, /a/ and /a:/ are low open vowels, and /e:/ and /o:/ are mid front and mid back round vowels respectively.⁵² Generally speaking, in terms of the distributional patterns of the short vowels, it is worth noting that /i/ < *a commonly occurs in initial open syllables in CV.CVC words where the second vowel is an /i/ or /a/, and CV.CV:C words where the long vowel is predominantly an /i:/, or an /a:/, e.g., /si.kʰan/ ‘place of residence’, /mi.liħ/ ‘salt’, /ħi.li:b/ ‘milk’, and /bi.na:t/ ‘girls’. On the other hand, /i/ < *u is common in initial open syllables in CV.CVC words in which the second vowel is an epenthetic vowel, and in CV.CV:C words where the long vowel is /u:/, e.g., /ma:j ʃ-ʃi.rib/ ‘drinking water’, /sʰi.biħ/ ‘morning’, /t-ti.mu:r/ ‘DEF-dates’, /s-si.bu:ħ/ ‘DEF-shower; i.e., showering’. In addition, an epenthetic short vowel is commonly inserted to break a CC cluster, e.g., in *CVCC words, such as /ta.mir/ ‘dates’, /ʒa.sʰir/ ‘afternoon’, /di.bis/ ‘date molasses’, /sa.min/ ‘ghee’, and /l-ʎi.sil/ ‘DEF-washing’. In terms of quality, short vowels are affected by the consonantal environment, e.g., /a/ is backed in the vicinity of labials and emphatic sounds and is realised as [ɑ] as in [ntʰɑħ] ‘traditional ankle bangles’, [nxaħ] ‘palm tree’; short /i/ is backed, more like [ɨ], probably due to the effect of emphatic, guttural, and labial sounds, in words like /bigar/ ‘cows’, /gabiħ/ ‘before’, /nixla:n/ ‘palm groves’.⁵³

⁵² The discussion here is based on preliminary impressionistic analysis of the vowels in the interview data. The qualities of the /a/ ~ /a:/ phonemes vary between a low front open position to a low back unrounded open position, but most commonly non-front.

⁵³ Emphatic sounds here also include sounds that have become emphatic due emphasis spread.

Final /a/

Final /a/ is also found to be conditionally realised as [i] (or [e]) in the absence of emphatics, as in /mistaʃfi/ ‘hospital’, /sge:fi/ ‘traditional game’,⁵⁴ /ʕiʃi/ ‘dinner’, and /ʃti/ ‘winter’. This process is known as *imāla* in the Arab grammarian tradition, and according to which, it is defined as a process of vowel raising and fronting.⁵⁵ In modern spoken Arabic, raising final -*a* is a common phenomenon. Within Oman, it is characteristic of the Bedouin speech along the Bāṭina coast, as well as the sedentary dialects of villages and towns that are located at the seaside of the mountains like al-Ristāq (Holes, 2007: 481). In the dialect described here, raising final /a/ is a salient feature. It is also noteworthy that, currently, [i] often varies with the mid vowel [e].

*Short unstressed vowel deletion*⁵⁶

Unstressed short vowels are likely to be elided in open non-final syllables. They can be deleted 1) before a long (stressed) syllable, like in /bla:d/ ‘town’, /bju:t.PL/ ‘houses’, and /hba:l/ ‘rope.PL’; 2) before a short (stressed) closed syllable, e.g., /b-sar.ʕe/ ‘quickly’; 3) in a succession of more than two short syllables, e.g., /ʃi-nt.hi/ ‘3-finish.IPFV.3SGM’, and /ki:l.mi/ ‘word’; and 4) in a succession of two short syllables, e.g., /ʃti/ ‘winter’, cf. /ʕi.ʃi/ ‘dinner’ and /ʕi.di/ ‘lunch’.

⁵⁴ Young girls choose five pebbles for the game. At the start of the game, they put four on the floor and they throw the fifth in the air; then, they try to get one of the four before they catch the one they threw, and so on.

⁵⁵ The Old Arabic long and short low back vowels *ā* and *a* in certain environments are realised as the long and short high front vowels [i:] and [i] respectively. Some also argued that *imāla* is a form of ‘vowel harmony’ because the occurrence of an adjacent front vowel and the semi-vowel /j/ triggers it (Sībawayhi, as cited in Levin, 2006: 311).

⁵⁶ Holes (2007: 482) for short unstressed vowel deletion in Omani Arabic.

*Long vowels, diphthongs, and the treatment of etymological *aw and *ay*

Long /i:/ occurs in CCV:C,⁵⁷ CV:C, CV: syllables in different positions, as in /tʰi:n/ ‘flour’, /ni.xi:l ~ nxi:l/ ‘palm groves’, /ji:r.a:n/ ‘neighbours’. Long /u:/ also occurs in different syllabic structures, e.g., in CV:CC as in /jnu:z/ ‘storerooms’, /gru:s/ ‘pancakes’, and CCV: /ʔtʰu:.ra:t/ ‘perfumes’. Long /a:/ also occurs in CCV:C, CV:C, CV: syllables, as in /hja:r/ ‘traditional head jewellery’, /ja:s/ ‘myrtle’⁵⁸ and /sa:.g-o:-h/ ‘carry.PFV-3PLM-3SGM.ACC; literally: they carried it’. In addition, like its short counterpart, the long vowel /a:/ is a low front open vowel; it is generally realised as [a:] but is backed in the vicinity of emphatic sounds and /w/ and realised as a [ɑ:], e.g., /ɣla:mijj-a:t/ ‘traditional earring-PLF’, /wa:jid/ ‘a lot’, /xuwa:bi/ ‘canals’.⁵⁹

Furthermore, word-medially, long /o:/ and /e:/ are the traditional realisations of the etymological diphthongs *aw and *ay in words like /jo:m/ ‘day’, /sʰo:tʰ/ ‘sound/voice’, /be:t/ ‘house’, /θo:b/ ‘dress’ and /ze:t/ ‘oil’. They also occur in other dialectal words such as, /he:jf/ ‘what’, /l-ʕo:d/ ‘DEF-grandfather; literally: the big (one)’, /bo:f/ ‘camels’, /ntʰe:l-a:t/ ‘ankle bangles.DIM-PLF’, and /ro:f/ ‘kindness/compassion’. In addition, long /o:/ occurs as a variant of /u:/ in:

- 1) The 3rd person plural masculine suffix of a perfective verb when a direct object suffix is directly attached to it, as in /ja:f-o:-h/ ‘see.PFV-3PLM-3SGM.ACC; i.e., they saw him’ and /nðʰid-o:-h/ ‘dry.PFV-3PLM-3SGM.ACC; i.e., they dried it’.
- 2) The imperfective weak and strong verb plural masculine subject suffixes, e.g., /ji-bn-o:n/ ‘3-build.IPFV-PLM; i.e., they build’, /ji-mf-o:n/ ‘3-walk.IPFV-PLM; i.e., they

⁵⁷ The consonant cluster here is a product of short unstressed vowel deletion in word-initial open syllables.

⁵⁸ A local flower plant of the *myrtus communis* family, which traditionally has been used by (married) women due to its benefits for their hair and to its nice scent.

⁵⁹ These canals are built in the storerooms where ripe dates are stacked in order to collect the syrup used for making date molasses.

go/walk’, /ji-ʃtayl-o:n/ ‘3-work.IPFV-PLM; i.e., they work’, and /ti-bd-o:n/ ‘2-start.IPFV-PLM; i.e., you start’; cf. /t-ʃarf-u:n/ ‘2-know.IPFV-PLM; i.e., you know’, and /j-sʕaḥ-u:n/ ‘3-pray.IPFV-PLM; i.e., they pray’. It is noteworthy that the suffix [-o:n] is generally less frequent than [-u:n].⁶⁰

On the other hand, long /e:/ also occurs as variants of /i:/ in the 2nd person singular feminine suffix attached to an imperfective IIIy verb stem, e.g., /ti-b-e:n/ ‘2-want.IPFV-SGF; i.e., you want’ and /ti-mʃ-e:n/ ‘2-go.IPFV-SGF; i.e., you go/walk’; cf. /t-kītb-i:n/ ‘2-write.IPFV-SGF; i.e., you write’.

In addition, it is noteworthy that the diphthong /aw/ is used for the 2nd and 3rd masculine plural suffixes, as in /int-aw/ ‘you-2PLM’, /ra:h-aw/ ‘go.PFV-3PLM; i.e., they went’, and /bid-aw/ ‘start.PFV-3PLM; i.e., they started’. It seems that the plural forms with the final diphthong are currently in variation with forms with the monophthong /u/, e.g., /la:za:l-u/ ‘still.IPFV-3PLM; i.e., they still are’, /kin-tu/ ‘be.PFV-2PLM; i.e., you were’, and /ka:n-u/ ~ /ka:n-aw/ ‘be.PFV-3PLM; i.e., they were’. On the other hand, the diphthong /aj/ is also used for the 2nd singular feminine suffix as in /int-aj/ ‘you-2SGF’.

2.2.3 *Ghawa* syndrome and Najdi re-syllabification⁶¹

The *ghawa* syndrome is a Bedouin feature of Arabia; it can be found in the Syrian Desert, in Najd, and in the Bedouin descendant varieties in eastern Arabia. Holes states that the syndrome occurs sporadically and irregularly in Oman, with more consistency in the Bedouin dialects near the border areas towards the north (2007: 481). The *ghawa* syndrome is a by-product of the *gahawa* syndrome and the Najdi re-syllabification rule. In the first stage, a short open vowel /a/ is inserted after a ‘morpheme internal back spirant’ which itself is

⁶⁰ Holes talked about [-o:n] and [-e:n] being the respective realisations of /-u:n/ and /-i:n/ prefix-stem suffixes in final-weak verbs with theme-vowel /a/ (e.g., 2016: 210).

⁶¹ See De Jong (2007: 151-153) for more on this syndrome and Ingham (1982) on Najdi re-syllabification.

preceded by /a/, resulting in forms like /ga.ha.wa/ for **qah.wa* ‘coffee’ and /ba.ħar/ for **bahr* (De Jong, 2007: 151). This process feeds into another re-syllabification rule known as the Najdi re-syllabification by which CaCaCV type words are re-syllabified resulting into CCvCV forms (ibid: 152). This means that lexical items originally affected by the *ghawa* syndrome lose the etymological stem /a/ vowel of the initial syllable in CaCaCV forms, resulting in CCa.CV forms like /gha.wa/ (via *ga.ha.wa* < *gah.wa* < **qah.wa*) ‘coffee’, where the word initial CCa syllable here includes the epenthetic /a/ and the original back spirant (in this case guttural) coda (ibid).

In the dialect under study, forms affected by the *ghawa* syndrome, vary with CaC- forms, e.g., /a-ʃrab gah.wa/ ‘1SG-drink.IPFV coffee’ vs. /gha.wat s^s-s^siβih/ ‘coffee DEF-morning’. It is interesting to note that my informants made a distinction between the two forms, in a way that *ghawa* is semantically different from *gahwa*. One informant said that *gahwa* is the specific actual drink that is made with roasted coffee beans, whereas *ghawa* is more general referring to the whole traditional custom of *fwāla* ‘coffee time’.⁶² Other examples of the variation in this type of forms include /nax.la/ ~ /nχ.ała/ ‘palm tree’, /n-ʃarf/ ‘1PL-know.IPFV’, /n-xadim/ ‘1PL-serve.IPFV’. There is also variation in forms affected by the Najdi re-syllabification rule in this dialect; CvC.CV(C) forms vary with CCa.CV(C) forms, e.g., /ʃajrah/ ~ /ʃjarah/ ‘tree’, /bugrah/ ~ /bgarah/ ‘cow’ (see Holes, 1989: 452-454).

⁶² The semantic difference associated with using different forms of the word for coffee as the drink itself, i.e., /gah.wa/, and coffee as the morning coffee with a snack, i.e., /gha.wat s^s-s^siβih/, is intriguing and can probably be explained in terms of the lexicalisation of /gha.wa/ in certain fixed contexts or environments—maybe suggesting that this is the older, now fossilised form.

2.3 Morphology/ morphophonology

2.3.1 Suffixation

Gender distinction is persevered in the 2nd and 3rd singular and plural suffixes. The dialect of *Yāl Sa‘ad*, like other Bedouin dialects in Najd, Ḥijāz, Mesopotamia and the Arabian Gulf, retains etymological [-i:n] ‘2SGF’ and [-u:n] ~ [-o:n] ‘PLM’ in the imperfective verbs, e.g., /t-gu:l-i:n/ ‘2-say.IPFV-SGF, i.e., you say’, and /j-gu:l-u:n/ ‘3-say.IPFV-PLM; i.e., they say’. The suffixes are also retained when suffixing the direct object, e.g., /t-kitb-i:n-hi/ ‘2-write.IPFV-2SGF-3SGF.ACC; i.e., you write it’, and /i-ji:b-u:n-hi/ ‘3-bring.IPFV-PLM-3SGF.ACC; i.e., they bring it’.

When the third masculine singular direct object pronoun is directly added to a vowel initial suffix, like [-an] ‘PLF’ and [-it] ‘3SGF’, the last consonant of the suffix is doubled, e.g., /j-labs-inn-i/ ‘3-wear.IPFV-PLF-3SGM.ACC; i.e., they wear it’.⁶³ Furthermore, there is an obligatory [-in(n)] infix between the direct object pronoun and the active participle with an agentive force, as explained in § 1.5.2 in Chapter One; e.g., /ħa:fið^s-t^s-inn-ah/ ‘memorise.ACT.PTCP-SGF-IN-3SGM.ACC’, /la:bis-t-inn-ah/ ‘wear.ACT.PTCP-1SGF-IN-3SGM.ACC’, and /fa:jf-in-hum/ ‘see.ACT.PTCP.SGM-IN-3PLM.ACC’.

In addition, the 1st person singular object and possessive suffix is [-je] ~ [ji] ~ [-ja], not [-i], e.g., /ʕumur-ja/ ‘age-1SG.GEN’, /uxu:t-jih/ ‘brother.PL-1SG.GEN’, and /abu:-je/ ‘father.SG-1SG.GEN’. This type of suffix is not very common in modern Arabic dialects in general. It also seems to be in variation with [-i] in the speech of the younger informants.

⁶³ With the PLF, the suffix’s final *-n* is doubled when adding the 1SG and the 1PL direct object suffixes.

2.3.2 Assimilation of /h/ in the 3rd person suffix pronouns

The consonant /h/ in the pronominal suffixes [-ha:], [-hin], and [-hum], namely the 3rd feminine singular and the 3rd feminine and masculine plurals, totally assimilates to the preceding voiceless consonant creating a geminate, e.g., /bo:ʃ-ʃum/ ‘camel.PL-3PLM.GEN’, /ʃif-it-tin/ ‘see.PFV-2SGM-3PLF.ACC’, /n-ʃarif-fi/ ‘1PL-know.IPFV-3SGF.ACC’, and /ta:ri:x-xi/ ‘history-3SGF.GEN’. This type of assimilation is *mutas mutandis* reported for other dialects, namely, Holes (2016), Alaodini (2019) for east Arabian dialects, Al-Hawamdeh (2016), Herin (2010) and Bani-Yasin and Owens (1987) for *Hōrāni* Jordanian Arabic, and De Jong (2000, 2011) for the Sinai Desert. For instance, in the ‘Arab dialect of Bahrain, /h/ in the 3rd person enclitics *-ha* ‘3SGF’ and *-hum* ‘3PLM’ is assimilated to the /t/ in the 3rd feminine stem (perfective) verbs and feminine nouns, like /ʃrub-at-ta/ ‘drink.PFV-3SGF-3SGM.ACC’ and /rgub-at-tum/ ‘neck-SGF-3PLM.GEN’ (Holes, 2016: 77).⁶⁴ While the assimilation in the dialect described by Holes is limited to the phoneme /t/ in specific forms, in the dialect described here the assimilation occurs to all preceding voiceless sounds, and it is not restricted in distribution in terms of the stem the pronominal suffix attaches to. On the other hand, with forms ending with voiced consonants or vowels, this /h/ does not assimilate to the preceding sound, e.g., /maʃ-hi/ ‘with-3SGF.GEN’, /be:n-hum/ ‘between-3PLM.GEN’. This seems to be the case with most function words, which either end in voiced sounds or vowels. It is noteworthy that, in the data, forms with the geminate vary with forms where /h/ is not assimilated (unassimilated forms); so, one can hear both /luʃat-tum/ ~ /luʃat-hum/ ‘language-3PLM.GEN’. Based on preliminary analysis of the *Yāl Sa‘ad* data, younger and educated speakers show more variation, whereas older speakers are generally conservative in their use of the assimilated forms.

⁶⁴ The transcription and the glossing is adapted to be in line with the thesis.

2.3.3 The definite article

The definite article morpheme in the dialect of *Yāl Sa‘ad* is generally realised as *l ~ il ~ li*, like many other Arabic varieties. This *l-* can have an epenthetic vowel before or after depending on the morphophonology of the noun it attaches to, e.g., /li-s‘ya:r/ ‘DEF-children’, /il-liwa:dim/ ‘DEF-people’, and /ruwa:ʕi l-ħa:ra/ ‘people DEF-village’. However, this dialect also has a null definite article, i.e., some noun phrases that can potentially show syntactic definiteness through the use of the overt definite article variably occur without it, e.g.,

1) n-ridd ila Ø-bla:d
 1PL-return.IPFV to town.DEF
 ‘We return to [the] town.’

2) n-ku:n hnī f-li-bla:d
 1PL-be.IPFV here in-DEF-town
 ‘We [would] be here in the town.’

So, forms with the null definite article are in variation with those with an overt article in this dialect. The type of null definite article constructions shown in the examples provided here are more frequent in the speech of older speakers, and especially in the speech of older women in certain villages in the study area. A thorough multivariate examination of this feature is presented in Chapters Six and Seven.

2.4 The apophonic ‘internal’ passive (AP)

The internal or ‘apophonic’ passive is a form of the verb where the passive voice is expressed through ‘ablaut’ which is an internal modification of the stem’s vocalisation patterns (Retsö,

1983: 21).⁶⁵ In other words, the apophonic passive's vowel sequence is a type of an inflectional morpheme. In Arabic, the apophonic passive or the *majhūl* is a special case of apophony. The *majhūl* forms can be used for other functions like qualitative passives, and agentless passives (ibid: 30). The other main way of expressing the passive in Arabic is using 'derivational' passive, also known as the *mut'āwif*, which involves the use of derived verb patterns, namely the *n*- and *t*- derivational verb forms (V, VII, VIII), to express the passive sense of the action described (Ryding, 2005: 657). The derivational passive can convey the passive voice, but it also has other functions (Retsö, 1983: 30; Ryding, 2005: 657). The semantic distinction between the two forms of passive is that the agent of the action in the case of the apophonic passive is 'unknown', i.e., such verbs point to 'actions', while the affixed passive verbs are resultative in that they reflect the 'state of the patient' of the action described, i.e., the 'result' of the action (Holes, 1998: 354). The following are examples of the two passive forms; (b)-(d) are extracted from my data:

Apophonic passive

(a) **kutib**

write.PFV.PASS.3SGM

'it was written.'

(MSA; cf. /**katab**/ 'write.PFV.ACT.3SGM; i.e., he wrote')

(b) **j-samm-an**

3-call.IPFV.PASS-PLF

'they are called.'

⁶⁵ Retsö (1983) presents a survey and a synchronic 'syntactico-semantic' account of the AP construction in Arabic.

(c) **ji-ðbaḥ**

3-slaughter.IPFV.PASS.SGM

‘it is slaughtered.’

(d) **liggit⁶**

gather.PFV.PASS.3SGM

‘it was gathered.’

(Bedouin Omani Arabic; my data)

Derivational passive

(e) **in-ṣaqada**

PASS-hold.PFV.3SGM

‘it was held.’

(MSA; Ryding, 2005: 657)

(f) **in-wilad-it**

PASS-give birth.PFV-1SG

‘I was born.’

(Omani Bedouin Arabic; my data)

In dialects where the apophonic passive still functions as part of the speakers’ grammar and where it is used alongside the derivational one, the difference in the nature of apophony as a ‘conservative’ passive morpheme in such dialects is that they would have lost the ‘narrative’ vs. ‘resultative’ semantic distinction between the two types (Holes, 1998: 355-56).⁶⁶ In addition, other means of expressing the passive include the use of the passive participle, and the use of active verbs with non-specific subjects to avoid mentioning a specific agent (Holes 1998: 351), as in:

⁶⁶ The passive in this case reflects the sense of ‘flexibility’ and ‘potentiality’ (ibid: 357).

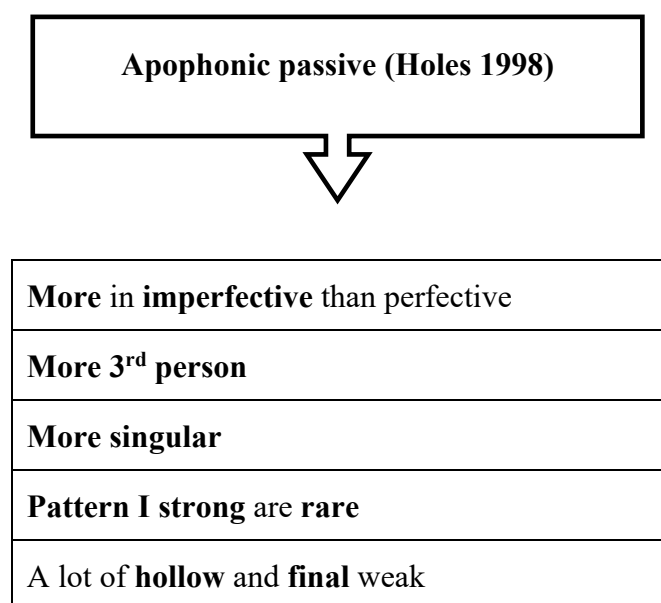
spoken Arabic is an eventual consequence of the changes in the verb vowel system in the dialects, and that the apophonic passive survives as a more or less functional category in certain Arabian dialects. Within the peninsula, the apophonic passive was reported for Northeast Arabian (Ingham, 1982), and Central Arabian (Ingham, 1994) dialects, Yemeni Arabic (Watson, 1993), and Omani Arabic (Holes, 1998; Eades, 2009). On the other hand, the majority of Gulf dialects have lost the internal passive and they compensate for that by a) using the derivational (affixal) passive verbs shown in examples (e) and (f) above, or b) using active verbs with impersonal or unspecified subjects as in examples (i) and (j) above (see Holes, 1998: 357-9).⁶⁷

2.4.1 Apophonic passive in Oman

In the context of Omani Arabic, the apophonic passive is one of the ‘conservative’ linguistic features that are shared by S-type and B-type dialects and which sets Omani Arabic as a separate dialect group in the peninsula (Holes 1998: 348). Holes maintains that “Omani dialects in general have a fully functional internally voweled pattern I passive verb” (2004: 158). He further argues that the apophonic passive is more productive in the sedentary dialects than Bedouin ones, a finding which is not surprising considering the fact that the widespread assumption about sedentary dialects being ‘innovative’ and Bedouin dialects being ‘conservative’ that is applied in the majority of Arabia, does not carry on in southern Arabia (ibid). Based on data collected from three sedentary dialect speakers in three towns in Oman, Holes found 57 apophonic passive tokens of 29 different verbs in 5,000 words of transcription; 53 were in the 3SG imperfective, and 4 perfective tokens only (three 1SG and

⁶⁷ Eades (2009: 18) hypothesises that using active verbs with unspecific agents is the first stage of the disappearance of AP forms and the eventual replacement of internal passive forms with productive affixal forms.

one 3SG).⁶⁸ In addition, all of the tokens were theme I and II. As for theme I, theme I strong verbs were rare (only 5 tokens of five different verbs), and a small number of theme I weak verbs accounted for nearly 50% of theme I forms. The rest were theme II weak and strong verbs (ibid: 349-350). Holes explains the reason for the abundance of the apophonic passive in the data in the light of the methodological biases in data collection; the type of topics discussed in the interviews entail that the speakers would give descriptions and explanations for processes in which the agent is irrelevant (ibid: 350). Nevertheless, the findings suggest that the apophonic passive is still functioning as part of the speakers' grammar and is similar to that of the Bedouin dialects in the interior of Oman (ibid: 359).



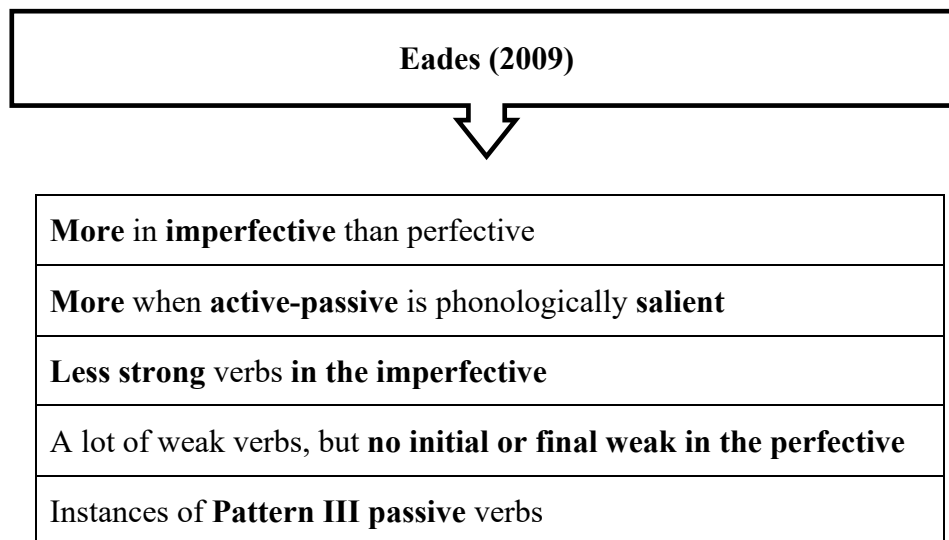
On the other hand, data collected from older Bedouin speakers along the coastal line and from Bedouin-sedentary 'transitional' towns in Oman shows transitivity with more

⁶⁸ In Holes's data, the most occurring perfective apophonic passive verbs were in the 1SG, like /xliq-t/ 'create.PFV.PASS-1SG; i.e., I was born; literally: I was created' or /wliid-t/ 'give birth.PFV.PASS-1SG, i.e., I was born'. These forms again are being levelled out in favour of active verbs with unspecified subjects, like /wild-o:-ni/ 'give birth.PFV.ACT-3PLM-1SG.ACC; i.e., they gave birth to me' (personal communication; transcription is mine); more on these forms in (Holes, 1998: 356, 358-9).

preference for the use of the affixed passive and active verbs with non-specific subjects (ibid: 359-60). In these areas the apophonic passive is ‘recessive’; it is largely replaced by ‘strings of active verbs using a non-specific ‘you’, ‘we’, ‘they’, ‘a man’, etc.’ (ibid: 351). For the few tokens of the apophonic passive in the Bedouin data (only 8 tokens in more than 5,000 transcribed words), Holes argues that these are ‘formulaic’, ‘ clichéd’, and ‘proverbial’ phrases, which are not to be considered as part of a fully productive AP system (ibid: 352).

In addition, in approximately 3,300-word-long text from two elderly men from the *Hidyīwī* Bedouin tribe residing in the hinterland of northern Oman, Eades (2009) found 56 apophonic passive tokens. To summarise, in his data, the imperfective apophonic passive verbs were more present in the data than the perfective ones; the apophonic passive verbs show singular and plural agreement; more occurrences of the apophonic passive are found in verbs where the passive-active phonological distinction is most salient, i.e., when the vowel quality is different in the two forms (ibid: 9); the apophonic passive was also found in Pattern I strong verbs. Eades also reported the use of strings of active verbs with impersonal subjects as a functionally equivalent alternative to passive verbs particularly with strong verb stems (ibid: 10). Eades maintains that the sedentary-Bedouin distinction in the interior of Oman is marked by a functioning apophonic morphology in strong verbs marked by the height of v2 unstressed vowels both in the perfective and imperfective. In the perfective, the passive-active distinction is either marked by the height of v2 or by altering both v1 and v2 in the case of Pattern II verbs (ibid: 11); although he found Pattern III strong and hollow perfective passives, there were no instances of perfective passives with an initial or a final weak radical (ibid: 12). In the case of the imperfective passives, Eades found fewer instance of strong imperfective passives, the majority being weak verbs (ibid: 12-15). In the case when the stem vowel is /a/, both the passive and the active have the same form; the context is needed to clarify the status of the verb (ibid: 14). Eades maintains that passivisation by affixation is

another productive alternate for the apophonic passive; it is increasing massively in coastal and interior dialects (ibid: 16). In the *Hidyīwī* dialect, Pattern VII forms have developed to convey or emphasise different functions, namely a) the ‘state of the patient’ where ‘the involvement of the agent is not necessarily implied’, as in /ba-j-int^ʕabax/ ‘FUT-3-cook.IPFV.SGM; i.e., will cook’ (ibid: 15-16)⁶⁹, b) the state of the patient where the involvement of the agent is not necessarily ruled out, as in /infataħ/ ‘open.PFV.3SGM; i.e., opened ~ was opened’ in, c) a potential passive, i.e., implying an involvement of an agent, in which case, Pattern VII passive verbs have also generalised to transitive verb stems, e.g., *s-k-r* > /ma: t-in-sikkir/ ‘NEG 3-PASS.close.IPFV.SGF; i.e., cannot be closed’ (ibid: 16-17).⁷⁰ It is noteworthy that Eades has stated that although Pattern VII is productive and has expanded its semantic meaning, it does not ‘productively form passive verbs’, so AP is still the main passivisation method in the *Hidyīwī* dialect (ibid: 17-18).



⁶⁹ Refer to Eades (2009: 15-17) for the complete version of the examples cited here; the transcription here is adapted to be in line with the thesis.

⁷⁰ By contrast, the AP is specific to marking the passive function only and does not convey potentiality.

2.4.2 Apophonic passive in the study area

Apophonic passive in this variety occurred more frequently in the speech of my older speakers. Generally speaking, forms like /j-samma/ ‘3-call.IPFV.PASS.SGM’, /j-ba:ʕ/ ‘3-sell.IPFV.PASS.SGM’, /ja-ʕtʕa/ ‘3-give.IPFV.PASS.SGM’ are common. The following is an excerpt from an interview with a 60+ year old woman from al-Miladda village in al-Miṣin‘a, narrating her experience with the Malaria vaccine when she was an adolescent. She uses five 1PL perfective passive Patterns I and II verb forms (strong and geminated) out of seven verbs in total.

w **niddirni**, w **riggijʕni**- tra humma alhi:n ma: be:ʕillu:nni, xa:jfi:n
 min l-maraðʕ- w **jille:ni** w **ðʕiribni** w **bifitʕni** l-waħda baʕtʕa, w
 ragigʕiʕu:nni

‘and we were taken out [of the hospital] but were brought back-
 because if they [the family] do not take us [there] they are concerned
 about the disease [Malaria]-so, we were taken, we got vaccinated, we
 each got a mark [on the arm], and then we were brought back [home].’

Furthermore, preliminary analysis of a one-hour interview with four Bedouin women from al-Bārda village in al-Suwaiq suggests that the apophonic passive for these speakers is still widely productive, and very robust at least in these speakers’ grammars not limited to formulaic use.⁷¹⁷² In fact, there is a relative abundance of the apophonic passive forms in this interview; there were **95** apophonic passive tokens in total for 34 different Pattern I and II verbs, expressing a number of morphosyntactic features.

⁷¹ This analysis is presented at the North Atlantic Conference of Afroasiatic Linguistics (NACAL 47) held at Université Sorbonne- INALCO, Paris on the 24th-26th of June 2019.

⁷² These women have lived locally throughout their lives; they are uneducated and aged 50 and above years old. Apophonic passive tokens were extracted in an Excel sheet and coded for different features, e.g., theme vowel, root, transitivity, number, gender, person, tense; the verbal roots are coded according to Holes’s (2001) glossary of eastern Arabia.

Verb type	Perfective (N and Example)	Imperfective (N and Example)	Total for verb type
strong	2 /liggitʰ/ ‘gather.PFV.PASS.3SGM’	28 /n-xibbir/ ‘ 1PL-tell.IPFV.PASS’	30
weak	7 /yitʰtʰ-at/ ‘cover.PFV.PASS-3SGF’	30 /tʰ-xa:tʰ/ ‘3-tailor.IPFV.PASS.SGF’	37
doubly weak	3 /siwwi:-l-hi/ ‘made.PFV.PASS-DAT- 3SGF’	15 /j-sawwa/ ‘3-make.IPFV.PASS.SGM’	18
geminated	1 /ʕigg-at/ ‘throw.PFV.PASS-3SGF’	9 /j-dagg/ ‘3- beat.IPFV.PASS.SGM’	10
Total for tense	13	82	N=95

Table 2.4 Apophonic passive tokens distributed across different verb types.

Morphological category	Perfective	Imperfective	Totals for morph. category
1PL	none	3	3
3SGM	7	47	54
3SGF	6	29	35
3PLM	none	1	1
3PLF	none	2	2
Pattern I	6 (4 different verbs)	34 (17 different verbs)	40
Pattern II	7 (5 different verbs)	48 (16 different verbs)	55

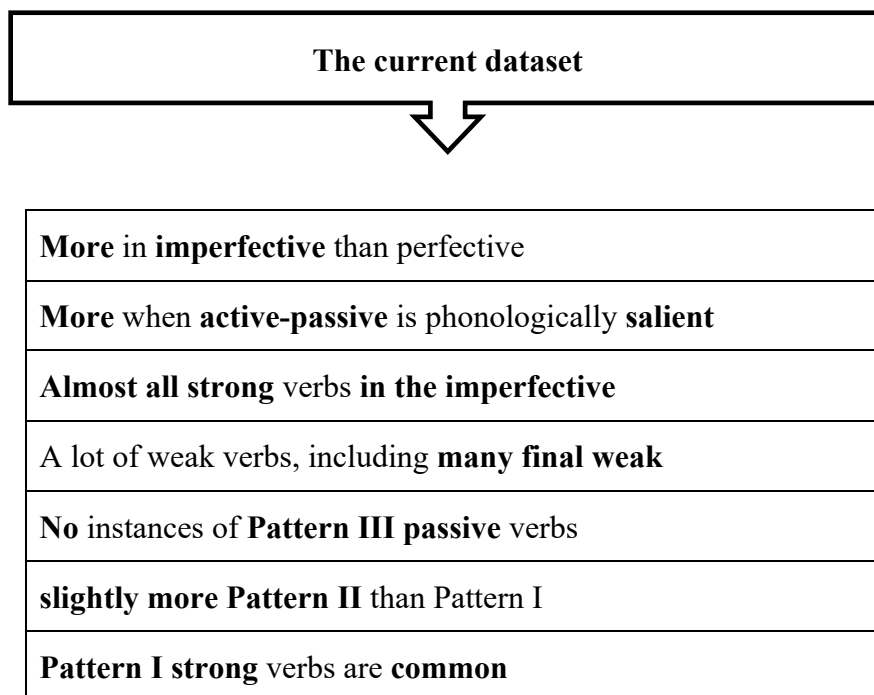
Table 2.5 Apophonic passive tokens distributed across morphological categories.

Table 2.4, and Table 2.5 generally show that:

- 1) Although the proportion of weak verbs in the dataset is high (68%), there is a relative abundance of strong passive verbs in the dataset (31.6%).

- 2) There are more apophonic passive verbs in the imperfective than in the perfective tense.
- 3) The overwhelming majority of the apophonic passive forms are in the 3SG form (93.4%).
- 4) There are more Pattern II than Pattern I verbs, although the proportion of the latter is relatively high (42%).
- 5) Many strong verbs Pattern I are verbs where the passive-active is not phonetically distinct, like /t-nað^sð^saf/ ‘3-clean.IPFV.PASS.SGF’ ~ /t-nað^sð^saf/ ‘3-clean.IPFV.ACT.SGF’, *cf.* /ji-ntaf/ ‘3-pick.IPFV.PASS.SGM’ ~ /ji-ntif/ ‘3-pick.IPFV.ACT.SGM’.
- 6) Pattern I strong verbs are not as frequent as Pattern II strong verbs, but they are not uncommon as found by Holes (1998: 394-50) since they constitute 26.7% of apophonic passive strong verbs (N=8/30). Some verbs in this category have different active-passive vowel sequences, yet others do not.
- 7) The 1SG, 1PL, 3PL barely occurred in the dataset (total N=6). This is probably to do with the nature of language in general since we do not normally talk about ourselves in the passive (Ingham, 1994: 26); *cf.* the excerpt at the beginning of this section.⁷³

⁷³ Instances of non-CLA passives of a combined Pattern VII-II form, like /inʕawwar/ ‘he hurt himself/ he got hurt’, as found in Holes’s data on Oman and Bahrain (personal communication) and Ingham’s data on Najd (1994: 74), did not occur in the data collected for this particular analysis, and although I have not analysed the internal passive in the speech of my larger sample, I could tentatively say that this is not a feature of the dialect under study; there are forms like /atʕawwar/ ‘I get hurt’, /ribe:t/ ‘I grew up’, /wild-o:n-jəh/ ‘give birth.PFV.ACT-3PLM-1SG.ACC; i.e., I was born, literally: they brought me (to life)’, and /in-wilad-it/ ‘PASS-give birth.PFV-1SG; i.e., I was born’. One informant confirmed the use of *n-* or *t-* affixal forms for the above mentioned Pattern VII-II verb types, e.g., /ma: ji-n-git^siʕ/ ‘NEG 3-PASS-cut.IPFV.SGM, i.e., it cannot be cut’, and /ma ji-t-waggaf/ ‘NEG 3-PASS-stop.IPFV.SGM, i.e., it cannot be stopped’.



The preliminary analysis suggests that the apophonic passive is still functioning as part of the dialect's grammar at least with some speakers, just like the sedentary and the Bedouin of the interior. It is noteworthy that few tokens of derivational passive verbs were found in the interview. The speakers either used strings of active verbs with non-specific subjects, or strings of apophonic passive verbs in their descriptions, which is not surprising since other features of this dialect are quite archaic and conservative, like the null definite article introduced in § 2.3.2 above. Although situated on the coast and in a Bedouin-Sedentary 'transitional' area, it can be argued that the dialect and the speech community have preserved to a long extent their Bedouin linguistic identity.

2.5 Negation in the dialect

This section presents a preliminary analysis of the negation paradigm in this Bedouin variety. It is descriptive with the aim of providing further insights on the traditional negative particles

am, ab, mā ~ ma, lā ~ la ~ ʕan lā,⁷⁴ which constitute the main means for negation in this variety. In addition to interview data, this description makes use of direct elicitation, and a written quiz on acceptability judgment.⁷⁵ The descriptions and the examples are mostly based on a discussion with a young woman from al-Khadhra locality, who also did the acceptability quiz used here.

2.5.1 Negating with *mā*

The particle *mā* is the typical verbal negator in Omani Arabic (Holes 2007: 485). In the Bedouin type varieties, like those of the Bāṭina region, which typically have *mu* and *muhu* for non-verbal constructions, *mā* functions as a verbal negator and a negator of prepositional expressions of possession (ibid). In this dialect, it negates verbs and helping verbs e.g.,

a. Negating the present tense:

1) *ma:* *t-sidd*
 NEG 3-suffice.IPFV.SGF
 ‘It is not enough.’

2) *ma:* *t-haḏīr* *ʃarwa* *ḏe:la:kʰ*
 NEG 3-talk.IPFV.SGF like DEM
 ‘She does not talk like those [people].’

⁷⁴ The vowel quality of the negative particle /ma:/ differs across the study area. It is a short [a] in the villages closer to al-Miṣin‘a, but a long [a:] in the centre of al-Suwaiq.

⁷⁵ The quiz is designed based on the discussion of negation and the examples provided in Al-Balushi (2016).

b. Negating the past tense:

- 3) ma: k̄tib-ah
 NEG write.PFV.3SGM-3SGM.ACC
 ‘He has not written/did not write it.’

- 4) ma: k̄tab-il-hi faj
 NEG write.PFV.3SGM-DAT-3SGF thing
 ‘he didn’t write her anything [in his will].’

c. Negating the future tense when a future particle *b-* is used, e.g.,

- 5) ma: b-a-kalm-a hu:
 NEG FUT-1SG-speak.IPFV-3SGM.ACC 3SGM.ACC
 ‘I will not talk to him.’

- 6) ma: b-a-ru:h
 NEG FUT-1SG-go.IPFV
 ‘I will not go.’

d. Negating coordinated clauses, e.g.,

- 7) ma: k̄allam-t aḥmad w la: k̄allam-t xa:lid
 NEG talk.PFV-1SG Ahmad CON NEG talk.PFV-1SG Khalid
 ‘I have neither talked to Ahmad nor to Khalid.’

2.5.2 Negating with *ab*

The particle *ab* is used with many ‘non-verbal’ forms of speech and constituents to mean ‘not’ or ‘is not’, e.g.,

- 8) xa:lid ab mudarris
 xalid NEG teacher
 ‘Khalid is not a teacher.’

This particle is probably related to the non-verbal Bedouin negative particles in eastern Arabia, including the more common *hub* which is mentioned for Najdi descendent dialects in Bahrain and some Bedouin varieties in Oman (see Holes, 2016: 104-107, and Holes, 2007: 485 for more on the Bedouin-type non-verbal negators in Bahrain and Oman).⁷⁶ In this dialect, *ab* can as well be viewed as a free morpheme, not part of a pronoun+NEG sequence; one probable evidence is that *ab* can be preceded by a free pronominal/personal pronoun:

- 9) hu: ab faqi:r
 3SGM.NOM NEG poor
 ‘He is not poor.’⁷⁷

Along with *ab*, the Bāṭina Bedouin variety also has *mu* ~ *ma-hu* ~ *-hu* which is arguably a pronominal that has come to serve as a copula (Al-Balushi, 2016: 113).⁷⁸ In addition, Al-Balushi (ibid: 112-113) describes ‘*-ab*’ as a negative marker used by the speakers of eastern Bedouin varieties in Oman, along with *m-ab*, *ma-hu*, *ma-hu-b*. He argues that ‘*-ab*’ is a

⁷⁶ It is noteworthy that, in Bahrian, the ‘Arab variants are distinctive from the Baḥārna ones in that they all include a *-b* element, and that the ‘Arab Bahraini variety also has feminine counterparts, *hīb* and *mā hīb*, although the masculine forms are more common and used for both the feminine and the masculine genders (Holes, 2016: 106). Based on the data I have, *ab* in this dialect can be used for both genders, and there is no feminine counterpart.

⁷⁷ Holes mentioned the possibility that all of the former variants along with *ab* and *am* (see § 2.5.3) are ultimately derived from the Bedouin *mā-hū-b* (personal communication).

⁷⁸ One of the informants used *mu* in /mu t-gu:l-li/ ‘NEG 2-say.IPFV.SGM-DAT-1SG, i.e., don’t come and tell me’, but this is very uncommon in the data I collected.

negative copula which probably originates from a combination of the negative particle 'a- and a copular element -b, and it appears in expressions like /ʕa-b ki:h/ 'not like this' and /ʕa-b ka:k/ 'not like that'.

In this dialect, the negative particle *ab* is used to negate non-verbal predicates and complements, such as nominals, including nouns (common and proper), and adjectives; it also negates locative adverbs, prepositional phrases, possessive and genitive constructions, pronouns, demonstratives, equational sentences. Following are examples of the uses of *ab* in the dialects under study:

a- Negating personal pronouns, e.g.,

10) /ab ani/ 'not I/me'

11) /ab humma/ 'not them (M)'

12) /ab hu:/ 'not him'

13) /ab hi:/ 'not her'

14) /ab hinne/ 'not them (F)'.

b- Negating the quantifier **kull* and quantifier phrases, e.g.,

15) ab kill-hin
 NEG QUANT-3PLF
 'not all of them.'

c- Negating noun phrases:

1- Proper nouns, e.g.,

16) ab yajda:ʔ
 NEG Ghaida
 'not Ghaida.'

2- Toponyms, e.g.,

17) ab l-xaḏ^srah
 NEG DEF-Khadhra
 ‘not al-Khadhra.’

3- Common nouns, e.g.,

18) he: ab kirsi
 DEM.SGM NEG chair.SGM
 ‘This is not a chair.’

19) ha:j-i ab ʃjar.ah
 DEM-SGF NEG tree.SGF
 ‘This not a tree.’

20) he: ab wilad
 DEM.SGM NEG boy.SGM
 ‘This is not a boy.’

21) ab l-filfil
 NEG DEF-pepper
 ‘Not the pepper,’

d- Negating possessive phrases, e.g.,

22) ha:j-i ab sajja:rt-a, ha:j-i sajja:rat xa:lid
 DEM-SGF NEG car.SGF-3SGM.GEN DEM-SGF car.SGF Khalid
 ‘This is not his car; this is Khalid’s car!’

e- Negating the analytic construct, e.g.,

23) ab ħa:l l-wilad, ħa:l uxu:-h
 NEG for DEF-boy for brother-3SGM.GEN
 ‘It is not for the boy; it is for his brother.’

f- Negating demonstrative phrases and demonstratives (far and near), e.g.,

24) ab ðe:-la
 NEG DEM-PL
 ‘not these ones.’

25) ab hað-i:k^ʃ
 NEG DEM-SGF
 ‘not that one.’

26) ab ha:j-i sa:ʕat-jeh
 NEG DEM-SGF watch.SGF-1SG.GEN
 ‘My wristwatch is not this one! lit. not this one my wristwatch.’ *cf.*

27) ha:ð-i ab sa:fat-jeh
 DEM-SGF NEG watch.SGF-1SG.GEN
 ‘This is not my watch.’

g- Negating prepositional phrases, e.g.,

28) ab fi-l-be:t
 NEG in-DEF-house
 ‘not in the house.’

29) ab fi-l-gja:mʕe, yajda: hni:h
 NEG in-DEF-university.SGF Ghaida here
 ‘[...] not in the university, Ghaida is here [at home].’⁷⁹

30) he: l-qitʕa:r ab le:n ko:lʕester bass, ...
 DEM.SGM DEF-train.SGM NEG towards Colchester only
 ‘This train does not stop at Colchester ...; lit. this train not to Colchester
 only.’

31) ab min l-be:t ja:jb-e li-ghawah
 NEG from DEF-house bring.ACT.PTCP-SGF DEF-coffee.SGF
 ‘I have not brought the coffee from home; lit. not from the house I
 brought the coffee.’

⁷⁹ To negate a previous statement about Ghaida’s whereabouts.

32) hu: fi-l-mat^ʕbax, ab fi-l-k'ibat, bas fi-d-dirig^j
 3SGM.NOM in-DEF-kitchen NEG in-DEF-cupboard CONJ in-DEF-drawer
 '[...] It is in the kitchen; not in the cupboard, but in the drawer.'

h- Negating the semi-preposition *šarwa* < OA **mitl*, e.g.,

33) ab šarwa l-ħi:n
 NEG like DEF-now
 'It is not like nowadays.'

i- Negating adverbs, e.g.,

34) ab hni:h
 NEG here
 'It is not here.'

35) ab fo:g s^ʕ-s^ʕat^ʕiħ, xalle:-na:-h fi-l-ħadi:qa
 NEG on DEF-roof put.PFV.ACT-1PL-3SGM.ACC in-DEF-garden.SGF
 '[... It is] not on the roof, we put it in the garden.'⁸⁰

j- Negating colour adjectives: e.g.,

36) ab as^ʕfar
 NEG yellow
 '[It is] not yellow.'

⁸⁰ An answer to the question 'Where did you put the chair?' where the expected place would be the negated NP, in this case 'the roof'.

k- Negating the existential *šay* ‘thing’ when followed by an adjective in a predicate clause, e.g.,

37) he: ab faj tʿabi:ʕi
 DEM.SGM NEG thing normal
 ‘This is not normal; lit. not thing normal.’

l- Negating *kīh*<OA **hākadā*, e.g.,

38) ab kī:h
 NEG DEM
 ‘not like this.’

m- Negating adjectives of quantity, e.g.,

39) ab wa:jid
 NEG much
 ‘not a lot/much.’

n- Negating interrogatives (yes/no question) and tag questions, e.g.,

40) ab hu: ðe: r-rajja:l illi ḥasʕsʕaʕ-na:-h ða:k l-jo:m
 NEG 3SGM.NOM DEM.SGM DEF-man.SGM REL find.PFV.ACT-1PL-3SGM.ACC
 DEM.SGM DEF-day
 ‘Isn’t this the man whom we found the other day?’

41) he: nafs l-fusta:n illi la:g-i:n-n-i ði:kʰ l-jo:m, ab hu:?
 DEM.SGM same DEF-dress.SGM REL find.ACT.PTCP-PLM-IN-3SGM.ACC
 DEM.SGF DEF-day NEG 3SGM.NOM
 ‘This is the same dress which we found the other day, isn’t it?’

- 42) aḥmad ab ʕind-a sajja:rat xa:lid
 Ahmad NEG at-3SGM.GEN car.SGF Khalid
 ‘Does not Ahmad have Khalid’s car?’

o- Negating stative predicative adjectives, e.g.,

- 43) ha:j-i ab faqi:r-ah
 DEM-SGF NEG poor-SGF
 ‘This one is not poor.’

p- Negating adjectives of quality, e.g.,

- 44) ha:j-i ab ḥilw-ah/ya:wj-eh
 DEM-SGF NEG pretty-SGF/nice-SGF
 ‘This one is not pretty/nice; lit. this not pretty.’

q- Negating expressions of possession; this entails a certain order of *ab* in relation to the possessive ‘*ind*, e.g.,

- 45) sajja:rat xa:lid ab ʕind aḥmad
 car.SGF Khalid NEG at Ahmad
 ‘Khalid’s car is not with Ahmad ~ Ahmad does not have Khalid’s car.’

Finally, it is noteworthy that the *ab* particle is currently in variation with *mā* in all of the above constructions, e.g.,

- 46) ɔe: l-be:t ma: wa:ʒid ʕo:d ab inn-i sʕyi:r bass ma: min kubr-ah
 DEM.SGM DEF-house NEG very big NEG COMP-3SGM small CONJ NEG that size-3SGM
 ‘This house is not too big; not that it [is] small, but it [is] not that big’.

2.5.3 Negating with *am*

In addition, Al-Balushi (2016: 113) states that a negative particle *'am* exists in the *Maj'ali* branch of the *Jnaba* tribe residing in the mainly sedentary interior town of Manaḥ but does not further elaborate on the nature of its distribution. In this dialect, *am* is used with the participles to denote the sense of ‘will not.../not ...ing/never ever will’ and to imply a progressive aspect, e.g.,

47) ani	am	ra:jħ-i	l-be:t
1SG	NEG	go.ACT.PTCP-SGF	DEF-house
‘I am not going to go to the house; I will not go to the house.’			

More examples of *am* include other attitudes to the action like assertiveness, certainty, intensity, impossibility, challenge/daring, disdain, sarcasm and give extra shades of meaning to the context like ‘as simple as that!’, ‘no way this action will happen’, ‘will make sure this action doesn’t happen’, ‘don’t think this will happen’, ‘will not happen anymore’. Table 2.6 and Table 2.7 show different *am*+participle agreement patterns.⁸¹

Active participle negation with /ra:jħ/ ‘go.ACT.PTCP’	
SGM	am ra:jħ
SGF	am ra:jħih
PLM ⁸²	am ra:jħi:n
PLF	am ra:jħa:t

Table 2.6 Negation of the active participle with *am*.

⁸¹ The active and passive participles in Arabic are nominal forms; nominal forms are inflected for gender and number, but not for person.

⁸² The dual-plural distinction is lost in this dialect, just like many other Arabic dialectal varieties.

Passive participle negation with /masq'u:n/ 'imprison.PASS.PTCP'	
SGM	am masq'u:n
SGF	am masq'u:nih
PLM	am masq'u:ni:n
PLF	am masq'u:na:t

Table 2.7 Negation of the passive participle with *am*.

Following are some remarks on negating with *am*:

First, we have seen that the particle *am* negates the participles. The negated participles can come in different number, and gender forms. The active participles have verbal force. They are derived from transitive verbs, e.g., /k'a:tib/ 'write.ACT.PTCP.SGM', /ba:jiʕ/ 'sell.ACT.PTCP.SGM', and intransitive active verbs, e.g., /ra:jiħ/ 'go.ACT.PTCP.SGM', /mizzawwig/ 'marry.ACT.PTCP.SGM'. Active participles of the former type can have the direct object pronoun suffixed and the Omani Arabic infix [-in(n)-] separating the two arguments probably to mark this transitivity relationship; *am*+active participle strings probably also emphasise the unwillingness of the speaker to undertake the action, or the speaker's certainty that the addressee is not going to undertake the action; *cf.* the use of the negative particles in examples 48-50.

48) am ka:tb-inn-i
 NEG write.ACT.PTCP.SGM-IN-3SGM.ACC

Literally: 'I am not the writer of it' or in the sense of 'I will make sure he will not write it/there is no way he is going to write it.'

49) ma: b-a-kītb-i
 NEG FUT-1SG-write.IPFV-3SGM.ACC
 ‘I will not write it.’

50) ma: ja:j a-kītb-i
 NEG PROG 1SG-write.IPFV-3SGM.ACC
 ‘I am not going to write it.’

More examples of the negated passive participle forms derived from transitive verbs used in the sense of ‘not going to’ include e.g., /am mafru:b/ ‘NEG drink.PASS.PTCP.SGM; i.e., not going to be drunk’, /am mabna:j/ ‘NEG build.PASS.PTCP.SGM; i.e., not going to be built’, /am maʕtʰa:j/ ‘NEG give.PASS.PTCP.SGM; i.e., not going to be given’, /am mazju:d/ ‘NEG add.PASS.PTCP.SGM; i.e., not going to be added ~ increased’, /am mardu:d/ ‘NEG return.PASS.PTCP.SGM; i.e., not going to be returned’, /am maʕlu:l/ ‘NEG carry.PASS.PTCP.SGM; i.e., not going to be carried ~ lifted up’. In addition, constituents of *am*+participle in the examples elicited appeared as predicates in nominal sentences where the subject is a demonstrative or a personal pronoun.

Furthermore, whereas *ab* has a wider distribution, preliminary analysis shows that *am* is restricted to negating participles only, both in the active and the passive forms. However, the case with the participles is not clear cut. The participles can also be preceded by the ‘nominal’ negator *ab* and, in this case, *mā* since the latter and the former are in variation. It seems that the difference between *ab*+participle and *am*+participle is that in the former string, the participle does not have a transitive force, i.e., derived from an intransitive verb and thus, does not take an object; in other words, this would be an example of the participle used substantively as a nominalised adjective (verbal predicates vs. nominal predicates). Also, the *ab*+participle construction could be said to negate the participle in the present tense

rather than in the future tense; in the example in number 51), for instance, *ab* is used to negate the fact that the speaker is not inside the market right now and that he is/has gone somewhere else:

51) *ab* *ra:jiħ* *s-su:g*
 NEG go.ACT.PTCP.SGM DEF-market.SGM
 ‘he is not in the market’, but not ‘he will never go to the market.’

This difference is clearer in the examples below:

52) *he:* *ab* *mizzawwig^j*
 DEM.SGM NEG marry.ACT.PTCP.SGM
 ‘he is not married’ a negative declarative statement denoting a current state.

53) *he:* *am* *mizzawwig^j*
 DEM.SGM NEG marry.ACT.PTCP.SGM
 ‘he will never get married’ to negate a possibility/potentiality in future.

A final note to add here is that in the interviews with the larger pool of informants for the current study, it can be noticed that *ab* often occurred in the interviews, while *am* was virtually absent,⁸³ which could be understandable considering the linguistic and semantic ~ pragmatic distributional nature of the two particles as presented here.

⁸³ A careful examination of these two particles in the larger interview set as well as with regards to how frequently they are actually used in natural conversations in general is needed for a holistic view of their distribution. Apart from the interview data, I did not have much access to these features as used in natural settings.

2.5.4 Negating with *lā* ~ *la*

There is also the prohibitive/negative imperative particle *lā* ~ *la* ‘don’t’. The particle *lā* is also used a) as a co-ordinated negative (*lā... w lā...*), in the sense of ‘neither this nor that’, as shown in example 54 below, b) as an emphatic negative (*w lā ...*) in the sense of ‘not once’ or ‘not a single...’, as example 55 shows, and c) in the sense of ‘so that not’ or ‘lest’ (*an lā ~ anlā*), as shown in example 56.⁸⁴ It can also be repeated (*lāla*) which is similar to ‘no, no’ in English, where the repetition is a token of emphasis, as shown in example 57.

54) *ma: a-ħibb la ðe: w la: ðe:*
 NEG 1SG-like.IPFV NEG DEM.SGM CONJ NEG DEM.SGM
 ‘I don’t like either of these.’

55) *w la: marra ga:l-an kī:h*
 CONJ NEG once say.PFV-3PLF DEM
 ‘they never once said so.’

56) *ʕanla: tʕ-tʕi:ħ*
 NEG 2-fall.IPFV.SGM
 ‘lest you fall.’

57) *b-a-gu:l la:la ma: b-a-tʕlaʕ*
 FUT-1SG-say.IPFV.ACT no NEG FUT-1SG-go out.IPFV
 ‘I will say no. I don’t want to go out.’

⁸⁴ This variant is reported for Bahraini (*Arab*) and described as an instance of *an* ‘*ana*’, a shift from Old Arabic *ʕ→/ʕ/ (Holes, 2016: 396). Holes questions this point since one of the shades of the meaning of the CA preposition *an* is ‘avoidance’ (ibid). I second this hypothesis.

To summarise this chapter, the Bedouin variety under study shows a combination of B-type 1 and B-type 2 features in Oman. This dialect is conservative despite the fact it is spoken in a coastal transitional area. For the Bedouin speech community, a significant factor that probably helped maintain such conservative features is that, within their tribal territory which they have lived in for centuries, they are predominant, not just in numbers, but also in the socio-economic and socio-political status. We also have seen very conservative features in the data from older women who would be found to be generally more linguistically conservative with regard to variation in this variety, as we will see in Chapters Five and Seven.

Chapter Three: Methodology

3.1 Preliminaries

The methodologies used in sociolinguistic research are varied but are mainly designed to “uncover the regularity in interpersonal and intrapersonal linguistic variability that typifies every community” (Milroy and Gordon, 2003: 23). The data sociolinguists collect are the basis of the analyses and conclusions they draw on the speech variation patterns they observe (ibid: 3). For variationists, the role is not merely observing and coming with general conclusions on how the different sub-groups in the communities they study behave with regard to the use of the variants of certain variables; variationists’ ultimate mission is to uncover and to order the constraints (both intra-linguistic and extra linguistic) that “lead to one choice rather another”, and that are at work in a given speech community (ibid: 5). For this aim, variationists: 1) need to collect varied and adequate ‘linguistic data’ which 2) they situate within the ‘social context’ in which this data is obtained, and which 3) they ensure is enough for analysing variation per speaker and per sample (ibid: 23-24). It is also essential that the researchers define the methods they use for data collection, and this is the primary role of this chapter. Chapter Three focuses on the data collection process for this study. It aims to provide the rationale for the sample selection and stratification (§ 3.2), and an overview of the fieldwork and the interview procedure which is the main source of data (§ 3.3-3.6); it also highlights some methodological concerns and issues that may arise due to the nature of the fieldwork and data collection (§ 3.7), and finally provides a brief introduction to the external predictors and dependent variables in this study (§ 3.8-3.9).

3.2 The sample

3.2.1 Defining the sample

One of the essential steps prior to proceeding to the data collection stage is identifying and selecting the subjects for the study (Milroy and Gordon, 2003: 24). This selection is motivated and defined by the research objectives and governs the research design. Ideally, the sample chosen for a variationist study should be ‘representative’ of the larger speech community in question, although this can prove to be a hard task since ‘biases’ in subject selections are difficult to overcome and are sometimes ‘unpredictable’, even with the most ‘randomly’ designed samples (ibid: 24-26). Thus, variationists mostly resort to what is called ‘quota’ or ‘judgement’ sampling through which they choose a limited and well-defined subset of speakers that are pre-defined and stratified along some main social characteristics relevant to the speech community that are hypothesised to influence interspeaker variation (ibid: 30-32). Two main issues are pertinent to the sampling process and are essential steps before the researcher decides on this subset and the size of the sample for the study, namely:

- 1) Defining the boundaries of the speech community under investigation, and
- 2) Assessing and identifying the social characteristics that stratify the speech community (Sankoff, 1980, via Milroy and Gordon, 2003: 26).

With regard to the first step, a number of decisions were made before the final pool of subjects is identified. The study area under investigation has a mixture of Bedouin (B-type dialect) and sedentary population (S-type dialect). Since this study focuses on variation in the Bedouin dialect of the area, informants who speak an S-type dialect are excluded by default. Although the Bedouin community in the area are collectively reported to speak what is locally described as the dialect of al-Suwaiq ‘*lahg^yat swēg^y*’, the tribal origins of the Bedouin community are diverse, and thus, a decision was initially made to limit the study to one

Bedouin tribe to control for tribal origin. Since the *Yāl Sa‘ad* tribe represents the main tribal element in the area (see Chapter One), it seemed reasonable to choose it for the study. In addition, the informants chosen for the interviews have *Sa‘di* parents;⁸⁵ the father has to be *Sa‘di*, but informants with non-*Sa‘di* Bedouin mothers from the area are included as well. Most of the informants eventually interviewed have *Sa‘di* fathers and mothers. On the other hand, if the mother is an S-type dialect speaker, the informant is not considered for the interview. I have one informant whose mother turned out to be an S-type dialect speaker from al-Rustāq, a predominantly sedentary neighbouring town; I found this fact after I interviewed her; her interview is very informative and thus, it is included for comparison purposes.

With regard to the second step which involves defining the social characteristics of the speech community that may influence variation in the dialect, I based the social stratification of the sample on my native knowledge of the study area and on the literature on variationist studies in the Arabic-speaking world in general.⁸⁶ In addition, another decision which was made prior to the fieldwork and which was based on preliminary assumptions on the dialect as spoken by this group, based on discussions with a non-*Sa‘di* Bedouin couple from al-Suwaiq, is to limit the choice of informants to two main localities, namely al-Batha and al-Tharmad, where locality here refers to a cluster of villages. During the interviews with the couple and a number of informal family gatherings, I had the chance to observe and discuss a number of B-type features, and the social parameters that may be at work in this speech community, which I kept in mind when I planned my initial fieldwork.

3.2.2 The sampling method

In the pilot interviews, I had to rely on a friend of mine from the *Yāl Sa‘ad* tribe as well as my sister’s friends and colleagues (see § 3.4 below). Afterwards, these informants helped in

⁸⁵ Belonging to the *Yāl Sa‘ad* tribe.

⁸⁶ Some relevant works are reviewed in Chapter 4 (§ 4.4)

facilitating the interviews with some relatives. When I did my main fieldwork, I contacted the informants from the pilot interviews to facilitate interviewing their relatives and neighbours. The informants' relatives also introduced me to their *Sa 'di* friends. Some of my older informants were encouraged to do the interviews once they knew that I had also interviewed some other 'respected' figures from the speech community. This type of method in allocating and finding potential informants is very common in sociolinguistic research and is widely known as 'the snowball method' (Milroy and Gordon, 2003: 32).

3.2.3 Sample stratification

The sample is stratified according to age, gender, and locality. Age as a stratifying social dimension is well-established in the variationist paradigm (e.g., as discussed in Eckert, 1997). The researcher has to carefully decide on the age group divisions when stratifying their sample. A study on the elderly age group in the Sultanate prepared by the National Centre for Statistics and information (NCSI) in 2012 showed that the average life expectancy of Omanis in 2010 was 76.1 years old.⁸⁷ This study defined elderly as people who are 60 years old and above, in accordance with the UN definition of the elderly. In addition, another publication of NCSI in 2016 defined Omani youth as Omanis who are 15-29 years old.⁸⁸ This means that one way I could stratify my sample was 'etic' (Eckert, 1997) in that it would include the informants aged 60 years old and above in my old age group, and young informants up to 29 years old as my young age group. This leaves informants aged 30 to 59 who would be expectedly included as the middle age group. However, in stratifying age in the sample, I have decided to follow an 'emic' distribution (ibid) because it makes sense for dividing age

⁸⁷ Available at:

https://www.ncsi.gov.om/Elibrary/LibraryContentDoc/ben_Facts%20and%20Figures%202011_6b8b6d1a-aea5-47b1-b158-1fe8b16f89b6.pdf, last accessed [25-04-2015].

⁸⁸ Available at:

https://www.ncsi.gov.om/Elibrary/LibraryContentDoc/bar_Youth%20in%20Figures_5c270ff7-be44-4a41-9921-0cf415db2c10.pdf, last accessed [25-04-2021].

groups in order to better capture the effect of age as a social variable in this speech community. Therefore, the age division here is centred around and reflects the main events in the course of development in Oman since the 1970s in general, and the study area in particular, mainly with regard to the introduction of formal and later on adult education in the study area (see § 1.2.3 in Chapter One). The age groups are:

The older age group= 55+ years old.

The middle age group= 30-54 years old.

The younger age group= 18-29 years old.

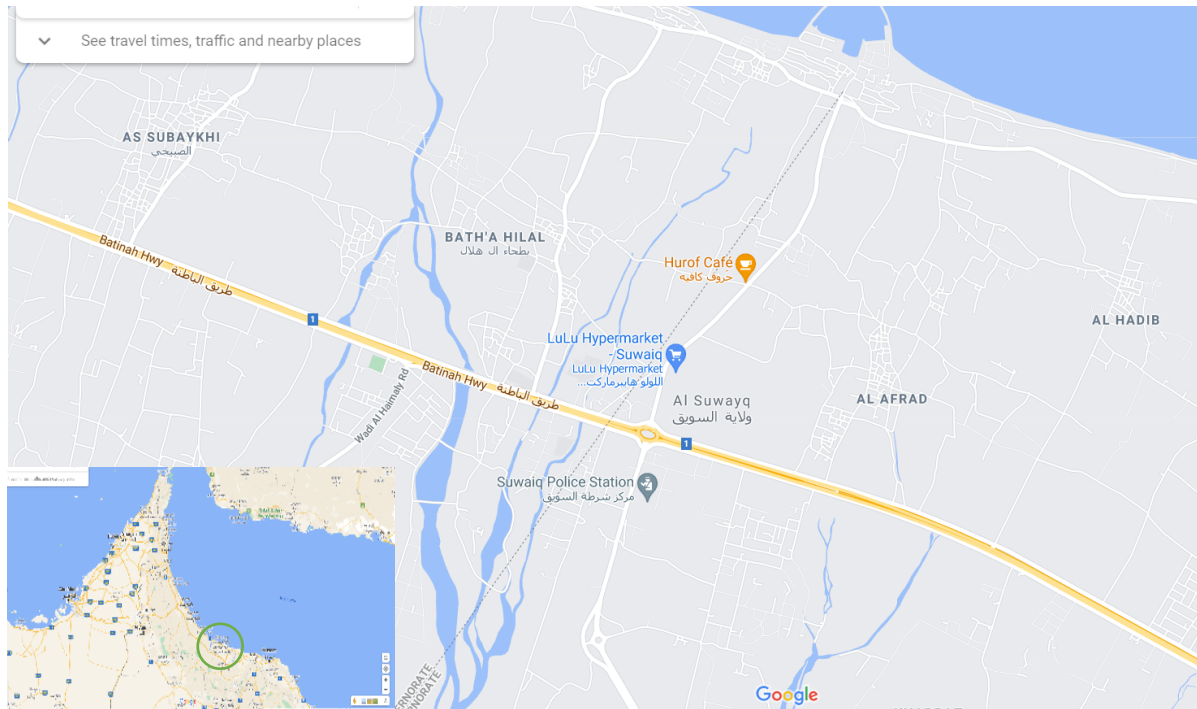
What sets my older age group apart from the rest is that the majority of them did not have a formal education; they could be semi-literate or functionally illiterate.⁸⁹ On the other hand, the informants in the other two age groups are literate and have been formally educated in local schools. What sets the younger age group apart from the middle age group is that the informants in my middle age group were born right before the 1970s to the first half of the 1980s; they would have started school aged seven to ten, at a time when there were few schools in the area. On the other hand, the younger age group participants were born during the second half of the 1980s and the beginning of the 1990s; this group would have started school aged six or seven and their schooling was part of the basic education programme, which was initiated as part of the educational reform in the whole of the country. The reason why this age division is relevant to the study area in general and to the speech community in particular is that at the time when the participants in the middle age group went to school, there were fewer local elementary and secondary schools in the area. This means that students from different villages in the town (and thus different dialectal backgrounds) would go to the same school. Also, at the time when schooling started in the study area, there were

⁸⁹ This does not exclude the possibility of outliers in this and the other two age group divisions.

very few Omani teachers; most of the teachers were Arabic speakers from outside Oman and foreign expatriates. On the other hand, when the participants in the younger age group started school, it was a time when there were many more schools in the town, so these participants mostly went to local schools in the immediate vicinity of their villages, in particular for the primary or elementary stage. By this time, there were relatively more local and non-local Omani teachers than expatriates working in these schools. So, the initial assumption was this background would have an implication on the type of contact the participants were exposed to as children and throughout their schooling years. The participants in the middle age group may have more pressure (from their non-Bedouin peers or non-Omani teachers) to conform to certain linguistic norms which are different from those of their home or social group. This type of pressure may be less intensified in the case of the younger age group due to the nature of the student and teacher composition of their schools.

Furthermore, I have initially decided to divide my sample into two main localities within al-Suwaiq. Locality here stands for a concentration of villages rather than an administrative division. The first one is al-Batha which is representative of the villages in what is locally known as the town centre where the main administrative, business establishments, and main market in the town are located (see Map 8). This locality includes the village of Batha Hilal, where the tribal leaders of *Yāl Sa 'ad* reside, along with adjacent villages with a relative concentration of the tribe.⁹⁰ There are other villages in this area of the town which are mainly populated by non-*Sa 'adi* and non-Bedouin families. The latter group includes S-type dialect speakers who originally came from the mountainous areas close by the town; it also includes Arab and non-Arab expatriate families and other Omanis with non-Arab origins.

⁹⁰ In Batha Hilal (*Baṭḥa Hilāl*) itself, there are *Ḥaḍari* families who have settled in the village for a long time.



Map 8 Location of al-Batha locality (Google Maps, 2021)

The other locality is al-Tharmad which is representative of al-Tharmad area located twelve kilometres away from the centre towards the east (see Map 9). This area and the nearby villages are mainly populated by Bedouin tribes, including the one under study. The original assumption that led to stratifying the sample into these two localities is that since the demographic (and tribal) make-up of the centre of the town (al-Batha and the surrounding villages) is heterogeneous in nature, unlike the more homogenous predominantly Bedouin area of al-Tharmad, more variation is expected to be characteristic of the former, whereas the latter would probably show a higher degree of linguistic conservatism.



Map 9 Location of al-Tharmad locality (Google Maps, 2021)

Another important variable along which sample stratification is designed is gender. For this speech community and the Omani society in general, a division based on gender is realistic and meaningful. Gender is hypothesised to influence the variation at hand especially with the older age group where men and women differ in the level of contact and access to other forms and varieties. During the 1960s, a lot of Omani men went to the Gulf and other neighbouring countries for work. Some stayed there for long periods of time, yet others went back and forth, staying one to two years each time. The situation is different for older women who mostly lived locally. On the other hand, men and women in the middle age group are mostly educated; a lot of the women aged 30-54 are working women who may also have undergraduate degrees. The younger group would consist mostly of educated men and women who may be working or not; they may have or are pursuing undergraduate degrees. In addition, the intersection between age and gender is thus also hypothesised to have an effect on the variation at hand.

3.3 The researcher

I am an insider to the town of al-Suwaiq in that I am native to the area and have lived in the town most of my life. My father was a native speaker of an S-type dialect; my mother, on the other hand, was a speaker of a coastal B-type dialect in the town, which shares some main features with the dialect of *Yāl Sa‘ad*.⁹¹ Nonetheless, all of these dialect varieties are mutually intelligible. I also attended local schools where both dialect types were spoken and had teachers and friends from both backgrounds. I may consider myself to be a ‘close outsider’ to the speech community under study since although I am a local, I do not belong to the ‘camel and cattle rearing’ Bedouin community in the area. I may also be considered as a stranger to the informants in that I met the majority of them for the first time during the interviews, although this may not be true for the people who assisted me with arranging and conducting the interviews (§ 3.6). But what helped break the ice is the fact that many of my informants knew my father or other members of my immediate and larger family.

3.4 The pilot study

Pilot interviews have proven to be a useful tool in sociolinguistic research, since they help in forecasting challenges and obstacles the researcher may face while conducting the larger research project and thus, give the researcher the advantage of editing and improving their research design accordingly (Milroy and Gordon, 2003: 141). This is the reason why I went

⁹¹ Interestingly, the Bedouin ‘camel and cattle rearing’ community in al-Suwaiq differentiates between their dialect and the B-type dialect of the villages along the coast in the area. They claim that people living in the immediate adjacency of the shore are ‘*hal s-sāḥil*’, which literally means the people of the coast. This group traditionally mainly worked as fishermen and many still do. In fact, a quick survey of the villages closer to the coast, which happens to be where I live, shows indeed that the people who are residing or have originally inhabited this area come from various tribal origins and different places in Oman. For example, my mother’s family moved from another predominantly Bedouin area up to the north of al-Suwaiq and settled in a coastal village to the southeast of the main market area, whereas my father’s family moved many decades ago from a sedentary mountainous village in the town of Ṣaḥam in the north-west of the Bāḥina coast before they founded and settled in a coastal village to the other side of the market. I speak my father’s variety which is the main variety spoken in the three coastal villages, in the immediate adjacency of and to the northwest of the centre of the main market area: [q], [gʲ], and [kʲ] as the traditional reflexes of *q, *g, and *k, and [-iʃ] for the 3rd SGF suffix.

to Oman in May 2017 to conduct a pilot study for a week. There were eight interviews in total, with twelve informants. The interviews were conducted in the informants' houses except for one group interview which took place in a local primary school. The main reason I conducted this pilot study is to observe the linguistic features of the dialect as spoken by the *Sa'di* population, and further discuss some of the issues that were raised and commented on by my two initial non-*Sa'di* Bedouin informants.

In addition to the topics typically discussed in sociolinguistic interviews (§ 3.6), the pilot interviews conducted were also exploratory and metalinguistic in nature; there were questions about the tribe, the differences- if any- between the different localities, and the salient features of the dialect. The pilot interviews helped me to identify an extra set of potential variables, but most importantly they helped me to narrow down my initial list of variables that I wanted to explore, since I was able to assess which variables would be feasible for thorough analysis in this study and which would not. For example, I originally was intrigued by the negation paradigm in the dialect and the pilot interviews confirmed that indeed there is variation in the negation system in this dialect, and this variation seems to be motivated to a large extent by locality; however, I noticed that one of the variants, namely *am*, almost did not occur, due to the nature of linguistic and pragmatic distribution of this variant (see § 2.5 in Chapter Two); such low occurrence in the interview data would pose an issue for a multivariate analysis of this syntactic variable. Another important contribution of the pilot interviews in this study is identifying locality as a speaker variable. The informants stressed the fact that there are indeed dialectal differences between the two localities of al-Batha and al-Tharmad. Although they mainly highlighted lexical differences, e.g., *šāwar* (al-Tharmad) and *barra'* (al-Batha) 'outside', I noticed some variation in the use of some phonological and morpho-phonological features, such as the vowel harmony in the 3SGF object/possessive suffix [-ha] ~ [-hi] which is realised as [-ho] following the long mid-back

vowel of the 3PLM subject suffix [o:] as in /ʕammur-o:-ho/ ‘build.PFV-3PLM-3SGF.ACC’. The informants also differentiated the localities as ‘more’ or ‘less’ Bedouin with regard to the demographic make-up and to the dialect as spoken in different areas of the town. They also stated that the null definite article is mainly distinctive of al-Tharmad locality; this variable is eventually chosen as one of the variables for this study.

3.5 The fieldwork

The main data collection was carried out in the winter of 2017-2018. Although the initial plan was to limit the research area to al-Batha and al-Tharmad localities in al-Suwayq, due to the unavailability of a satisfying number of informants from these localities (see § 3.6), I decided to extend the research area to include the neighbouring locality of al-Khadhra (15 kilometres to the west of the town’s centre; see Map 10) and the neighbouring villages of al-Qaraṭ and al-Miladda which are administratively part of the bordering town of al-Miṣin‘a. The latter two were included in al-Tharmad locality since they are closer to al-Tharmad village itself.



Map 10 Location of the three localities in this study (Google Maps, 2021)

Figure 3.1 shows the distribution of the villages where my informants come from across the three localities. I decided on this division with the help of a *Sa‘di* informant and according to the subdivisions of al-Suwaiq as mentioned in Al-Shibli’s (2010) study on the town.

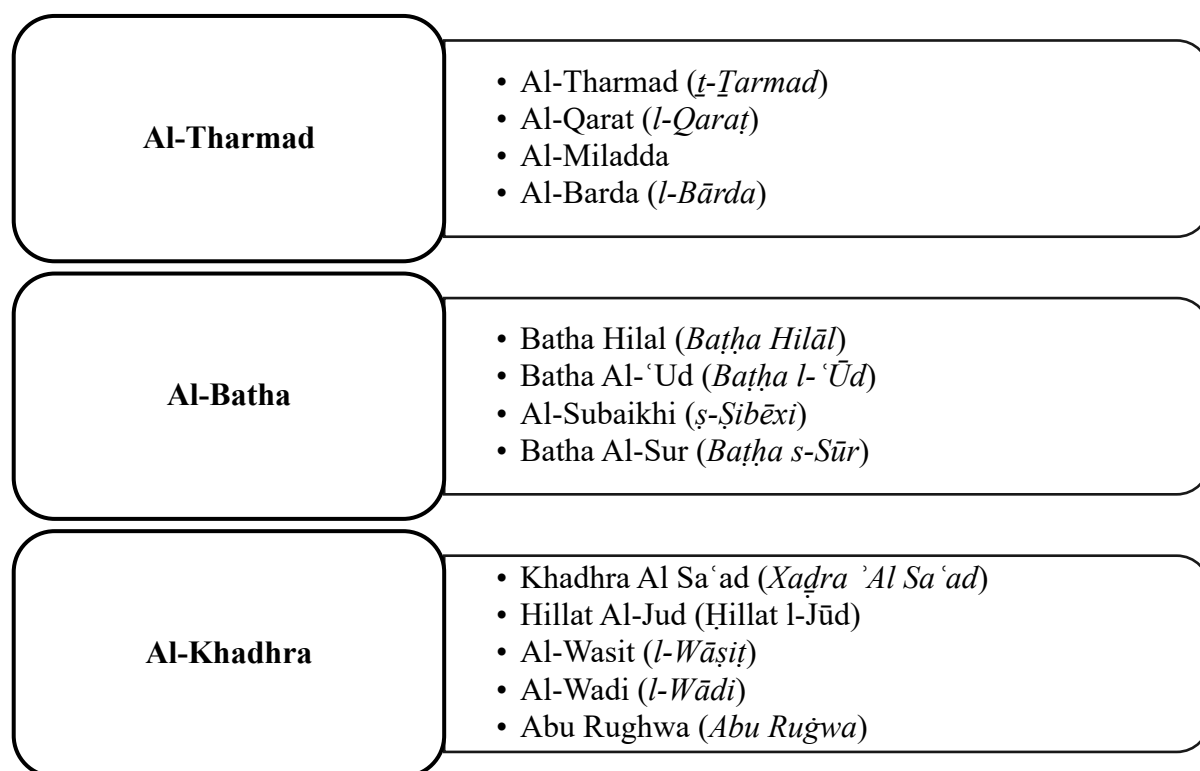


Figure 3.1 The distribution of informants’ villages across three localities.

3.6 The interview procedure

The data collected for variationist research, in addition to being varied and adequate, it ideally needs to be 1) representative of the language as used by the speech community under study, and 2) naturalistic and spontaneous. Naturalistic speech is normally obtained when the speaker uses his or her ‘vernacular’, a key concept in variationist studies which is broadly defined as the speech that the speakers produce when they are not paying attention to how they talk, particularly because they pay more attention to what they say [i.e., content] than to how they say it [i.e., the language] (Labov, 1984: 29). Obtaining this kind of naturalistic data “provides the most systematic data for linguistic analysis”, and it is the main purpose of

‘face-to-face’ interviews as a data collection tool (ibid). Face-to-face interviews are also significant in order to obtain large amounts of speech samples sufficient enough for carrying quantitative analysis (ibid). However, the type of speech obtained from face-to-face interviews can be subject to the formality constraints imposed by the sociolinguistic interview procedure itself due to the ‘observer’s paradox’, since the speakers most likely will not use their vernacular in first time encounters with ‘strangers’ (ibid: 29-30).

Twenty-seven interviews were conducted with speakers of the *Yāl Sa‘ad* tribe from different localities in Suwaiq and the neighbouring town of al-Miṣīn‘a. Together with the pilot study conducted in May 2017, I have a total of 36 interviews and 44 speakers. In total, I have ten group interviews with two to three informants per interview. The total time of the interviews was about 33 hours and the average duration of the interviews was 55 minutes.

At the beginning of the trip, I could not obtain any speakers, because most of the people I was planning to interview were busy with different social obligations, and college-age students were busy with end of term exams. Other speakers work out of town, so I had to wait for weekends to interview them. I conducted my first set of interviews end of Dec-2017. Another challenge besides time limitations, is that some members of the community refused to conduct the interviews for different personal reasons and others did not want to be recorded. In general, women were more willing to be interviewed than men. The informants who agreed to be interviewed were contacted beforehand, because I found that most people approached without prior arrangements or not through an intermediary refused to be interviewed. For example, my cousin arranged an interview with her daughter’s teacher. There are five *Sa‘adi* teachers in that school, however, I managed to talk to two only, whereas the other three teachers did not wish to take part in the study.

I had a male assistant who conducted some of the interviews with male informants with me and others by himself. I chose him because of his knowledge of the Bedouin culture

in the town. I trained him on the procedures of conducting sociolinguistic interviews and the use of equipment. He is a speaker of an S-type dialect local to the area. Not all of the interviews were conducted by the male assistant and me. Few of the interviews were conducted by others. In addition, the informants were informed about the topic of the study and the interview procedures. An oral consent was obtained prior to the recording, followed by a written consent at the end of the interview; the participants were also informed that they have the right to withdraw their participation at any time if they wish to. They were given an information sheet explaining the topic, the interview, and data management procedures. Some informants were illiterate, so we had to explain the procedures of the interview and data management to them.⁹² The recorders used in the interviews were Zoom H4n Pro for the main field trip to ensure a good quality recording, and a Sony recorder ICD-UX543F for the pilot interviews; the interviews are stored in an external hard disk and on my personal laptop, both of which are protected by passwords. I also recorded the interviews on my iPhone as a backup but then immediately erased them once transferred to my laptop.

The interviews were generally semi-structured, in that while the questions are designed to obtain certain type of information or speech, the speakers have the chance to speak or elaborate on subjects of interest to them to divert their attention from language and thus, encouraging the flow of naturalistic data (Schilling, 2013: 108). The questions were mainly based on the culture and everyday life of the speakers. Demographic information about age, education, and the places where the informants lived were obtained first, followed by general questions which ranged from inquiries about past times, childhood memories and games, marriage customs, dress and jewellery, houses in the past, social relations in the village, and the main social celebrations, to questions about the study area before and after

⁹² Eckert (2013) and Milroy and Gordon (2003: 79-87) for a discussion of the ethical considerations and data management for conducting this type of research.

the 1970s. Similar topics are discussed in group interviews. In addition, some speakers spoke more about other personal experiences like going abroad for study. Younger speakers were also asked about their interests, their studies, their specialties, favourite food, and leisure time activities. Towards the end of the interviews some informants were directly asked about the history and current demographics of the tribe, and the dialectal differences in the different localities.⁹³

3.6.1 Distribution of the informants

Locality	Gender/age	Old	Middle	Young	Totals per gender	Total per locality
Al-Tharmad	F	3	3	3	9	14
	M	2	2	1	5	
Al-Batha	F	3	4	4	11	16
	M	2	2	1	5	
Al-Khadhra	F	2	2	1	5	10
	M	1	2	2	5	
Totals per age		13	15	12	Total N=40	

Table 3.1 Distribution of the informants across age, gender, and locality.

Three interviews were excluded from the analysis. In one interview, the informant used MSA most of the time. The second interview was excluded because it was very short.

⁹³ This is primary meant to elicit the speakers' attitudes towards the dialect as spoken by this community, and their perception of what social context is relevant to my study (Schilling, 2013: 104-107).

The third interview included two younger men from al-Batha; it was excluded because there was a lot of overlap. The remaining forty speakers are distributed as in Table 3.1.

3.7 Methodological considerations

One of the limitations of the fact that I am considered a stranger to many of my informants is that the type of demographic and social data that I could obtain from the interviews may be limited. This would have implications on the range of social predictors that may have an effect on variation in this speech community but which I would not be able to thoroughly investigate due to the lack of certain information, e.g., as in the case of creating elaborate social network designs or contact indices. In addition, being an outsider to the speech community may add to the ‘observer’s paradox’ (Labov, 1972) in that the informants may consciously or sub-consciously shift from the vernacular to a more formal or more accessible level of language, which in turn is likely to affect the flow of naturalistic speech. Another methodological issue is the type of bias that results from using semi-structured sociolinguistic interviews as a data collection tool. One source of bias stems from the fact that maintaining a smooth and free flow of speech for different types of participants, i.e., different demographic and social backgrounds, would influence the type of data obtained. The same topics that are intriguing for the older speakers are not necessarily so for the younger ones. In other words, the data gathered may be biased towards a certain type of lexical forms, or certain variants in this case, since the data gathered from the older vs. the other younger age groups, for instance, are relatively different. For example, in the case of linguistic variables which are sensitive to the lexical status that is associated with the traditional dialect, the distribution of the traditional variants would probably be wider in the speech of the older speakers, since most of the interview questions for this group focus on the traditional life style, as in the case of the DEF variable, for instance; this is not necessarily the same with younger informants who would prefer to talk about other things, thus triggering less instances of the traditional

form, making any inferences on the variation based on the proportion of use of the variant not totally objective.⁹⁴ In the same respect, the more innovative the variants are, the more they are normally evident in the speech of the younger informants due to the nature of interview questions, as well speakers' demographics, an in the case of the affricate variant of (dʒ).

3.8 The external predictors in this study

Figure 3.2 presents the external predictors in this study:

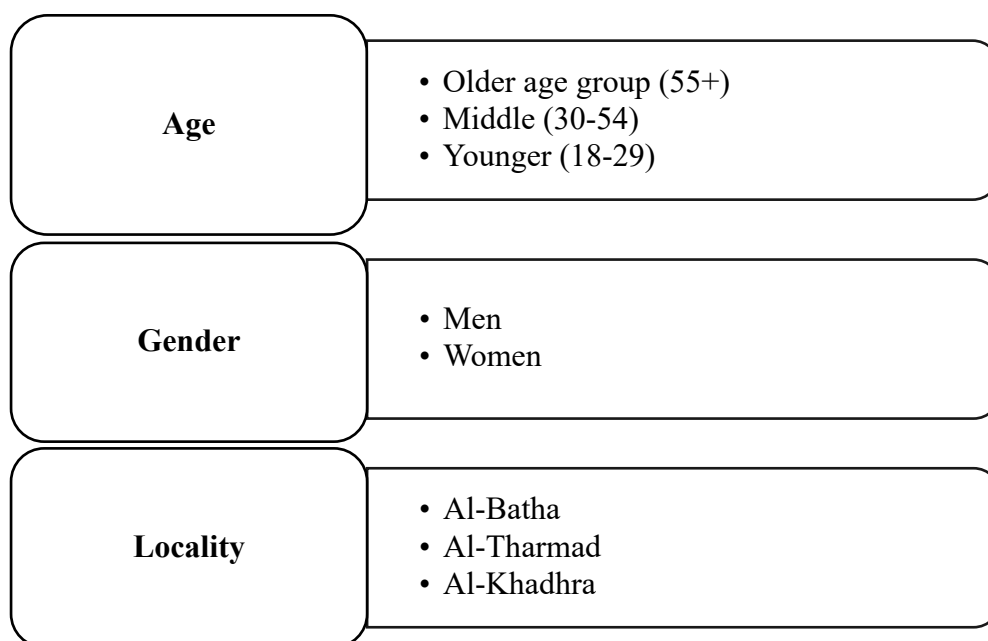


Figure 3.2 The external predictors in this study.

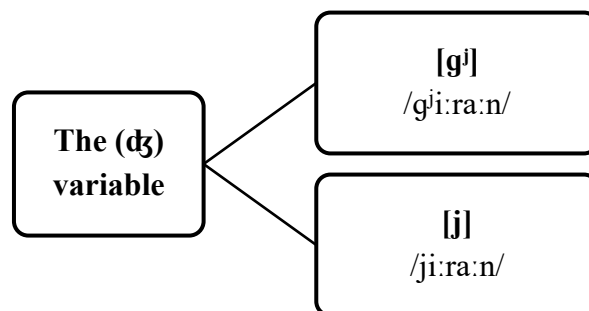
⁹⁴ In the case of the passive verbs, as another example, the type of questions addressed to the older speakers may have resulted in higher occurrences of the apophonic passive forms. Some of the questions were about processes, like date harvesting and storing, or explaining the way of making something like a traditional meal where the agent is not important. This methodological bias resembles that which Holes described for his (1998) study of the internal passive in Oman.

3.9 The dependent variables in this study

In the treatment of the data for the two variables statistically analysed in this study, namely the (dʒ) variable and the DEF variable, both variables are treated as ‘categorical’, i.e., the values for the dependent variable belong to “distinct categories rather having a range of intermediate possibilities”, and ‘binary’, i.e., containing two categories for the dependent variable in this case, that can include the ‘presence’ ~ ‘absence’, or the ‘variation’ between two possible responses (Johnson, 2013: 290). Following is a brief introduction to these variables.

3.9.1 The (dʒ) variable

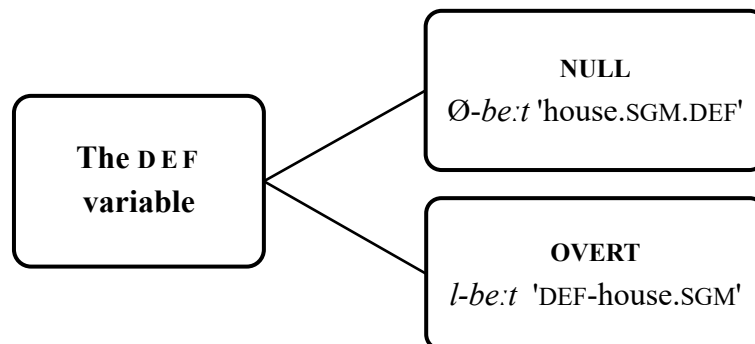
The (dʒ) variable is one of the salient phonological variables in Arabic sociolinguistics. Its importance probably stems from the fact that it generally encodes or indexes the macro divisions that stratify the Arabic speaking world, e.g., urbanity vs. rurality, Bedouinness vs. sedentariness, sectarian and religious affiliation, and ethnicity. In the Gulf area, the main variant of (dʒ) is the glide [j]; this is also the traditional (dʒ) variant in the dialect under investigation. The other main variant which is local to the study area in particular and to Omani Arabic in general is a palatalised velar [gʲ]. The glide is currently in variation with the latter, as in: /ji.ra:n/ ~ /gʲi.ra:n/ ‘neighbours’.



3.9.2 The DEF variable

An intriguing feature of the dialect of the *Yāl Sa‘ad* tribe in the area is the null definite article, which is another variant of the definite article in this dialect, in addition to the overt definite article /l-/. So, the DEF variable involves variation between the overt article /l-/ generally used to mark definiteness in the dialect, and a null article which is also used as a DEF category. This means that some NPs that have the potential to show the definiteness marking through adding the prefix /l-/ variably do not show it, as the following example from an interview with a young woman from al-Tharmad shows:

umm-je ma ti-tʔaʕ, ti-jlis fi Ø-be:t ʕaʕa:n ij-j-in hni: j-sallm-an hiri:m l-ħa:ra
 mother.SGF-1SG.GEN NEG 3-go out.IPFV.SGF 3-sit.IPFV.SGF in house.SGM.DEF CONJ
 3-come.IPFV-PLF here 3-greet.IPFV-PLF woman.PL DEF-village
 ‘My mother does not go out (at Eid); she stays in [the] house so the village women come here to greet her.’ (S22FTH)



To my knowledge, Alqahtani’s (2015) treatment of the definite article in *Tihāmi Qaḥṭāni* Arabic is the only quantitative variationist analysis of the definite article in Arabic (§ 6.8 in Chapter Six). Her study deals with the phonological variation between two different forms or reflexes of the article, namely [l-] and [m-]. However, as far as this study is concerned, /l-/ is not a phonological variable, rather a morphosyntactic one where the two variants in question are coded in terms of whether the DEF category is marked overtly (OVERT) or zero marked (NULL).

Chapter Four: Literature on the (dʒ) variable

This chapter presents the background for the analysis of my first sociolinguistic variable, (dʒ), which is a phonological one. The (dʒ) variable is one of the main variables that are commonly studied in the Arabic-speaking world. An overview of the relevant descriptive and sociolinguistic literature is presented in this chapter, with a focus on the historical development of the Arabic phoneme /dʒ/, and the current trends and trajectories of variation and change.

Overview of the literature of Arabic /dʒ/

4.1 Diachronic development(s)

Dialectological and historical studies that focus on Arabic /dʒ/ more or less agree on certain hypotheses in the history and development of this phoneme:

- The etymology of Arabic /dʒ/ goes back to a reconstructed proto-Semitic voiced velar plosive *g, part of a homorganic dorsal consonantal triad which also includes the voiceless velar plosive *k and a glottalised *q (see Cantineau, 1966: 88-9).⁹⁵
- Arabic /dʒ/ witnessed a number of shifts that resulted from several (internal) phonological processes throughout its history, ending in a wide range of realisations for this phoneme in the Arabic dialects as spoken today.

⁹⁵ Watson (2002: 2). Consonants forming a triad cannot co-occur in the etymon of the word (Boudelaa and Marslen-Wilson, 2001 and Watson, 2002, via Embarki, 2013); also Greenberg (1978, via Holes, 2004: 99) on distributional constraints on homorganic consonants within the Arabic system.

- Historically, /dʒ/ in Arabic has developed in different trajectories, but mainly through two main routes (visually represented in Figure 4.1 below):

⇒ /g/ → /gʲ/; or /g/ → /gʲ/ → /j/

⇒ /g/ → /dʲ/

(Cantineau, *ibid*; Al-Nassir, 1993: 44)

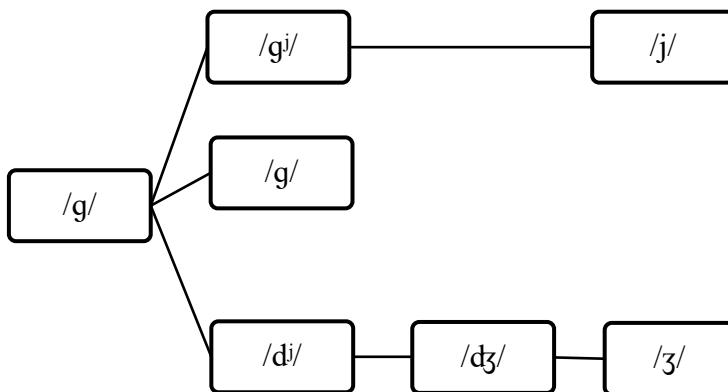
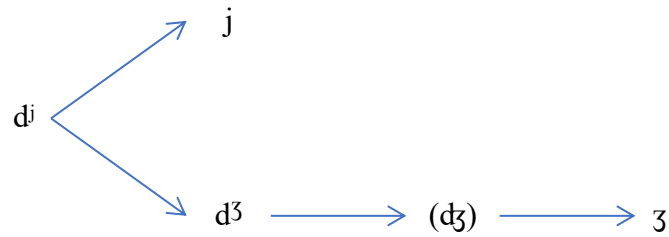


Figure 4.1 Diagrammatical pattern of the phonetic changes of Old Arabic /dʒ/ (adopted from Al-Nassir, 1993: 44).⁹⁶

According to Cantineau (*ibid*: 89), /dʲ/, the alveolar with the off-glide in *g > dʲ was either: a) lenited, resulting in the glide /j/ through dropping the stop part, or b) fortified through changing the off-glide secondary articulation to a fricative one, consequently resulting in an affricate /dʒ/ which may lose the stop rendering a fricative realisation /ʒ/, as the following diagram shows:

⁹⁶ All of the variants listed in this figure are voiced, regardless of the change in the manner and place of articulation (Al-Nassir, 1993: 43).



- In addition, the /dʒ/ of Classical Arabic (CA), which is described as a ‘*majhūr*’ ~ voiced, and ‘*šadīd*’ ~ stop sound is produced “in the middle part of the hard palate [placed] together with the *šīn* [i.e., /ʃ/] and the *yā*’ [i.e., /j/]” (Sibawayh, in Al-Nassir, *ibid*: 41) is a palatalised velar stop /gj/ (also in Cantineau, 1966: 90; *cf.* Kaye, 1972).⁹⁷⁹⁸

4.2 Current geographical distribution of /dʒ/ and key issues

Sources on Arabic /dʒ/ have established that variation in /dʒ/ has not developed recently and that since earlier times other variants co-existed with those of pre-Classical and Classical Arabic (e.g., Cantineau, 1966: 89; Johnstone, 1965; El-Gindi, 1983: 458-61).⁹⁹ This dialectal variation carries on in contemporary Arabic varieties. This section sums up the different documented realisations of /dʒ/ and their main geographical distribution within the Arabic-speaking world.

⁹⁷ Kaye contends with the argument that there existed an Old Arabic koiné in pre-Islamic times which had the fricative /ʒ/ as its realisation of /dʒ/; so, for Kaye, the affricate variant of the Standard Arabic /dʒ/ has diachronically resulted from this fricative: Semitic /g/ → pre-classical /ʒ/ → (late) Classical /dʒ/.

⁹⁸ Blanc (1969: 18) states that Arabic /dʒ/ in Sibawayhi’s account is ‘an unaspirated *majhūra* [=voiced] stop with an audible release; it is medio-palatal and has the same place as the glide /j/ and /ʃ/; it also has ‘a spirant and perhaps affricated allophones’.

⁹⁹ Parrado (2019) for an analysis of the history of /dʒ/ in the Magribi (North African) dialects. Parrado argues that variation was indeed characteristic of the Arabic that was first introduced to the area (*ibid*: 139).

[g]

The proto-Semitic variant *g, i.e., /g/ as in /gamal/ ‘camel.SGM’, was a voiced velar stop. This stop is the /dʒ/ reflex in a number of Semitic languages like Akkadian, Aramaic, Hebrew, Ethiopian (Al-Nassir, 1993: 41), and Modern South Arabian (Rabin, 1951: 31). As a dialectal reflex, [g] was disliked by earlier Arab grammarians whose accounts describe this variant as a *kāf*¹⁰⁰ which is between the *kāf*, i.e., Arabic /k/, and *jīm*, i.e., Arabic /dʒ/, and which they considered an unacceptable form (Sībawayh, 1982: 432).¹⁰¹ The markedness of [g] as a variant of /dʒ/ is reflected in its rather restricted distribution in modern day Arabic; it is described for literary and colloquial Cairene Arabic, and for some Yemenite dialects in the southern coast of Arabia,¹⁰² along with the sedentary interior dialects of Oman (e.g., in Al-Nassir, *ibid*: 42; Rabin, 1951: 31; Cantineau, 1966: 90; Holes, 1989: 451-54 and 2004: 59; Woidich and Zack, 2009; Blanc, 1969: 21).¹⁰³¹⁰⁴

Studies concerning this realisation in Egypt adopt opposing views regarding the occurrence of [g] for /dʒ/ in various Egyptian dialects including Cairo. The first view is that of Hary (1996; previously adopted by Blanc, 1969 and Kaye, 1972) in which the contemporary [g] realisation of /dʒ/ in Egyptian Arabic is an innovation. Hary further proposes a historical timeline that traces a ‘linear’ A>B>A phonetic development of /dʒ/ in this variety (summarised in 4-2 below), by which the development from [g]→[dʒ]→[g] included some phonetic shifts and stages of variability ~ instability, and is not a merely phonetically

¹⁰⁰ A reflex of an underlying /dʒ/ in Al-Nassir (*ibid*).

¹⁰¹ The reflex [g] is considered by Arab grammarians as the *correct* realisation of early CA /dʒ/ which had been influenced by other variants and eventually started to be commonly realised as [dʒ] or [ʒ] (Cantineau, 1966: 90).

¹⁰² For Ta‘izz and Huggriyyah (Watson, 2002: 16). See Rabin (1951: 31) for /dʒ/ in the Yemenite dialects.

¹⁰³ Cantineau also reports restricted ‘unexplained’ occurrence of [g] as a /dʒ/ variant of some Bedouin nomads’ Arabic in North Arabia (*ibid*: 91).

¹⁰⁴ Il-Hazmi reports this variant as well for some fishermen groups of the Banu Şubḥ (1975: 45).

motivated shift, but may also be motivated by social factors making it a sociolinguistic shift as well.

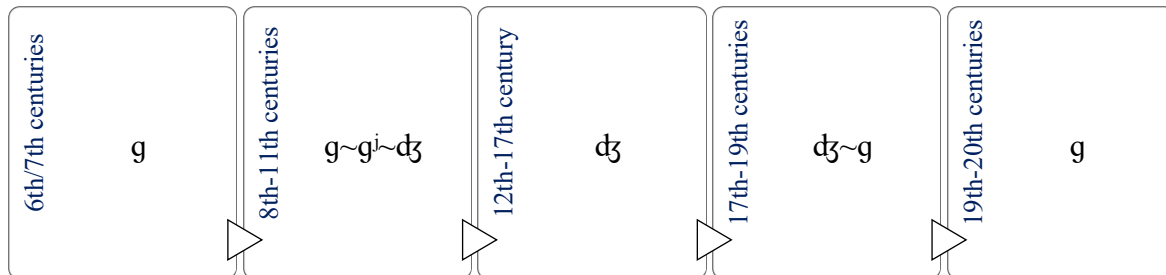


Figure 4.2 The development of the /g/ reflex of /dʒ/ in Urban Egyptian Arabic (based on Hary, 1996).

On the other hand, Woidich and Zack (2009) reject the scenario proposed above partly on the grounds of the credibility and reliability of the evidence used by Hary (1996), but mainly on the basis that:

- [g] in Lower Egypt is a relic feature; it is not an innovation, nor a result of a ‘linear development’ or ‘back-shifting’; it can be traced to the time of the Arab conquests;
- Speakers of [dʒ] co-existed alongside speakers of [g] for a long time in this area; a comparison between old and new dialect maps for the former two variants reveals a similar geographic pattern in terms of isoglosses which are still evident today. Based on this evidence, the researchers conclude that [g] must be an archaic feature since it takes a considerable long-time span for such geographical patterns to emerge and develop (ibid: 55).

[gʲ] and [j]

As mentioned in § 4.1, the pre-classical Old Arabic reflex of /dʒ/ was most probably a palatalised stop [gʲ]. Cantineau comments on the articulation of this variant as a lenited pre-palatal that it is hard to distinguish from /dʲ/; it is basically a velar followed by a palatal off-glide (1966: 91). This realisation is documented for some modern varieties. Cantineau reports this variant for the dialects of the nomadic Bedouins of the north of the peninsula, especially *Šammar* and *Anaza*, e.g., in /gʲabha/ ‘forehead’, /nʃagʲa/ ‘goat’, and /gʲefen/ ‘eyebrow’ (ibid: 91). Il-Hazmi also talks about [gʲ], a ‘free’ variant of /dʒ/ among the Northern Central group,¹⁰⁵ which he describes as ‘a hard palatal plosive’ that occurs in all positions alongside the main variant [dʒ], e.g., /tirgʲaʃ/ ‘you/she return(s)’, /rigʲa:lin/ ‘men’, and /gʲa:ʃ/ ‘he came’ (1975: 52-5). In addition, Kaye reports that this variant is used in Oman but rarely in Najd (1972: 42).

On the other hand, a palatal plosive /j/ is reported as another variation of the former realisation of Arabic /dʒ/. Watson states that early Classical Arabic /dʒ/ was realised either as a voiced palatalised velar stop [gʲ], or as a voiced palatal stop [j] (2002: 15-16, and the references therein). In modern day Arabic, the phoneme is realised as a voiced palatal stop [j] in parts of the Arabian Peninsula, as well as Upper Egypt, parts of Sudan, and some northern Yemeni dialects (ibid). Rabin comments on the instability of the early Yemenite /dʒ/, the equivalent of the CA variant as depicted by the Arab grammarians and reports a ‘pure palatal’ [j] which has a rather ‘squeezed’ pronunciation of the velar [g], almost [gʲ] and which is still used in southern Yemen (1951: 31, and the references therein). In addition, Ingham comments on this variant in Meccan speech as bearing “the strongest resemblance to the speech of the Sudan” and that the “number and realisation of [these] phonemes are in fact

¹⁰⁵ Namely for a group of *Harb* in north central Najd, but also, for the *Hijāzi Rubuga*, *Miṣāʾila* and *Sihliyya* groups (ibid).

almost identical to the koiné dialect of Khartoum as described by Trimmingham (1946)” (1971: 273).¹⁰⁶ Holes also mentions a palatal realisation for /dʒ/ in the Šī‘i Būri and ‘Āli villages, which also have a velar [g] for this phoneme and either [c] or [k] for /k/, unlike the rest of the Šī‘i community who use the affricate counterparts (1980: 80-81).

[dʲ]

The other variant of the /dʒ/ with the palatal off-glide, namely [dʲ], seems to be even more restricted distribution-wise than the three variants described above. Diachronically, this reflex has emerged as a result of the fronting of the older form /qj/ (see § 4.1). In modern Arabic varieties, it is reported for some ‘inland districts’ of Palestine and Syria and also can be found in Upper Egypt, but rarely in Lower Egypt (Kaye, 1972: 43). It is also documented as a variant of /dʒ/ for the camel rearing Bedouin Ḥarbs of Ḥijāz and some nomads of the Northern Central group of the *Ḥarb* tribe (Il-Hazmi, 1975: 53-55), where this reflex is said to be accompanied by affrication and occurs mostly in word-initial position, like in /dʲba:l/ ‘mountains’ and /dʲimal/ ‘a camel’ (ibid).

[dʒ] and [ʒ]

The voiced palato-alveolar affricate variant of /dʒ/ is believed to have developed from [dʲ] through spirantising the secondary palatal articulator into a fricative one (see § 4.1). According to the traditional grammarian accounts, this variant came to be used as the ‘normative’ realisation of /dʒ/ in Classical Arabic around the 11th century AD (e.g., Ibn Sīna, in Blanc, 1969: 23).

¹⁰⁶ Transcriptions in the source.

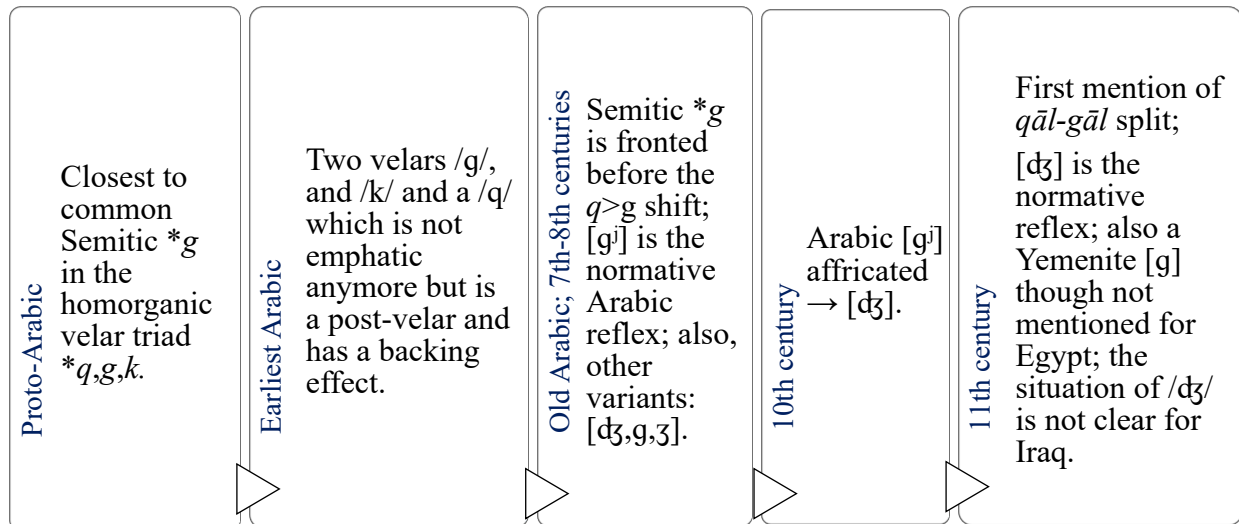


Figure 4.3 A reconstruction of /dʒ/ (based on Blanc, 1969).

The diagram in Figure 4.3 presents a reconstructed timeline for /dʒ/ as depicted in Blanc (1969). In contemporary dialectal Arabic, the affricate [dʒ], as in /dʒabal/ ‘mountain’ and /dʒaras/ ‘bell’, is perhaps the most common modern realisation of /dʒ/ in the Arabic-speaking world. It is the current variant in MSA and occurs in a number of modern Arabic dialects where it mainly features as a west peninsular feature. Within the peninsula, it is found in Najd (e.g., Ingham, 1982; 1994). It is found in some parts of the Gulf (e.g., Holes, 1980, for the Baḥārna in Bahrain) where it has intricate prestige statuses, and to a lesser extent in Oman (Holes, 1989: 451-53). It is also found in Iraq, Yemen,¹⁰⁷ and in rural and Bedouin centres in the Levant region,¹⁰⁸ in other parts of the Arabic-speaking world, [dʒ] is primarily a desert feature found in the deserts of Libya and Algeria, and the Sinai desert.¹⁰⁹

¹⁰⁷ In the central region of northern Yemen (Behnstedt, 1985, via Watson, 2002: 15). Cantineau (1966: 91) comments that this variant is the most common in Yemen, and also it is prestigious in Iraq.

¹⁰⁸ In the Syrian desert, /dʒ/ is categorically an affricate as spoken by small nomadic Bedouin tribes, Cantineau (ibid: 92).

¹⁰⁹ Cantineau (ibid).

Furthermore, a related variant is the fricative [ʒ]. It is the norm in Lebanon, in major urban centres along the Mediterranean in the Levant and North Africa.¹¹⁰ This realisation is also criticised by early Arab grammarians,¹¹¹ along with the velar realisation [g], based on which Cantineau infers that these two variants are probably very old, as old as or maybe older than the CA analogue (1966: 89). However, the [dʒ]→[ʒ] change is considered as a relatively recent one for the Levant, taking place around the 18th-20th centuries (Al-Wer, 1991: 179; more in § 4.4). The fricative realisation is also reported for Medini Arabic, where it is gaining prestige as an urban feature (Hussain, 2017; more in § 4.4).

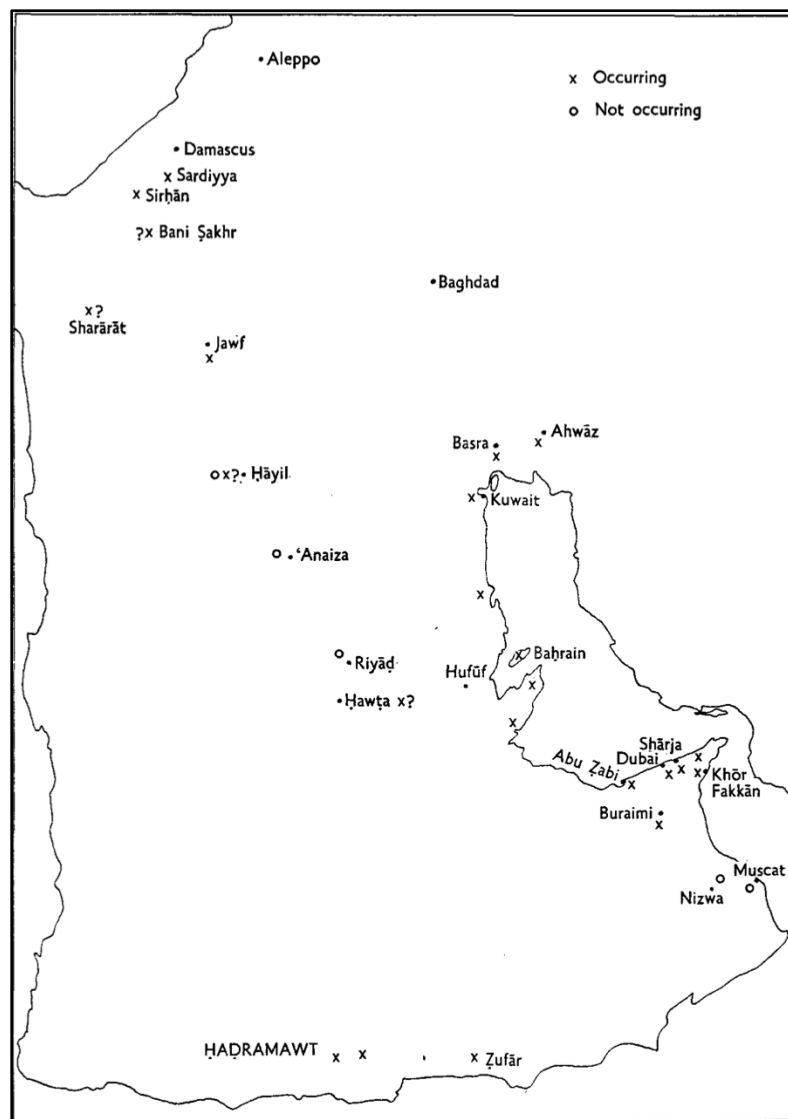
[j] and Ø

One of the earlier developments of /dʒ/ in Arabic is the palatal glide /j/ which resulted from a historical stop→glide lenition process. As shown in § 4.1, this lenition process could have taken two main routes: 1) through a direct development from the older Arabic form [qʲ], or 2) as an end product of a lenition process of the former that involved an intermediate stage where [dʲ] was the norm. Arabic varieties which underwent this change also retained the Semitic glide *y. In the older dialectal peninsular varieties, [j] as a reflex of /dʒ/ occurred sporadically in the speech of some Najdi tribes, primarily Tamīm, where it also triggered variation in the quality of the preceding vowel, e.g., /ʃaɖʒar/ ~ /ʃijar/ ‘trees’ (Johnstone, 1965: 233-34). In addition, Johnstone states that it is not clearly indicated in Arabic philological references whether *g→[j] was a general sound change like the one evident in modern Arabic dialects, but “there are a number of specific cases which suggest this may have been a dialectal peculiarity at early times” (ibid: 234).

¹¹⁰ See Cantineau (ibid: 93) for a detailed distribution; also, in Holes (2004: 59). In addition, Parrado (2019) for a detailed current distribution of the /dʒ/ variants in North Africa.

¹¹¹ Another unacceptable form of the Arabic /dʒ/ by Arab grammarian, the *jīm* (i.e., /dʒ/) which is similar to the *šīm* (i.e., /ʃ/) (Sībawayh, 1982: 432).

In modern Arabic dialects, this variant is mainly a Bedouin dialectal feature which occurs along what resembles a curve in the Arabian Peninsula (Johnstone, 1965: 241). It is reported for some dialects for Ḥaḍramawt in Yemen, for some northern peninsular tribes, the tribes dwelling in the lower Euphrates area, for the eastern part of Arabia along the coast of the Gulf, and towards the south in Ḍufār (ibid), although it is absent in Central Najd (see Map 11). It is also found in the Algerian Sahara and Upper Nile regions in Egypt (Al-Nassir, 1993: 43).



Map 11 Distribution of the [j] variant of /dʒ/ (Johnstone, 1965: 235).

Furthermore, another related mostly concurrent variant of the palatal glide is the deletion of the glide altogether (Cantineau, 1966: 91; Johnstone, 1965: 239). Cantineau states that the glide variant could become a ‘kasra’, i.e., short front vowel, which is reported for a number of tribes of the north of the peninsula and lower Euphrates (1966: 91).¹¹² However, the examples cited, namely /iba:h/ ‘foreheads’, /ri:li/ ‘my leg’, and /ħa:ibe:n/ ‘eyebrows’,¹¹³ are likely to show a deletion of the glide rather than a lenition to /i/ (cf. /mi:da:f/ ‘oar’ and /mistaʕi:l/ ‘hastening’ in Johnstone, 1965: 239). This process is also reported for Qatari Arabic (Al-Amadidhi, 1985), and it occurs in the dialect under study as we shall see in § 5.1 in Chapter Five.¹¹⁴

Less common variants

Some less common variants are reported as well, e.g., [z], [d], [ʦ], [ʧ], [dʒ];¹¹⁵ many of these have been triggered by an assimilatory effect (Cantineau, 1960: 92-7 and Zaborski, 2006: 494 for more on the nature and the geographical distribution for the former variants). The diagram in Figure 4.4 summarises the different diachronic developmental stages of Arabic /dʒ/ and the primary sites of the main variants in modern Arabic varieties.

¹¹² For precise distribution, see Cantineau (1966: 91).

¹¹³ The transcription is adapted from the source: SG. *yabha*→*ibāh*.PL; *riy^l*→*rīli*.POSS; and *ħāyib*→*ħāybēn*.DUAL consecutively (ibid). There is a possibility the latter maybe not be the exact rendition of the word due to the fact that the short vowel followed a long /a:/ which may trigger a hiatus constraint.

¹¹⁴ The labio-velar and palatal glides *w and *j are preserved in all modern dialects; however, in some varieties, e.g., Muslim Tunisian dialects, Damascene, and Central Sudanese, these glides are realised as [u] and [i] between consonants (Watson 2002: 19, and the references therein).

¹¹⁵ Also [k], as in /as^snak/</as^snadʒ/ ‘deaf’ in the Bedouin dialect of Ḥawāzīm of Banu Sālim (Il-Hazmi, 1975: 45).

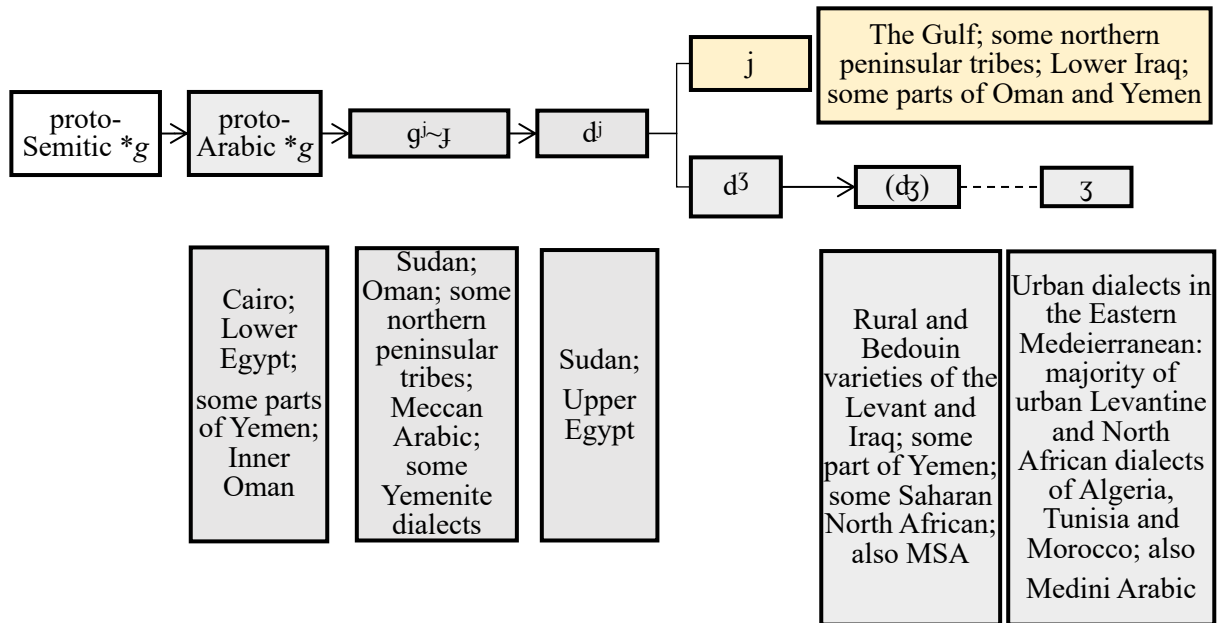


Figure 4.4 A summary of the developmental stages of Arabic /dʒ/ and the contemporary distribution of the main variants in the Arabic-speaking world.

4.3 /dʒ/ in Oman

As we have seen in § 1.5.3 and Table 2.2 in Chapter Two, Holes (1989) uses Old Arabic /q/, /dʒ/, and /k/ in addition to other phonological and structural variables as points of reference to the sedentary-Bedouin contrast in Omani Arabic. The /dʒ/ variants available are the glide [j] which is a Bedouin feature, in contrast to the velar stop [g], the palatal stop [ɟ], and the affricate [dʒ] which are sedentary in distribution.¹¹⁶ Holes (2007: 479) comments on the geographical distribution of /dʒ/ in Oman saying that:

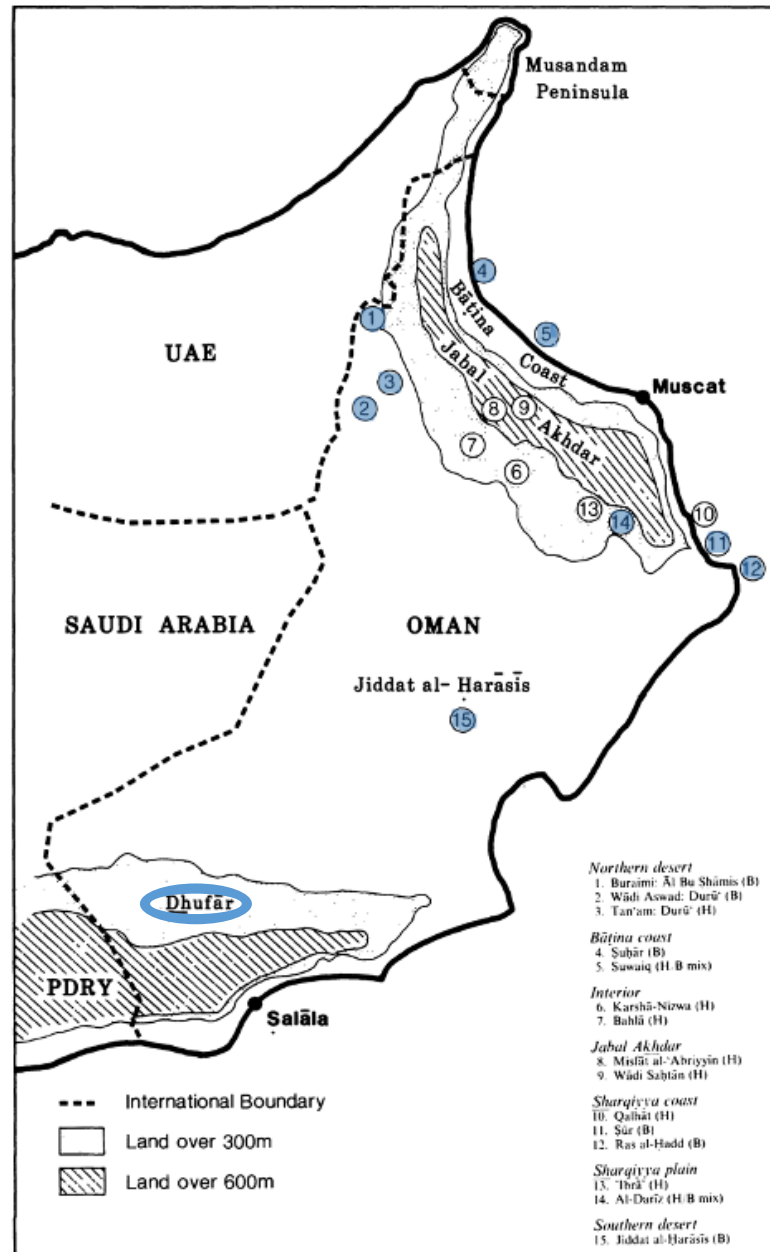
“The most common sedentary Omani Arabic reflex of Old Arabic /j/ is a velar stop /g/. In the Bedouin dialects of the west and southeast, it is a glide /y/. In parts of the Šarqiyya and in some parts of the Empty Quarter, it is an alveolar /j/. In this feature, as in so much else, the Bāṭinah is a mixed area.”¹¹⁷

¹¹⁶ This does not apply to al-Buraymi area.

¹¹⁷ Transcription in the source.

In Map 12, I highlighted the areas with the [j] variant in Holes's (1989) survey of Omani Arabic. According to Holes, this variant occurs in the northern desert in Buraymi and the Ḥaḍari and Bedouin branches of the Durū' tribe, in the southern desert in Jiddat al-Ḥarāsīs, along the coast in the eastern (Šarqiyya) region in Šūr and Rās al-Ḥadd, and along the Bāṭina coast, in Šuḥār and al-Suwaiq. I also highlighted Ḍufār (Rhodokanakis, 1911, as cited in Johnstone 1967: 236). To my knowledge, the glide variant could be, *grosso modo*, generalised for the coastal tribes along the whole of the Bāṭina coast, from Barka to Šuḥār, and then, it seems, all the way up to the Kumzār in Musandam.¹¹⁸

¹¹⁸ Based on my notes on a twenty-seven minute long interview with a bilingual Kumzāri older man conducted as part of the TV programme *Sawālif Šiyyābna* 'Stories of our elderly'; the informant uses the glide reflex for common words like /ja:j-i:n/ 'come.ACT.PTPL-PLM', and /rija:ji:l/ 'men' which varies with the affricate [dʒ] (YouTube link: <https://www.youtube.com/watch?v=myIIDOiXQBM>).



Map 12 The distribution of the [j] variant of /dʒ/ in Oman (reproduced from Holes, 1989: 450).¹¹⁹

Furthermore, another study to mention here is Hamid *et al*'s (2008) remarks on /k/<*k and /g/<*g as part of a study that focuses on the phonology of the dialect of Nizwa, in the heart of the northern Omani hinterland. What is intriguing about this study is that the researchers argue that the velar stop realisations of /dʒ/ and /k/, i.e., [g] and [k], in this S-type

¹¹⁹ The shaded numbers are locations where the normal reflex is the glide.

dialect, are much more restricted in terms of distribution than the palatal allophones, i.e., [j] and [ç], for the same phonemes (ibid: 57), contrary to the available literature that describes the velar [g] as the main reflex of /dʒ/.¹²⁰ According to them, the velar allophones are mainly triggered by back consonants and emphatics; in other words, velars are the exception here (ibid). On the other hand, they grouped the palatal allophones together because they are homorganic, ‘palatal’, and share similar characteristics, ‘plosive’, the only difference being voicing (ibid: 56). In terms of their articulation, these allophones are produced by the middle of the tongue totally closed and touching the roof of the mouth at the palate region; the velum is raised to stop the airflow through the nasal cavity, and just like in the production of other plosives, there is a sudden release of air afterwards as the tongue returns to its original position producing a plosive [ç] in the case of non-vibrating vocal cords, or a plosive [j] if the vocal cords vibrate (ibid). With regard to allophony, these phonemes can be produced with a slight variation: a) towards the back in the direction of the soft palate ‘l-ḥanak l-qasīy’ and in the vicinity of back vowels /a, a:/, they are respectively realised as [k] and [g]; however, b) in the vicinity of front vowels like /æ, æ:/, they are palatal [ç, j] or pre-palatal [ç̣, j̣], e.g.,

- 1) /k/→[ç]: [cæ:nət] ‘be.PFV-3SGF’; [dacæ:ci:n] ‘shop,PLM’; [siccæh] ‘an alleyway.SGF’,
- 2) /k/→[k]: [dukka:n-əh] ‘shop.SGM-3SGM.GEN’; [ʔil-kull] ‘DEF-all’; [ʔakbar] ‘bigger/greater’,
- 3) /dʒ/→[j]: [ʔil-faʔər] ‘DEF-dawn.SGM’; [wa:ʔd-a:t] ‘a lot-PLF’, and
- 4) /dʒ/→[g]: [ħuɡrah] ‘room.SGF’; [mawgu:d-a:t] ‘existent-PLF’. (ibid: 58)¹²¹

¹²⁰ Although Holes mentions the occurrence of a voiced palatal stop /j/ in the sedentary dialects falling within system 5, i.e., Bahla, Ibra, Qalhāt, as well as Nizwa, he does not elaborate on its phonological characteristics nor distribution (1989: 451).

¹²¹ The transcription is modified to be in line with the transcription in this thesis. I also provide the glossing and the English equivalents. The study cited here is written in Arabic with some technical terms provided in English.

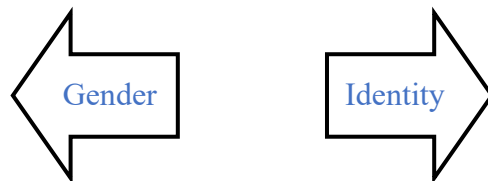
4.4 (dʒ) as a sociolinguistic variable: Trends and issues

The fact that /dʒ/ can be realised in different ways in terms of place and manner of articulation highlights the great instability of this phoneme; this is evident through the descriptive studies reviewed above, and via the sociolinguistic treatments of the (dʒ) variable conducted so far in a number of Arabic-speaking communities. Al-Wer conducted a study on this variable in the speech of 116 Jordanian women from three provincial Jordanian towns, namely Aljoun, Sult, and Karak, based on which, she generally maintains that (dʒ) in the Levant is a sociolinguistic indicator, a low salient variable which does not entertain a similar status to that of (dʒ) in the Gulf region (1991: 174).¹²² Her results fall in line with this argument, since the proportion of the innovative variant [ʒ] is quite low compared to other more salient sociolinguistic variables in her study, supporting the conclusion that the low salience of a linguistic variable in a given speech community can result in low accommodation to non-local variants. Al-Wer's study investigates the role of age and level of education as stratifying social parameters in the context of the study, and explains the variation patterns she observed in light of two main premises:

- 1) There are two major competing norms in the Jordanian society, namely the indigenous Jordanian norm and the Palestinian norm. These norms bear identity connotations, and at the same time, have emerged as gender markers: the indigenous norm → 'a male norm', whereas the Palestinian norm → 'a female' norm.

¹²² Al-Wer (ibid: 175-76) elaborates on the reasons for the difference in the saliency of (dʒ) as a sociolinguistic variable in the Gulf area and in the Levant based on Trudgill's (1986) notion of phonemic contrast as an explanatory factor for salience of linguist variables. She maintains that in addition to the phonetic distinction by which the glide is markedly different from the affricate, the glide ~ affricate distinction is phonemic, unlike in the case of the fricative ~ affricate distinction which is less distinct phonetically and is non-phonemic in nature. There are minimal pairs in Arabic for the first set of variants, i.e., [j] and [dʒ], where the use of the glide would mean the lexical contrast is 'neutralised' when both items of the pair are realised with the glide, and this is probably why the (dʒ) variable is more salient in the Gulf area, but this is not necessarily the case since there are other communities where the variable is not salient despite the phonemic contrast as in the case of the velar reflex /g/ in Cairo.

- 2) These identity-based connotations along with gender influence the linguistic behaviour of indigenous Jordanian women, because they constitute two pressures pulling in opposing directions:



Al-Wer also found that education is a significant predictor; the uneducated speakers are almost categorical in the use of the local variant. They use the incoming fricative variant 1% of the time in contrast to the educated speakers who use it 22% of the time, a result which is explained in the light of education being a measure to the amount of social contacts outside the local community, and thus the amount of exposure to the urban varieties (ibid: 183); the more educated the speaker is, the more [ʒ] is used to index ‘urban life, and urban emancipated and modern women’ in this case, as opposed to the ‘old-fashioned tough’ and masculine connotations of the affricate (ibid: 183-84). These associations are also reflected in the results on age with a considerable difference in the use of [ʒ] between the oldest and the youngest age groups who relate themselves to the local as well as the larger urban community and lifestyle which they aspire to.

Contrary to the sociolinguistic status of (dʒ) in the Jordanian speech communities where Al-Wer conducted her study, the high saliency of the (dʒ) variable in the Gulf and the surrounding peninsular region is driving linguistic change in the communities studied. Al-Essa (2019: 155-7) provides an overview of the literature which deals with the (dʒ) variable in the Arabic-speaking world. She states that, in the Gulf area, results of the studies on (dʒ)

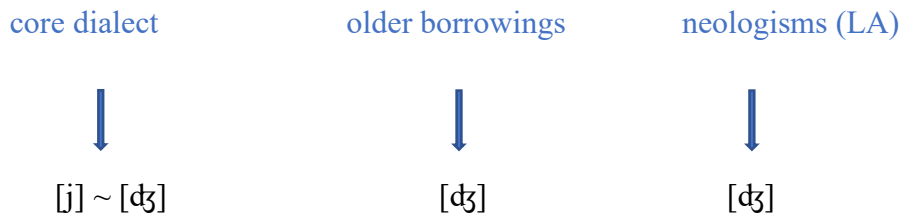
indicate that there is a general shift towards the [j] variant of (dʒ), one of the characterising features of a regional standard variety currently in the process of formation in the area (ibid: 155), giving the variation and change in the use of (dʒ) in Bahrain and Kuwait as case studies.

Holes's study of Bahraini Arabic (1980; 1986; 1987) reveals varying patterns of variation that is community specific. It investigates the [j] ~ [dʒ] variation patterns in the Sunni ('Arab) and Šī'i (Baḥārna) groups' speech, by exploring the role that 'lexical categorisation' as an internal constraint, and literacy, age, the speaker's sex and sectarian affiliation as external predictors play with regard to variation in (dʒ) in an original sample of 108 Bahraini men and women. An important distinction in Holes's study is whether a word belongs to the 'core' dialect ~ vernacular or not (1980: 77); in Bahraini Arabic, the [j] ~ [dʒ] variation always occurs with core dialectal words and never in borrowings (ibid: 77-9); although older (foreign) borrowings, e.g., /dʒalbu:t/ ~ /dʒawa:lbi:t.PL/ 'jollyboat', show a greater deal of internal morphological adaptation to the Bahrain Arabic variety, they never show variation in their realisation of /dʒ/; they are always realised with the affricate. Another essential aspect of this variation is to do with the lexical status of the tokens in terms of whether the word in question has a Literary Arabic analogue, e.g., /dʒa:r/ ~ /ja:r/ 'neighbour.SGM' or not, e.g., a word like /naxxadʒ/ ~ /naxxaj/ 'chickpeas' (ibid: 72).¹²³ Added to the former list are newly introduced words to the dialect through exposure to Literary Arabic which are invariably realised with [dʒ], never with [j] (ibid).¹²⁴ Holes found that in words where the use of (dʒ) is variable, e.g., /wa:ʒid/ ~ /wa:dʒid/ 'much, many' and /ja/ ~ /dʒa/ 'he came', the norm for the Sunni community is to use [j], whereas the Šī'i community tend to use [dʒ] (ibid: 79), although the sectarian division of the two variants is not always clear cut when it comes to

¹²³ I adapted the transcription in the source to be in line with the current study.

¹²⁴ These neologisms are meant to replace older Bahraini Arabic equivalents or older foreign borrowings (Holes, 1980: 78).

the latter group, since Holes found some Šī‘i (illiterate) sub-groups using the Sunni glide variant (ibid: 80).



In terms of the results on the social predictors, Holes found a correlation between sect, sex and literacy: illiterate Sunni speakers are more conservative in their use of the glide, whereas literate Sunni speakers use [dʒ] more frequently, a gradual change which predominantly targets the core dialect’s words with LA analogues; Sunni men generally use it more than women, and literate men use it the most (ibid: 83-85). As for the Šī‘i sample, Holes found that within the [j]-speaking sub-groups, the literate men are the ones leading the change towards [dʒ], and similar lexical constraints on the distribution of the two variants found in the Sunni group are also found to be at work in the latter Šī‘i sub-groups (ibid: 85). In addition, although the use of the affricate within the [dʒ]-speaking Šī‘i sub-groups is predominant, literate informants use the Sunni variant [j] and men use more [j] than women; with literate village men from Al-Muḥarraḡ, the variation is quite considerable. So, the trajectory of change in the first two groups (Sunni and [j]-speaking Šī‘i) is [j]→[dʒ], whereas in the third one ([dʒ]-speaking Šī‘i), the trend is towards the glide, i.e., [dʒ]→[j].

Holes maintains that widespread literacy affects the way the younger members of the Bahraini society speak (ibid: 81). He explains these results with regard to two opposing parameters: ‘local prestige’ vs. ‘linguistic correctness’. The glide variant holds high social prestige which is borrowed from the societal status of its speakers; it is the variant of (dʒ) used by the ruling families and high-ranking Sunni figures in the Gulf area besides being one

of the characterising features of their Bedouin ancestry and heritage (ibid); this prestige is also reinforced by the predominance of this variety in the Bahraini media. The affricate [dʒ], on other hand, borrows its prestige from the MSA, due to its association with the ‘correct’ ‘supra-dialectal’ standard form of Arabic [dʒ]; however, in Bahrain, this variant is generally associated with the Šī‘i speech community which although is indigenous to the area, their speech holds a much lower local prestige (ibid).

In addition, in the light of these findings, Holes highlights two key issues with regard to the linguistic phonological development of the glide as a variant of /dʒ/ in Bahrain and the larger Gulf area:

- The initial historical change *g→[j] seems to have come to an end in these varieties; new words that have been introduced to the dialect afterwards ‘failed to undergo it’.
- The current variation is motivated by extra linguistic factors (including literacy in this case) which could foster an eventual disappearance of the glide variant (1980: 88).

Three decades after Holes did his study, Al-Qouz (2009) conducted another study on (dʒ) as a sociolinguistic variable in Bahrain which aims at investigating whether the ‘intercommunal’ dialect of the city of Manama has an influence on dialect acquisition patterns among 128 six- to seventeen-year-old Sunni and Šī‘i male and female children. The four groups in the sample are stratified according to age, gender, type of school (private, state), social class (upper, middle, working), and community membership (Sunni, Šī‘i). Al-Qouz found that while the Sunni children were categorical in their use of the Sunni [j] variant, the Šī‘i children showed a considerable amount of variation. The Sunni children of the oldest and the youngest age groups are categorical in their use of [j]; Al-Qouz concluded that [j] is a stable variant, which is not surprising since it remains the prestigious and supra-

local variant. The other significant finding is that the Šī'i children's use of [dʒ] gradually decreases as they grow older with some variation established even before they go to school. The fact that the Sunni children are categorical in their use of the glide variant, and the considerably high proportion of this variant in the Šī'i children's speech suggests that the glide variant is still very salient in Bahrain, and that MSA has no influence on the variation since the oldest Shī'i age group (15-17 years old) used the MSA-like form 40% of the time. Al-Qouz also found that the upper- and middle-class children in general disfavour the use of [dʒ], unlike lower class children who do not seem to have acquired [j] since they exhibit a high use of the traditional [dʒ] throughout; the youngest age group favours the glide, unlike the rest, and males slightly favour it more than females. The innovators in this case are primarily private school children who are the first acquirers of the [j] variant, followed by upper class children in state schools.

Furthermore, not far from Bahrain, Taqi (2010) found comparative results with regard to the variation between the 'prestigious' Najdi [j] and 'less prestigious' Ajami [dʒ] in Kuwait. Her sample is stratified according to age, gender, and ethnicity of the informants; it consists of 48 men and women of Najdi and Ajami origin (all educated) across three generations. Her study highlights the effect of social network, dialect contact and correlation between identity and levelling, along with prestige on the variation patterns observed. She maintains that the Najdi variety is more stable across generations; the Ajami dialect seems to be moving in the direction of the former, with the younger Ajami speakers resembling the Najdis the most. With regard to gender, she found that females in all groups generally showed less [dʒ] than their male counterparts, although within the Najdi group, there is not a significant difference across gender or age groups. With regard to age, Taqi found that older speakers within both groups use more [dʒ] than the rest, and younger speakers in both groups behave similarly, which suggests that [j] is gradually becoming the koiné form of Kuwaiti Arabic (and Kuwaiti

identity).¹²⁵ In addition to identity, ethnicity is another significant explanatory factor, since younger Ajami speakers affiliate themselves with Kuwaiti Arabic not Farsi (Persian descent), and they socialise more with their friends from different ethnic groups than with their family; they are not bilingual in Kuwaiti and Farsi like their grandparents.¹²⁶

The picture with regard to the status of the different variants of (dʒ) outside the Gulf is different; the traditional [j] variant is undergoing variation and change in favour of other supralocal and/or prestigious forms in places close to the Gulf itself. One case study is Alshawi (2019) on the Mishleb tribe in Qal‘at Sikr where a change in progress from [j], the stereotypical local variant, to the mid-Mesopotamian Baghdadi-like [dʒ] is taking place in Southern Iraq. The change is quite advanced and is constrained by a number of internal (see § 4.5.1) and external predictors (age, gender, and level of contact). Results from a multivariate analysis of a dataset based on interviews with 53 informants show that the overall use of [dʒ] is quite high, and that the social constraints are quite salient, with the ‘level of contact’ selected as the most statistically significant predictor as high contact speakers using the innovative variant almost as twice as the low contact ones (90% and 50% respectively) (ibid: 125). Also, the younger age group (20-34-year-old) is found to use the innovative affricate the most (77%), followed by the middle (70%; 37–56-year-old) and older groups (58%; 60+ years old) respectively. Men use the innovative variant 74% whereas women use it 62% (ibid). Alshawi explains the variation and change patterns observed mainly in the light of

¹²⁵ Johnstone (1965) also maintains that the reversal of the historical change *g→[j] in Kuwaiti Arabic is influenced by the process of koineization the dialect is undergoing. Generally speaking, the items that still hold on to the glide variant are common local dialectal words that do not have an equivalent in the koiné, although he does not explicitly mention what this koiné reflexes are (ibid: 238).

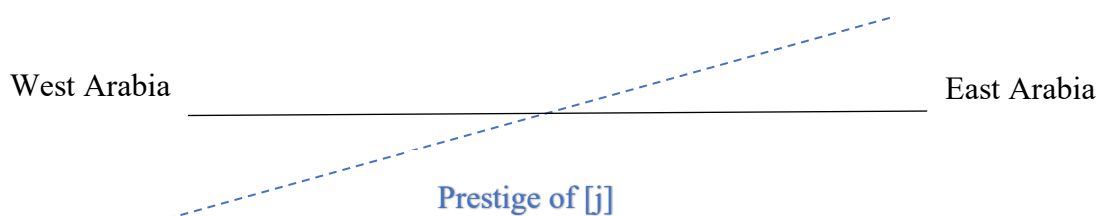
¹²⁶ The internal constraints explored in Taqi’s study are associated with the lexical status of the token in terms of ‘standardness’, ‘recency’, and ‘frequency of usage’. Similar analysis is conducted by Al-Amadhidhi (1985) and Al-Muhannadi (1991) on the (communal) patterns of variation in (dʒ) in Qatar; the two main variants in the area are the glide and the affricate. These studies also include the lexical status/class of the tokens as a main internal constraint governing the rules of the variation in (dʒ), and they explain the findings on the internal constraints in the light of a standardisation vs. colloquialisation approach which is primarily based on the assumption that the affricate variant of (dʒ) in this case directly borrows its prestige from the Standard Arabic variety.

dialect contact and levelling that is currently very relevant to the sociolinguistic situation in Iraq and which is facilitated through migration, socio-economical (urbanisation), and socio-political patterns (ibid: 132-35). Alshawi also highlights the notion of ‘indexicality’ and how the speakers’ choice of certain variants can index certain social meanings (ibid: 135-137).

Furthermore, one of the main parts of Alaodini’s (2019) work focuses on variation and change in [dʒ] in the dialect of the Dōsari families in Dammam, Saudi Arabia. Her study is another interesting case study of how population movements and migration patterns can trigger variation and change. In the course of its relatively recent history, the Dōsari speech community witnessed major population movements from Wādi Ad-Dawāsir in Najd to Bahrain and then back to Saudi Arabia where they settled in Dammam, and the phoneme /dʒ/ witnessed different changes both linguistic and sociolinguistic accordingly. This group of Dawāsir presumably used the affricate variant which is the end product of the historical changes that affected proto-Semitic *g before they immigrated to Bahrain, where they shifted to using [j], and now that they have settled back in Dammam, they are replacing this [j] by the prestigious Saudi counterpart [dʒ] (ibid: 115-16). In addition to a set of internal constraints (see § 4.5.1), Alaodini’s study investigates the effect of social network, age and gender as social predictors of linguistic variation and change in this speech community. Results from a multivariate analysis of a dataset collected from 39 informants generally show that there is a change in progress towards the more prestigious [dʒ] variant, which is characteristic of the supralocal dialect of the capital, Riyadh. Alaodini’s results show that the change is quite advanced ([dʒ]=76.3%), and it is almost complete for the younger speakers ([dʒ]=96%; 20-39 years old) (ibid: 112). Men ([dʒ]=90%) are found to use [dʒ] more than women ([dʒ]=61%), and informants with loose social networks (93%) use [dʒ] much more than those with tight social networks (57%) (ibid). The older women (60+ years old) in this study are the most linguistically conservative speakers (ibid: 125). The leaders of change in

this speech community are men across the three age groups and younger women. Alaodini interprets these results in the light of the different social meanings the two variants are indexical of, e.g., with regard to gender and identity connotations, akin to what Al-Wer (1991) has found for Jordan: the glide here is associated with women's speech especially within the older age group, and the affricate is associated with national identity (ibid: 126-28).

Farther from the Gulf area and further into the inland of the Arabian Peninsula, two primary sociolinguistic studies that discuss the change towards the prestigious supralocal or regional (dʒ) forms were conducted in Saudi Arabia. An earlier study by Al-Shehri (1993) investigates the patterns of variation and change related to the (dʒ) variable as used by the immigrant rural (or an urbanised tribal) Shehri speech community dwelling in the urban Jeddah area. His study investigates the role of social urbanisation, through age, education, length of stay, and the speaker's sex, on the linguistic 'convergence' or 'divergence' from the urban speech patterns in a sample of 84 speakers. Al-Shehri remarks that the prestige the glide entertains as a variant of (dʒ) in eastern Arabia is confronted/compromised by the prestige of the affricate variant in west Arabia (ibid: 79):



The affricate [dʒ] being the highly prestigious ‘urban’ and supralocal form is replacing [j], the highly marked and stigmatised traditional ‘rural’/‘tribal’ feature, with the youngest age group (15-30 years old) using the affricate 97% of the time. On the other side of the age grouping, the speakers (60+ years old) use [j] considerably more than the rest of the sample (28%), a result which Al-Sheri explains with regard to the lack of ‘sociolinguistic awareness’ of the older speakers which (according to the researcher) is normally obtained through education and mobility (ibid: 81); educated speakers are found to use the rural variants the least, i.e., education is responsible for abandoning the traditional feature (ibid: 89). An interesting result is that within the groups of educated informants only, there is more gradience in the proportions of the use of the affricate, unlike in other groups that have both educated and uneducated speakers, probably indicating more uniformity or consistency in the use of the urban variant (ibid: 90). Another noteworthy observation is that since the gap between men and women is small, Al-Shehri argues that the speaker’s sex is no longer relevant to the variation at hand, probably due to the high markedness of the traditional variant (ibid: 93), and that new rural migrants who spent less time in the area than older ones also show more use of the innovative feature since they also happen to be educated, unlike the older migrants who although have spent more time in the urban area, they have not experienced the same kind of contact, mobility and access to other varieties that the younger speakers have enjoyed via education.

The effect on urban centres of influence in Saudi Arabia is also the main focus of Hussain’s (2017) investigation of variation and change in Medini Arabic. In Hussain’s study, the urban and Bedouin Medini communities revealed a general trend toward de-affricating the local variant [dʒ], resulting in a change in progress towards the fricative [ʒ]. Stress, along with following and preceding segments were found to be statistically significant within the urban community, in addition to age which is the most significant predictor in her dataset; the

youngest age group used the fricative the most (mean of [ʒ]=0.626) (ibid: 201). Gender on the other hand is not selected as statistically significant in predicting the variation in the urban community (ibid). When both speech communities are considered for the multivariate analysis, gender, community, and the variable's position are selected as significant constraints, along with the former ones; age selected as the most significant predictor with both young and adult age groups favouring the use of the fricative and the middle-aged and older speakers disavouring it. The urban community uses the innovative variant more than the Bedouin community and women in both communities use it more than men. Hussein interprets the results in the light of the urban influence of the city of Jeddah due to the regular contact of the urban Medini speakers with the Jeddah community, and then this influence has extended to the Bedouin community due to the Bedouin-urban contact situation within Medina itself, where young Bedouin women are the ones advancing the change, probably due to their aspiration to the urban lifestyle and prestige.

So, the sociolinguistic literature on (dʒ) reveals that different variants carry different prestige or status in different parts of the Arabic-speaking world: we find [j] as the prestigious variant in the Gulf, [dʒ] as the most favourable in west Arabia, [ʒ] as the prestigious variant in many urban centres along the Mediterranean (Levant and North Africa), but we also find [g] as the prestigious form in Egypt. However, this notion of prestige as an explanatory factor is not clear-cut since: 1) the prestige of the different variants is borrowed predominantly from the social status a dialectal variety maintains within a geographical area, and 2) this variety may be susceptible to multiple prestiges rendering variation sensitive to opposing pressures that may pull in opposing directions; see e.g., Al-Wer (1991) on gender and identity/ speaker affiliation as opposing powers for variation in the Jordanian context, and Holes (1980: 8) on social prestige and linguistic correctness as forces pulling in opposing directions in the Bahraini context. We find that, within Saudi Arabia, for example, on the one

hand, there is a trend of change from [j] towards the supra-local Saudi form [dʒ] (Alaodini, 2019), but on the other, there is a change from a traditional form, which happens to be identical to the form found in the supra-local variety (viz. [dʒ]), towards the Jeddah form [ʒ] in Medini Arabic (Hussein, 2017). We also see that in the Gulf, generally speaking, the change is towards [j], the prestigious Gulf form. That is to say that certain regions have certain contrasting and sometimes overlapping forces that govern the mechanism of change in this and other sociolinguistic variables in the Arabic speaking world: koineisation, urbanisation, and supra-local prestige, to name a few (but could also be due to dialect contact and/or linguistic accommodation). It is noteworthy that due to the diglossic situation in the Arabic-speaking world, the prestige that is driving most variation in the region is not borrowed from MSA. Arabic sociolinguists have re-emphasised over the years that MSA or Literary Arabic variants, though they may contribute to the existing variation, are not the target of change (mainly Al-Wer 1991, and 1999; Ibrahim, 1986; Abdel-Jawad, 1986).

4.5 (dʒ) as a sociolinguistic variable: Trajectories of change

As we have seen above, some sound changes may be unidirectional, yet others are bidirectional, particularly where different sub-groups of a given speech community move in different directions with regard to a specific sound change. Linguistically speaking, while some trajectories of variation the (dʒ) variable has undergone or is undergoing involve lenition changes, as in the de-affrication of [dʒ]→[ʒ] (Al-Wer, 1991; Hussain, 2017), and the [dʒ]→[j] change (e.g., Holes, 1980; Al-Qouz, 2009; Taqi, 2010), others involve fortition, as in the fronting and affrication of [j]→[dʒ] (Alaodini, 2019; Alshawi, 2019; Al-Shehri, 1993).

In addition to focusing on investigating the external forces on variation and change and on interpreting the existing variation in the light of these predictors, recent variationist

studies (e.g., Hussain, 2017; Alaodini, 2019; Alshawi, 2019) also explore relevant internal ‘phonological’ and ‘morpho-phonological’ linguistic constraints and interpret the variation and change observed in terms of these constraints and in relation to the wider linguistic theory. The situation is similar to that of the external forces on variation and change; different internal constraints on related changes in the different dialects govern the variation in (dʒ), which is also true of diachronic phonological changes cross-linguistically. But before reviewing the findings on the internal constraints on the variation in (dʒ), it is important to discuss the main diachronic and synchronic phonological processes involving /dʒ/ in order to define some key concepts at work. Pertinent to this study are the key concepts of lenition and fortition which are the main focus of the following section.

4.5.1 Lenition and fortition: Internal constraints

Whatever the route that the traditional [j] reflex took, it is well-established by now that it has resulted from a lenition process of proto-Semitic *g, as previously described in § 4.1. Many phonological works have been devoted to the study of synchronic and diachronic phonological processes that are included under the larger umbrella of ‘lenition’, and different theoretical frameworks have developed varying approaches in their treatment of lenition;¹²⁷ however, there seems to be a general consensus on the basic definition of the term. Lenition is viewed as a weakening process, during which a ‘strong’ sound becomes ‘weaker’.¹²⁸ Another process pertinent to the discussion of lenition and a key concept in the current study is ‘fortition’ which is commonly defined as a strengthening process through which a ‘weak’ segment becomes ‘stronger’. Fortition is not as well-studied as lenition, due to a general

¹²⁷ For a review of the key terms, arguments and theoretical approaches in the treatment of lenition and fortition see De Carvalho, Scheer and Ségéral (2008).

¹²⁸ See e.g., Honeybone (2008) and Bauer (2008) for a discussion of the concept of ‘lenition’ and a review of its development in the history of phonology.

tendency to view it as the counterpart or mirror image of lenition, and also due to it being less common than lenition cross-linguistically (e.g., Honeybone, 2008: 10; *cf.* Bauer, 2008).¹²⁹

In synchronic phonology, lenition (and fortition) is commonly linked to consonantal strength, in terms of the differences in ‘sonority’,¹³⁰ ‘stricture’ or ‘constriction’, and/or ‘openness’¹³¹ involved in the production of consonants. A quick survey of the relevant works on lenition shows that languages behave differently in applying lenition rules (and by default fortition rules); they also differ in the constraints and conditioning environments that are relevant to these rules. For instance, syllabic position is a paramount factor in lenition; it depends on the Sonority Hierarchy, a key concept here, which is often used as a measure for consonantal (and vowel) strength or weakness. Lenition is basically understood as moving up the Sonority Hierarchy; on the other hand, fortition is commonly understood as moving down this hierarchy. The Sonority Hierarchy is mainly based on segment voicing and manner of articulation in that voiceless consonants are stronger than voiced ones and stop segments are stronger than others, as shown in the following scales:

(1a) A simplified scale:

Plosives>fricatives>approximants>zero

Aspirated>plain voiceless>voiced

(Spencer, 1996: 62)

¹²⁹ Honeybone argues that “[...] cases of real fortition are vanishingly rare, and it is by no means obvious that they really are the literal ‘opposite’ of lenition.” (2008: 10), *cf.* Bybee and Easterday (2019: 289).

¹³⁰ E.g., Crowley (1997: 36); Szigetvári (2008: 94); Lass (1984: 177-79).

¹³¹ Lass (*ibid.*).

(1b) A more detailed scale which includes voicing distinctions:

voiceless plosives>voiced plosives>voiceless fricatives>voiced

fricatives>nasals>laterals> rhotics>high vowels>mid vowels>Low vowels

(Szigetvári, 2008: 96)¹³²

The Sonority Hierarchy is relevant to syllabic position because it is based on a universal Sonority Sequencing Principle-SSP (e.g., Szigetvári, 2008: 94).¹³³ According to this principle, within a given syllable, the sonority in the onset position rises towards the nucleus of the syllable, then falls down on the right edge of the nucleus, e.g., in the case of the syllable's coda (ibid), so it would be reasonable to expect stronger segments in initial syllabic positions and weaker segments in final syllabic positions. In fact, there seems to be a general consensus that coda and intervocalic positions are leniting contexts, and post-vocalic and pre-pause positions are considered as weak contexts that trigger lenition; on the other hand, post-pause, pre-vocalic, and post-coda positions are considered as strong positions that trigger fortition (ibid: 113-114, and the references therein). In addition, a post-obstruent position is considered a stronger context than a post-sonorant one (ibid),¹³⁴ as well as whether the segment involved is a part of a consonant cluster.¹³⁵ To add to these constraints, stress is relevant to the discussion of positional strength and lenition since the “function of stress is to give strength to the vowel (or the syllabic consonant) on which it falls and to the consonantal environment” (Ségéral and Scheer, 2008b: 504). Bybee and Easterday (2019: 282) also state

¹³² The types of lenition proposed in Szigetvári (2008: 110) include: a) the loss of place features, b) the loss of laryngeal features, and c) sonorization, i.e., the increase in sonority. Different phonological processes are involved in each of the three types, e.g., spirantisation, gliding and rhoticism.

¹³³ This principle is considered as a ‘universal phonotactic constraint’, but this does not necessarily entail that it is always obeyed (ibid).

¹³⁴ Also discussed in Ségéral and Scheer (2008a), who state that the strength of the post-coda position depends on whether the coda in this position is a sonorant or an obstruent, since post-sonorant positions can be weak or strong, yet the post-obstruent positions are always strong (ibid).

¹³⁵ Szigetvári, citing Escure's (1977) ‘environmental hierarchy for lenition’, states that being part of a consonant cluster or not is one of the criteria necessary for a hierarchy of leniting positions (2008: 111).

that the segments involved play a role in a process like fortition where adjacent consonants increase the constriction of the segments in question.

4.5.1.1 *Glide* → *stop fortition*

Based on the Sonority Hierarchy, and for the purpose of the current study, the traditional variant [j], which is an approximant, is a ‘lenis’ sound when compared to the incoming stop variant [gʲ], which is a ‘fortis’ sound. This means that the current trajectory of variation in the speech community under study can be linguistically viewed as a case of fortition, a probable reversal of a completed historical phonological lenition process.¹³⁶ We have already seen that some Arabic dialects which have retained the traditional glide reflex are synchronically replacing it by an affricate ‘fortis’ variant of /dʒ/. A recent survey of fortition (here, viewed as an increase in the degree of oral constriction) found that, cross-linguistically, glide (primarily palatal) strengthening is the most common type of the fortition processes (Bybee and Easterday, 2019: 279), and that the fortition of palatal glides results in consonants “which are also produced in the general palatal(ized) region” (ibid: 297-80).¹³⁷ The most common favourable environment(s) for palatal strengthening according to this survey is a preceding high and/or front vowel typically accompanied by the glide itself being in a syllable initial position or a ‘domain initial’ position, i.e., in a word-initial or a syllable onset position, regardless of the nature of the following vowel, although in much fewer cases, the fortition happens in relatively weaker positions (ibid: 281-82). It is noteworthy that the glide fortition processes in Bybee and Easterday’s studies are found to mainly involve an increase in the oral constriction in terms of the articulation of the glide (ibid: 279, 288).

¹³⁶ This is not to say that the choice between [j] and [gʲ] is entirely purely phonological. The occurrence of one variant or another can sometimes be sensitive to or conditioned by the lexical status of the word (see § 5.1 and § 5.2 for more on this point).

¹³⁷ According to Bybee and Easterday’s findings, fortition “tends to be less about increases in magnitude in general and more about the particular places of articulation involved: glide strengthening, fricative strengthening, and buccalization all share palatal and labial articulations in common.” (ibid: 287).

In the context of the Arabian Peninsula and unlike in the case of the historical internal change of the **q* and **k* part of the Semitic triad which, in some Arabic dialects, ended in allophony with fronted and affricated allophones mainly occurring in front vowel environments, in eastern Arabia where the end product of the change of Semitic **g* is the glide, the change is not conditioned by the phonetic context, since it freely occurs in all contexts (Johnstone, 1965).¹³⁸ However, the occurrence of the glide is reported to be lexically conditioned in that it occurs with core and widely used common dialectal items; in certain classicisms, borrowings, neologisms or certain proper nouns the affricate is used, although this division is not always neat in that some common or ‘core’ dialectal words are never realised with the glide (Johnstone, 1965: 241;¹³⁹ see Holes 1980: 78 in § 4.4).

Furthermore, it is noteworthy that the treatment of (dʒ) as a sociolinguistic variable in the communities where the glide is the incoming variant rather than the traditional one, e.g., in the dialect of certain Bahārna of Bahrain in Holes (1980) and Al-Qouz (2009), and the ‘Ajamis of Kuwait in Taqi (2010), largely emphasises the lexical status of the tokens analysed, which means that it is not possible yet to know which internal phonological constraints are at work in these cases.¹⁴⁰ But we can keep in mind the historical trajectory of the [j] reflex of /dʒ/, which was argued to be a product of historical **g*→*g^j* palatalisation and *g^j*→*j* or *d^j*→*j* lenition process, and we know that cross-linguistically, palatalisation is triggered by following front vocoids: vowels and glides or in the context of front vowels, palatal glides, palatal or palatalised consonants (e.g., Zeroual, 2006: 525; Bhat, 1978: 49). However, we do have statistical evidence for the phonological environments that condition

¹³⁸ Cf. Zeroual (2006: 525): the glide reflex of /dʒ/ in the modern Arabic dialects of Kuwait and Basra “appears in general in the context of a low vowel”.

¹³⁹ Johnstone maintains that the synchronic change from [j] to [dʒ] is reversible at will but does not usually happen to common words (ibid: 241).

¹⁴⁰ The exception in these works is Holes’s discussion of internal vocalisation associated with the lexical status of the tokens in question (1987: 101-105).

the variation in (dʒ) where the glide is the traditional variant which is being levelled out in favour of the affricate [dʒ] (Alaodini, 2019; Alshawi, 2019). The two studies offer statistical results with regard to the preceding and following segments that are shown to favour the incoming ‘fortis’ variant. Alaodini found that preceding high vowels, both front and back, and following consonants and high front vowels favour the incoming variant, in contrast to preceding low vowels and consonants, and following high back and low front vowels which disfavour it (ibid: 112-13). On the other hand, Alshawi found that front vowels disfavour the incoming affricate, whereas front consonants and back sounds¹⁴¹ favour it both in the preceding and following contexts. With regard to the other internal constraints, Alaodini found that monosyllabic words favour the affricate whereas di- and polysyllabic words disfavour it. These results are contrary to Alshawi’s findings with regard to mono and poly- (and tri-) syllabic words since he found that the former disfavour the affricate and the latter favour it. The two studies also looked at stress in that whether the dependent variant occurs in a stressed or unstressed syllable, but in both studies, stress was not chosen as a significant predictor. In addition, Alshawi found gemination to be significant with singletons favouring the incoming affricate but not the geminates.¹⁴² Alshawi also looked at the position of the dependent variant in the syllable (onset or coda) and the structure of the syllable itself in terms of whether it is a light, heavy, or superheavy syllable. Syllabic position was not chosen as a significant predictor, whereas syllabic structure was found to be statically significant with light syllables favouring the incoming affricate as opposed to heavy and superheavy syllables which were found to disfavour it. It is noteworthy that in Alshawi’s study the social constraints were found to be more significant in predicting the variation than the internal

¹⁴¹ The back sounds here include back vowels and the bilabial glide /w/.

¹⁴² Alaodini initially coded for gemination as an internal predictor but eventually excluded it due to the unbalanced distribution of tokens in this group; the total geminate tokens were considerably lower than the singletons (personal communication).

constraints when all of the constraints (internal and external) were considered for the multivariate analysis ($R^2=0.354$). The following sound was selected as the most significant internal predictor, followed by the number of syllables, the preceding sound, syllabic structure and gemination respectively. The internal constraints selected as the most significant in Alaodini's study were the number of syllables followed by the following sound and the preceding sound respectively ($R^2=0.549$).

One very important point to raise here concerns the nature of the incoming variant in the dialect under study. The literature reviewed in § 4.1 does not present a precise description of the nature of the fronted Old Arabic palatal stop; it is sometimes described as a pre-palatal or a dorsal palatalised velar stop [gʲ] (a palatalised dorso-palatal plosive: Cantineau in Zeroual (2006: 525)), or simply as a plain palatal stop [j].¹⁴³ Nonetheless, whatever the precise description might be, Old Arabic /dʒ/ seems to be viewed as a stop that shares or approximates the place of articulation with the palatal glide. In Oman, the palatal stop [j] is reported as a reflex of /dʒ/ in the sedentary interior (see § 4.3). With regard to the context of this study, Holes (personal communication) elaborates on /dʒ/ in the speech of two of his informants on which the description of the study area in his (1989) survey is based. The older speaker on whose speech the data from al-Suwaiq is based varied between the glide [j] and another 'palatal sound', but always used [j] in common words.¹⁴⁴ The other is a 45-year-old speaker from Şaḥam, a town located to the north-west of al-Suwaiq. This speaker used a voiced palatal sound, e.g., *a-gʲi* '1SG-come.IPFV' and *sagʲgʲal-t* 'enrol.PFV-1SG' along with the glide [j] much more than the former speaker, and showed a voiceless palatal stop for /k/<*k in front vowel environment, e.g., *lācin* 'but' and *sācin* 'live.ACT.PTCP.SGM'.¹⁴⁵ In

¹⁴³ See Watson (2002: 15).

¹⁴⁴ This speaker has a mixture of B- and S-type dialectal features in that he shows B-type phonology and plenty of S-type morphological features, which could be explained in the light of the speaker's life experience and the fact that he spent thirteen years in the Gulf (Holes, 1989: 452).

¹⁴⁵ The examples are provided by Holes; I adapted his transcription to be in line with this research.

addition, the latter informant's speech matches the description of the velar stops of my Bedouin informants. This point is related to a noteworthy observation on the fronting of the velar stops in the dialect under study. Both velar stops /k/<*k and /g/<*q in this dialect show a palatal realisation which seems to be favoured in a front vowel environment although I have not thoroughly analysed it; this is a similar environment to the one that triggers fronting and affrication in other peninsular B-type varieties. The palatal realisation of the velar stops is not exclusive to the Bedouin dialects within the study area. It seems to be typical of the speech of the people (whether B- or S-type)¹⁴⁶ on the coast and the speech of the Bedouins living farther inland of Suwaiq. In addition, it is noteworthy that none of the former variants have been analysed acoustically. What is essential to this study though is that, impressionistically speaking, the incoming variant in this dialect is a 'fortis' voiced dorsal-palatal which sounds more like a palatalised velar stop represented here by the symbol [gʲ].

To summarise, this chapter provided a brief linguistic background to the current analysis of the (dʒ) variable, focusing on the literature on lenition, fortition, and the findings of recent variationist studies on the internal constraints that govern the variation and change. We have seen from the linguistic literature reviewed here that glide fortition favours strong segments, strong positions, stressed syllables, and certain vocalic contexts with regard to height and tongue position. The variationist literature (Alaodini, 2019; Alshawi, 2019) adds more to the picture; number of syllables, syllabic weight, and gemination are also examined

¹⁴⁶ I have not analysed the S-type dialect of al-Suwaiq itself; I am a speaker of an S-type, but my family originates from a mountainous village in the town of Ṣaḥam, the same town Holes's informant is from. They came to al-Suwaiq a long time ago and settled along the coast of al-Suwaiq near the main market (see § 3.3). The norm for the S-type speakers in this area is a palatal sound for both velar stops *g and *k; the plain velar [g] does not occur, neither as a variant of the etymon *g nor as a variant of *q which is traditionally realised as [q] for this speech community. However, in the mountainous villages in al-Suwaiq, the situation is similar to that of the interior dialects where allophony seems to be the norm as well, based on preliminary data I collected from a young Jahwari informant from al-Suwaiq. Also, the fronting of Arabic /k/ is previously reported for the Āl Wahība tribe, as explained in § 2.2.1.1 in Chapter Two. There are some families of Āl Wahība who live in al-Suwaiq. It is common knowledge that they have settled in the area a long time ago, but I don't have further details as to when this happened neither roughly nor exactly.

as internal constraints relevant to the fortition cases found in the speech communities studied. This background provides the primary basis for the linguistic coding and modelling stages in this study, as reported in Chapter Five.

Chapter Five: Data analysis and findings for the (dʒ) variable

This chapter explains the data preparation and analysis process of the (dʒ) variable. It presents the results of the quantitative analysis of the variable and reflects on the findings in relation to the different internal and external predictors selected for analysis in light of the previous literature and in relation to the speech community under study.

5.1 The dependent variable

In the dialect under study, the (dʒ) variable has two main variants: the glide [j] and a palatalised stop [gʲ]. There is some intra-speaker variation in the interview data where other variants occur: the affricate [dʒ] and to a much lesser extent, the velar stop [g]. In addition, the glide is deleted in some lexical items; this variant is coded as Ø. Below are some examples for each variant:

[j] ~ [gʲ]:

/ji:t/ ~ /gʲi:t/ ‘I came’

/l-jibal/ ~ /l-gʲabal/ ‘the mountain’

/jimal/ ~ /gʲamal/ ‘camel’

/ja:rni/ ~ /gʲa:rni/ ‘our neighbour’

/dja:j/ ~ /dgʲa:gʲ/ ‘chicken’

/ma:ju:d/ ~ /mawgʲu:d/ ‘existent’

As in the data from Bahrain (e.g., Holes 1987: 99) and Dammam (Alaodini, 2019: 104), this variation also occurs in proper names,¹⁴⁷ e.g., /l-jimʕi/ ~ /l-gʕimʕi/ ‘Friday’, and in personal names, e.g.,

/ma:jdi/ ~ /ma:gʕda/ ‘a personal name.F’

/jimʕa/ ~ /gʕimʕa/ ‘a personal name.M’

/jazmi/ ~ /gʕazmi/ ‘a local toponym’

/baʕʔha ja:bir/ ~ /batʕʔha gʕa:bir/ ‘a local toponym’

Arabic loan words:

/gʕe:f/ ‘army’

/aɡʕa:za/ ‘holiday’

/lahɡʕa/ ‘dialect’

/aɡʕa:r/ ‘rent’

/zo:ɡʕi/ ‘husband’

Ø:

Word-initially:

¹⁴⁷ The criteria for deciding on the lexical status of core vs. non-core dialectal items follows a similar logic to Holes’s (1980) analysis of his Bahraini Arabic data. For the purpose of this analysis, words that have traditional, non-standard-like dialectal equivalents are treated as either non-core dialectal items or as borrowings “into the dialect as MSA neologisms” (see Holes, 1980: 439). So, e.g., the word /zo:ɡʕi/ ‘husband’ above has a traditional dialectal equivalent /rajil/ (cf. /rajja:l/ ‘the dialectal equivalent for *man*’), the word /lahɡʕa/ ‘dialect’ has a dialectal equivalent /harji/, and the word /gʕe:f/ ‘army ~ soldiers’ also has a dialectal equivalent /ʕaskar/. Such lexical items are never realised with the traditional variant [j], probably because they had been introduced to the dialect through contact at a time when the historical change *j > [j] was no longer operative (see § 5.2). It is noteworthy that such categorical tokens are not included in the quantitative analysis of this variable, since the variation here is not phonologically conditioned, rather it is sensitive to the lexical status of the words (see § 5.3).

/de:da/ (< *jde:da) ‘grandmother.DIM.SGF’¹⁴⁸

Word medially:¹⁴⁹

/ma:-ʕa:b-ah/ (< *ma:-ʕaØab-ah < ma:-ʕajab-ah) ‘NEG-like.PFV.3SGM-3SGM; literally: he did not like it’;

/ri:l/ (< *riØil < rijil) ‘leg’

/na:ri:l/ (< *na:rji:l) ‘coconut’

Word-finally:

/xa:ri/ (< *xa:rij) ‘outside’

/l-biri/ (< *l-birij) ‘a local toponym’

/ji-nðʕa/ (< ji-nðʕaj*) ‘3-get done.IPFV.SGM’

[dʒ]:

/mawdʒu:d/ ‘existent’

/d-didʒa:dʒ/ ‘the chicken’

/li-ħdʒu:r/ ‘the villages at the rocky mountain foothills’

/l-dʒi:ra:n/ ‘the neighbours’

/l-dʒiba:l/ ‘the mountains’

[g]:

/mawgu:d/ ‘existent’

/raɣaʕ-t/ ‘come back.PFV-1SG’

¹⁴⁸ This is the only word I found. It literally means little grandmother and is probably meant to be more personal or endearing form of the noun /jadda/ ‘grandmother’.

¹⁴⁹ When [j] is deleted intervocalically, the preceding and following vowels seem to merge into one long vowel.

/gargi:r/ ‘rocket leaves’

/gi:ra:nne/ ‘our neighbours’

5.2 Notes on the lexical distribution of the variants

In line with the literature on the (dʒ) variable in dialects in which the glide is the traditional variant, both [j] and Ø occur in words which are considered as *core* dialectal or common dialectal words (see e.g., Holes, 1980; 1986; 1987). While they can be used for proper names (personal names and toponyms), traditional variants are not used in borrowed (foreign) lexical items; in such items [gʲ] is generally used. The latter variant is in variation with the glide in many of the words that are traditionally realised with [j]. It is also used in Arabic loan words and in cases of code-switching to the standard variety. This observation is quite intriguing since one would expect the MSA-like variant [dʒ] to be characteristic of the latter group. The older informants almost never produced the standard-like affricate variant even when they were using specialised words. For my younger informants, the default variant for formal words related to work or school is also [gʲ], not the standard-like variant. In addition, the affricate variant in the data occurs in what could be classified as core dialectal items, along with formal borrowings from MSA, seemingly in the same domains where the [j] ~ [gʲ] variation takes place (more in § 5.5).

Another key point is that the choice of the variation in (dʒ) in this dialect is not always a matter of a simple choice of one segment or another; the picture is more complicated. With some words, choosing one or the other variant is sensitive to the lexical status of the word and it means the structure of the word itself can change, phonologically speaking, as previously discussed in the literature (see § 4.1 and 4.5 in Chapter Four). The [j] ~ [gʲ] variation can be: 1) accompanied by a change in the quality of the following vowel, e.g., /jilas-it/ ~ /gʲalas-it/ ‘sit.PFV-1SG’, and /jibal/ ~ /gʲabal/ ‘mountain.SGM’, or 2) accompanied

by re-syllabification of the first part of the word, e.g., /ʃja.ra/ ~ /ʃaɟj.ra/ ‘tree.SGF’. There are of course other similar lexical items, where the variation is in the use of one phoneme (dependent variant or another) that does not entail unrelated accompanying changes, e.g., in the following vowel quality or the word’s syllable structure; so, a word like /qʌimʕi/ can be used for a man’s proper name (a core dialectal word), which in this case varies with /jimʕi/,¹⁵⁰ as well as in the difference between /jimi:ʕ/ ~ /qʌimi:ʕ/ ‘all’; so, in this case, the change can be argued to have sociolinguistic (intra- and extralinguistic) motivation. Although I have not done a thorough analysis of the nature of the lexical distribution of the variants of (dʒ) in the dataset, the token extraction process is governed by the above general observations and the literature on the lexical distribution of (dʒ).

5.3 Envelope of variation and token extraction

The envelope of variation¹⁵¹ includes instances of the glide resulting from the historical *g→[j] lenition of etymological *g, whether occurring as a singleton or a geminate, along with its current variants in the dialect under study. This means that the other instances of the palatal glide, e.g., words with the etymon *y, as in /l-jimi:n/ ‘DEF-right’ and /jo:m/ ‘day.SGM’, are excluded. Also excluded from the envelope of variation are instances of the glide with the etymon *q. These are a result of extreme lenition of Arabic /q/; an example of this is the word /ja:ʕid/ <*q-ʕ-d ‘a camel’s saddle’.¹⁵² Similarly, words which include the fronted variant of

¹⁵⁰ Cf. /qʌimʕi/ ‘Friday’ (a standard-like analogue).

¹⁵¹ Tagliamonte (2006: 13); Milroy and Gordon (2003: 152-154).

¹⁵² Dialects that show this pattern are normally Bedouin in type. The etymological *q in such dialects is realised as [q]. This stop is sometimes conditionally fronted and realised as an affricate (in this case, it is realised as [qʌ], since this is the allophone for fronted *q in this dialect as previously mentioned in § 2.2.1.1 in Chapter Two). This stop can be further lenited to [j], although this process seems to be much less common (see examples in Owens, 2006: 245; Johnstone, 1965: 241; and Holes, 1980: 77 for foreign borrowings where the underlying velar is lenited and ended up being realised as the glide). Holes (personal communication) argues that the lenition of etymological *q to /j/ (via < dʒ < g < q) occurs in Bahrain, but in very few words only, e.g., /tanji:l l-ma/ ‘transporting water’, and /jassam/ ‘divide’ (my transcription); however, the lenition of etymological *j (i.e., *g) is ‘virtually categorical’. He further hypothesises that the

Arabic /q/ in this dialect, i.e., words with /qj/ < *q, e.g., /wa:di qja:sim/ < *q-s-m ‘a local toponym’, /qjidda:m/ < *q-d-m ‘in front of’, and /qja:bl-i/ < *q-b-l ‘watch out-2SGF’, are excluded.

The (dʒ) tokens were extracted from the interviews using E-LAN versions 5.4 (2018) to 5.8 (2019). In E-LAN, I transcribed the (dʒ) tokens themselves, and the adjacent context, then coded for the dependent variants in separate tiers for each participant, as Figure 5.1 shows. I extracted all the tokens of (dʒ) throughout the interviews.¹⁵³ I then exported them as tab-delimited text files including the time stamp and duration of each token. These files were then saved as Excel files, which in turn were cleaned up by excluding all the tokens where there is no variation (categorical tokens), and repeated tokens beyond the maximum limit of three tokens of the same word per speaker.

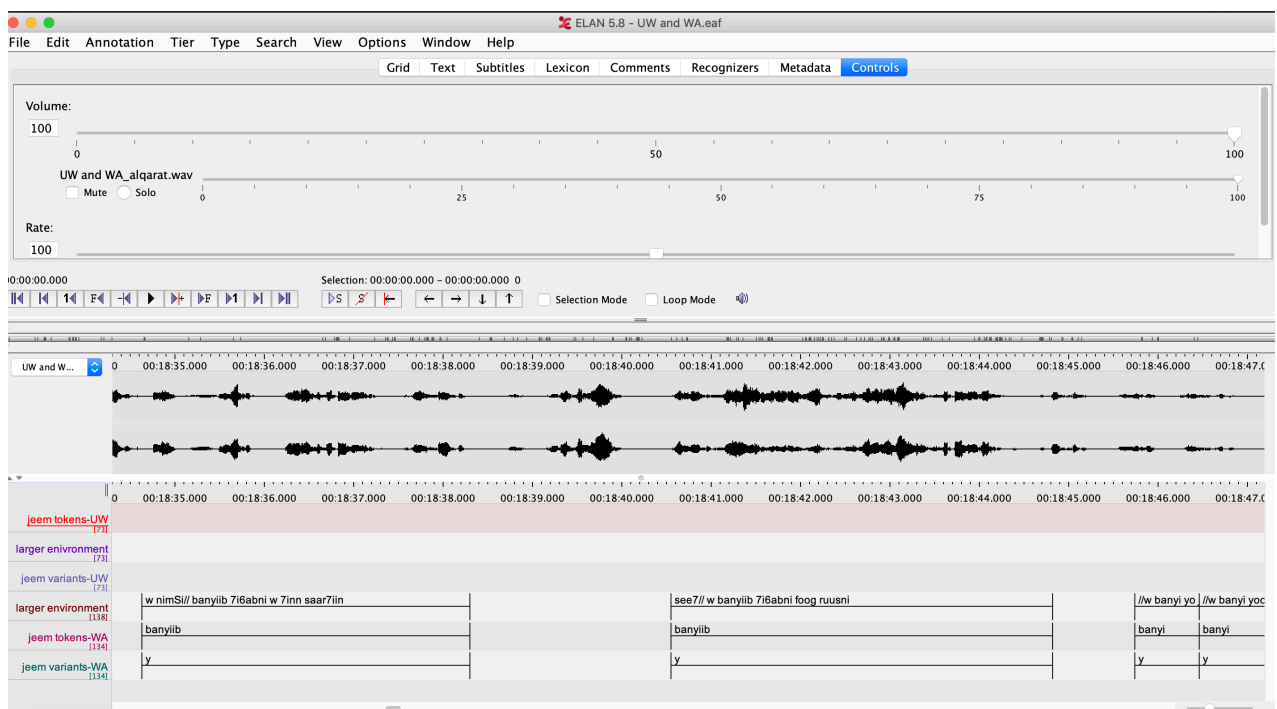


Figure 5.1 Screenshot of E-LAN as used in the analysis for this variable.

lenition of etymological *j to /j/ is an old change, older than the lenition of etymological *q to /j/, and that by the time the later lenition of etymological *q > /q/ and then to /dʒ/ occurred, the lenition etymological *j to /j/ was no longer operative; so, the fact that the speakers seemingly treat /dʒ/ differently on the basis of its etymological status is only apparent (ibid).

¹⁵³ The only exception is tokens in parts of the interviews when the informants commented on the variable (dʒ) itself, i.e., when they make metalinguistic comments on the variable.

One important note to mention here is that tokens where there is more than one instance of the variable are treated differently according to the nature of the segments involved. So, for instance, words with true geminates,¹⁵⁴ e.g., /ʕi:d l-ħijj/ ‘Eid al-Adha’ and /l-ħagʕi/ ~ /l-ħigʕi/ ‘the pilgrimage’ are treated as one token where the geminate is treated as a singleton, a long consonant that comprises the coda in the syllable with (dʒ). However, in a word like /d-di.ja:j/ ‘DEF-chicken.PL’, there are two singleton instances of (dʒ), one is an onset and the other a coda. These tokens are doubled in the Excel sheet. This is also the case with words including geminates resulting from derivation, like in /raj.ja:l/ ‘man’ and /a-xaj.jil/ ‘1SG-shy.IPFV’, where the geminate also comprises the coda of a preceding syllable and the onset of the following one.

A final Excel sheet for all of the informants was created, with N=2,627 tokens in total, and an average of N=65 tokens per informant. This Excel sheet was subsequently saved in a comma separated file format (.csv) in order to be used for Rbrul analysis. In addition, tokens of the three other variants of (dʒ), namely [dʒ], [g] and Ø were excluded from the multivariate analysis because their distribution is restricted; in the final Excel sheet, there were few tokens of each (N=69, N=6, N=42 respectively) compared to [j] (N=1,801) and [gʲ] (N=709). Thus, the (dʒ) variable is treated as a categorical binary variable with [j] as the traditional ‘conservative’ variant and [gʲ] as the incoming variant.

¹⁵⁴ See also Hussain (2017: 197); Alshawi (2019: 123); and Embarki (2013) for more on gemination in Arabic.

5.4 The coding procedures for (dʒ) and the modelling process

Based on the data collected, /dʒ/ as a phoneme occurs in various positions in the word, preceded and followed by various segments, but some trends could be noticed with regard to the distributional characteristics of this phoneme. For example, in terms of contextual positions, expectedly since we are dealing with a variety of Arabic, this sound mostly occurs in C-V (pre-vocalic) and V-V (intervocalic) positions, e.g., /l-jɑ:r/ ‘DEF-neighbour.SGM’ and /diʒɑ:j/ ‘chicken.PL’ respectively. Two other positions are also popular in the data, namely V-C (post-vocalic), e.g., /l-mɑjmar/ ‘DEF-frankincense burner.SG’, and to a lesser extent the post-pausal position, respectively. Also, this phoneme can freely appear in pre-consonantal and post-consonantal positions, as an onset or a coda. The C-C position is quite restricted in distribution in the data, and apart from words where it occurs as a geminate, /dʒ/ in a C-C position occurred across word boundaries, e.g., /da:xit jħɑ:l/ ‘inside pot.PL’. It is noteworthy that, generally speaking, the contexts in which the variation of (dʒ) happens in this dialect are varied in nature but mirroring those of the general distribution of the phoneme itself. Bearing these points in mind, the remaining tokens considered for this analysis, N=2,510, were coded for the following internal and external factor groups. The coding of the internal predictors is largely guided by the literature both linguistic and variationist as presented in § 4.4 and 4.5 in Chapter Four.

5.4.1 Internal predictors

Preceding and following segments

The preceding, /ʕ tʰ h sʕ ð ɡʲ kʲ ʔ ʃ θ a a: b d e e: f ɡ h i i: æ æ: k l m n o o: q r s t u u: w j/, and following, /ʕ h ð ɡʲ a a: b d e e: f h i i: æ æ: k l m n o: r s t u u: w x j z/, segments are impressionistically coded as they are to guide further groupings according to the place and manner of articulation. Pauses are also coded for in these factor groups (see Appendix A).

Number of syllables

The words are also coded as ‘monosyllabic’, e.g., /gʲa:t/ ‘come.PFV.3SGF’, ‘disyllabic’, e.g., /ja:t-ni/ ‘come.PFV.3SGF-1SG, i.e., she came to me’, ‘trisyllabic’, e.g., /ɫ.-gʲim.ʃi/ ‘DEF-Friday’ and ‘polysyllabic’, e.g., /it.-ti.ʃagʲ.gʲi.'b-e:n/ ‘2-wonder.IPFV-SGF, i.e., you wonder’.

Position in syllable

In other words, whether the variant is a ‘coda’ or an ‘onset’, e.g., the /j/ in /ni-j.lis/ ‘1PL-sit.IPFV’ is a coda, and that in /ji:.ra:n/ ~ /gʲi:.ra:n/ ‘neighbour.PL’ is an onset.

Stress

The tokens are also coded as ‘stressed’ or ‘unstressed’. Stress is assigned in the light of the stress assignment patterns discussed in Holes (2004: 62-3; 2007: 481). First, in di- and tri-syllabic words with light syllables (CV, CV:, CVC), the penult is stressed, e.g., /'fa.gʲir/ ~ /'fa.jir/ ‘dawn’, unless the antepenult is CV: and the penult is CV, then in this case the antepenult is stressed.¹⁵⁵ If both the antepenult and the penult are CV, then the antepenult is stressed, e.g., /'ji.la.s-it/ ~ /gʲala-sit/ ‘sit.PFV-1SG’. Also, if there is a heavy or a long (CV:C, CVCC) syllable in the word, then it is stressed, /'ja:l.s-a/ ‘sit.ACT.PTCP-SGF’. On the other hand, if there are two long syllables, then the second one is stressed, e.g., /ja:l.'s-i:n/ ‘sit.ACT.PTCP-PLM’ and /ji:.'ra:n/ ~ /gʲi:.'ra:n/ ‘neighbour.PL’.

In addition, in polysyllabic words (words with four or more syllables): a) if the word contains a CV:, then it is stressed; b) if there are more than one CV: syllables, then the one nearest to the end of the word is stressed, e.g., /ba-t.-'ji:.b-il.-ni/ ‘FUT-3-bring.IPFV.SGF-DAT-1PL; i.e., she will bring for us’; c) if there is no CV:, then the antepenult is stressed, e.g., /ji-t.'gʲam.mi-ʃ-aw/ ‘3-gather.IPFV-PLM; i.e., they gather’. Finally, if there is a long syllable

¹⁵⁵ In my data, this situation did not occur.

(CV:C, CVCC), the stress always falls on this syllable, e.g., /it.-ti.raj.ˈj-a:h/ ‘2-wait.IPFV.SGM-3SGM.ACC; i.e., you wait for him’ and /it.-ti.ʃaɟi.ɡi.ˈb-e:n/ ‘2-wonder.IPFV-SGF, i.e., you wonder’.

Weight of syllable

The weight of the syllable in which the variant occurs; I divided them into ‘light’ and ‘heavy’ based on Holes (2004: 62-3): ‘light’ also includes the more commonly known as the heavy syllables CV: , e.g., /ji:t-i/ ‘come.PFV-2SGF’ and /ɡi:ra:n/ ‘neighbour.PL’ , and CVC, e.g., /ni-jlis/ ~ /ni-ɡlis/ ‘1PL-sit.IPFV; i.e., we sit’, whereas ‘heavy’ describes the super heavy syllables CV:C, e.g., /ja:r/ ~ /ɡja:r/ ‘neighbour.SG’, and CVCC, e.g., /jamb/ ~ /ɡjamb/ ‘side’.

Syllable type

Whether the syllable is ‘open’, e.g., CV → /ɡi.hid/ ‘effort’, and CV: → /ma.ɡja:.lis/ ‘guestroom.PL’, or ‘closed’, e.g., CV:C → /ja:b/ ‘bring.PFV.3SGM’, CVC → /maɟi.las.-sum/ ‘gustroom.SGM-PLM.GEN’, and CVCC → /l-janz/ ‘DEF-store room’.

Gemination

The tokens are also coded as ‘geminate’ or ‘singleton’ (see § 5.3).

Table 5.1 provides an example of how the (dʒ) tokens are coded, adapted from the Excel sheet prepared for the Rbrul analysis.

Token	(dʒ)	preceding	following	gemination	stress	number	position	weight	type
l-gʒimfi	gʒ	l	i	singleton	stressed	tri	onset	light	closed
ja:tɲi	j	// ¹⁵⁶	a:	singleton	stressed	di	onset	heavy	closed
rajja:l	j	a	j	geminate	unstressed	di	coda	light	closed
rajja:l	j	j	a:	geminate	stressed	di	onset	heavy	closed
ma:gʒu:da:t	gʒ	a:	u:	singleton	unstressed	tri	onset	light	open
l-masgʒid	gʒ	s	i	singleton	unstressed	tri	onset	light	closed
batʒi	j	t	i	singleton	unstressed	di	onset	light	open
ja:jibtinje	j	i	a:	singleton	stressed	poly	onset	light	open

Table 5.1 Example of (dʒ) coding.

External predictors

<i>Age</i>	Older group (51+)	Middle group (30-50)	Younger group (18-29)
<i>Locality</i>	Al-Batha	Al-Tharmad	Al-Khadhra
<i>Gender</i>	Male	Female	

¹⁵⁶ The code used for a preceding or a following pause.

5.4.2 Modelling

This section sums up the modelling that led to the final model on which the results section is based. The application value in all of the modelling stages is [gⁱ].

Preceding and following segments

The preceding and following segments were coded according to their place and manner of articulation, and later on according to their contextual position. With regard to the place of articulation, I first ran the preceding and the following segments individually; based on these initial runs, I then conflated the segments according to their places of articulation. The first run of the preceding place of articulation showed a concentration of tokens around ‘dental-alveolar’, ‘high front vowel’, ‘low back vowel’, ‘labio-velar’, ‘pause’, and ‘palatal’.

Subsequent modelling was done to conflate low N values based on the similarities between the different factors, on the percentages of the application value, on the log odds ~ factor weights, but also guided by the existing literature. This resulted in an initial grouping of eleven factors namely ‘palatalised velar’, ‘back consonant’, ‘low back vowel’, ‘non low-back vowel’, ‘high front vowel’, ‘non-high front vowel’, ‘labial’, ‘labio-velar’, ‘palatal’, ‘coronal’, and ‘pause’. The palatalised velars were subsequently conflated with back consonants, the back vowels were all put in one group, the front vowels also were put in one group, and the labio-velar was included with the labials; the palatal is left as a separate factor, thus giving us a final grouping of the preceding environment consisting of seven values:

- ‘back consonant’: pharyngeal, laryngeal, uvular, velar, and palatalised velar;
- ‘labial’: bi-labial, labio-dental, and the labio-alveolar glide /w/;
- ‘coronal’: inter-dental, dental-alveolar, and palato-alveolar;
- ‘front vowel’: high, mid, and low front;
- ‘back vowel’: high, mid, and low back;

- ‘palatal’: the glide /j/;
- ‘pause’.

The following place of articulation went through a similar modelling process. The difference here is that due to considerably low token numbers of ‘coronal’ (N=184), ‘labial’ (N=49), and ‘back consonant’ (N=35), compared to the larger groups of ‘front vowel’ (N=1,081) and ‘back vowel’ (N=1,090), all of the following consonants were conflated in one group ‘consonant’. This latter group excludes the palatal glide /j/ (N=60) which was conflated with ‘front vowel’. One more change to the following environment involves the exclusion of ‘pause’ due to the relatively low number of tokens (N=11). Although the following pause was excluded earlier on in the analysis, the preceding pause was excluded in some stages of the modelling to reduce direct and indirect interaction with other groupings like syllabic position and manner of articulation, but eventually, it was retained in the final stage of modelling.

Furthermore, the preceding and the following segments were coded according to their manner of articulation into ‘fricative’, ‘plosive’, ‘approximant’, ‘glide’, ‘nasal’, and ‘sibilant’ for consonants, and ‘long’, and ‘short’ for vowels. These were then grouped and regrouped in an attempt to fit both the place and the manner factor groups in the same model, and at the same time, to avoid as much overlap as Rbrul can allow. This was possible when excluding preceding pause and with a conflated manner of articulation with two factors, namely ‘obstruent’ vs. ‘sonorant’; however, the gap in the token number between the two factors is significantly large especially in the following environment (‘obstruent’: N= 92; ‘sonorant’: N=2,269), in addition to the fact that this means excluding the ‘pause’ altogether, meant that it was best to exclude ‘manner of articulation’.

Syllabic position

‘Position in syllable’ was initially coded in terms of whether the dependent variable is in a coda or an onset position within the syllable where it occurs. This factor group always overlapped with the following place of articulation, even when it was recoded at a later stage as: ‘branching onset’, ‘single onset’, ‘branching coda’, and ‘coda’ in an attempt to fit it in the same model with the place of articulation. In any case, ‘syllabic position’ is excluded in the final model, because of the direct overlap with ‘place of articulation’ which is sometimes shown through statistical non-significance or primarily as a *vif* column (ranging from *vif*>2.5 to *vif*>20) in the results from the Rbrul runs. In his Rbrul manual, Johnson (2009) states that:

“VIF (variance inflation factor) is a measure of multicollinearity. Values above 5 (or at least above 10) are thought to show that a predictor is highly correlated with the others. This is not necessarily a problem, especially if the correlated variables are control variables or nuisance variables rather than those of interest.”

Number of syllables

In addition, in the number of syllables factor group, ‘polysyllabic’ (N=136) and ‘trisyllabic’ (N=791) tokens are conflated as ‘polysyllabic’; they both favour the application value, unlike the mono- and disyllabic tokens, as will be seen soon.

The linguistic context

At a later stage in the modelling process, a linguistic context factor group that involves the interaction between the preceding and the following environments was added to the analysis. Because the dependent variants were coded for the preceding and the following contexts in terms of whether they are preceded and/or followed by a sonorant, an obstruent, a glide, a vowel or a pause, and, later on in terms of whether they are preceded or followed by a

consonant, a vowel or a pause, the linguistic context factor group integrated both the syllabic position, and the manner of articulation. However, even when totally excluding contexts with ‘pause’, this factor group directly overlapped with different combinations of both of the preceding and following environment factor groups (*vif*>5 and higher).

Gemination

The number of ‘geminate’ tokens is very low (N=94) compared to the singletons (N=2,416). When incorporated in the modelling, it was chosen as non-significant; thus, it was excluded from the final model, primarily on the basis of the imbalance in the token number for the two values.¹⁵⁷

Table 5.2 summarises the internal and external predictors included in the final model.

Predictors	Factors
Preceding environment	front vowel, back vowel, back consonant, coronal, labial, pause, palatal
Following environment	front vowel, back vowel, consonant
Stress	stressed, unstressed
Number of syllables	mono-, di-, polysyllabic
Syllable type	open, closed
Syllable weight	light, heavy
Age	younger, middle, older
Gender	female, male
Locality	al-Khadhra, al-Batha, al-Tharmad

Table 5.2 Internal and external predictors included in the final model.

¹⁵⁷ This is a similar case to gemination in Hussein’s (2017: 197) study, as well as in Alaodini’s (2019) study; the latter initially coded for gemination but had to eventually exclude it from the analysis (personal communication).

5.4.3 Results and discussion

Table 5.3 (continued overleaf) presents the results of a step up/step down analysis in Rbrul of the variation in (dʒ) in the dataset for the *Yāl Sa ‘ad* tribe. The log odds and factor weights (FW) for the different factors are both presented here. These values correlate with the effect (significance) for the different factors on the variation at hand; positive log odds and FWs above 0.50 indicate that the factor in question favours the application value, whereas the opposite is true for negative log odds and FWs below 0.50. A log odd of zero or a FW of 0.50 represent neutral values, i.e., the effect of the factor in question on the variation is neutral; it neither favours nor disfavours the application value (see Tagliamonte, 2012; Johnson, 2009: 361).

Application value= [gʲ]; N= 2,499; overall proportion= 28.1%				
R ² = 0.31; log likelihood= -1244.844				
Predictor	Log odd	FW	N	% [gʲ]
Preceding environment (p<0.0001) (p=1.19e-25)				
Coronal	0.706	0.67	819	34.8%
Back vowel	0.552	0.635	635	33.7%
Front vowel	0.326	0.581	303	34.7%
Labial	0.260	0.565	175	23.4%
Pause	-0.079	0.48	138	19.6%
Back consonant	-0.235	0.441	50	20%
Palatal	-1.529	0.178	379	5%
Number of syllables (p<0.0001) (p=1.63e-19)				
Polysyllabic	0.748	0.679	922	40.2%
Disyllabic	-0.091	0.477	1251	22.7%
Monosyllabic	-0.656	0.342	326	14.1%

Following environment (p<0.0001) (p=1.52e-16)				
Consonant	0.440	0.608	268	42.5%
Back vowel	0.197	0.549	1090	32.8%
Front vowel	-0.637	0.346	1141	20.2%
Age (p<0.0001) (p=4.18e-14)				
Middle	0.438	0.608	941	35.9%
Younger	0.028	0.507	661	27.1%
Older	-0.467	0.385	897	20.5%
Gender (p<0.0001) (p=3.53e-10)				
Male	0.313	0.578	969	35.5%
Female	-0.313	0.422	1530	23.3%
Syllable weight (p<0.001) (p=0.000232)				
Light	0.254	0.563	1717	32.1%
Heavy	-0.254	0.437	782	19.1%
Syllable type (p<0.03) (p=0.0247)				
closed	0.136	0.534	1581	28.3%
open	-0.136	0.466	918	27.7%

Table 5.3 Results for the use of [gʲ] in the dataset.

The results based on the multivariate analysis of the data collected generally show that there is limited variability in the dataset, with 28.1% in total for the application value [gʲ] (N=701). In addition, the log odds and intercept of the model show that the application value is generally disfavoured in the dataset, with a ‘preceding palatal glide’ as the most disavouring environment for the [gʲ] variant at a factor weight of 0.178. On the other hand, the traditional variant [j] is largely preserved with an overall proportion of 71.9% (N=1,798). Rbrul returned all predictors as statistically significant apart from stress and locality:

preceding environment > number of syllables > following environment > age > gender > syllable weight > syllable type, in order of significance.¹⁵⁸ It is noteworthy that some predictors show greater differences in the FWs and log odds, whereas others are rather gradient and cluster around the neutral points, like syllable weight and type.

Furthermore, a closer look at the results for the internal predictors shows that the application value is favoured when preceded by a back vowel, a front vowel, a labial and a coronal, and when followed by a consonant and a back vowel. On the other hand, the use of [gʲ] is disfavoured when the dependent variable is preceded by a palatal, a back consonant and a pause, and when followed by a front vowel, although the preceding pause could be best described as neutral with a FW=0.48. The use of [gʲ] is also favoured with a) polysyllabic words, b) in light syllables, and c) in closed syllables. Polysyllabic words (FW=0.679) and preceding coronal sounds (FW=0.67) are the most favouring contexts for the application value; the use of [gʲ] is mostly disfavoured with a preceding palatal sound (FW=0.178) and in monosyllabic words (FW=14.1).

The results regarding the internal constraints very much go in line with what Alshawi (2019) has found for similar constraints with regard to the fortition of (dʒ) in the southern Iraqi variety of Qal'at Sikr. The same dissimilarities between Alshawi's study and Alaodini's (2019) study with regard to the number of syllables (§ 4.5 in Chapter Four) are also found to be true about this study in that mono- and disyllabic words disfavour the incoming variant whereas polysyllabic words favour it. With regards to the place of articulation, the picture is a bit more complicated since the grouping of the factors within the preceding and following segment constraints for Alaodini's (2019) study and this one are different.

¹⁵⁸ It is interesting that stress is also not selected as statistically significant in the fortition of (dʒ) in Alshawi (2019) and Alaodini's (2019) studies, whereas in Hussein's (2017), stress is selected as significant in the lenition of (dʒ) both in the urban and the Bedouin datasets (§ 4.4 in Chapter Four).

In addition, in relation to the linguistic constraints on fortition and lenition discussed in § 4.5 in Chapter Four, the results regarding the preceding and following environments do not necessarily agree with the phonological studies reviewed in terms of positional factors and their relation to fortition processes. For instance, the literature suggests that a preceding pause is a favouring environment for fortition; however, this factor is almost neutral in the variation at hand. It also suggests that pre-vocalic contexts are strong contexts; however, when the preceding and following segments were recoded as ‘vowel’ and ‘consonant’,¹⁵⁹ the use of [gʲ] is disfavoured in a following vowel context, although the results on the use of [gʲ] with regard to preceding and following consonants are in line with the literature that argues that a preceding and or a following consonantal environment is a favourable context for fortition.¹⁶⁰ Furthermore, the use of [gʲ] was favoured when preceded by a front vowel (FW=0.581), which is in line with the results from Bybee and Easterday’s (2019) study on the important effect of a preceding front vowel on fortition, but when looking at vowel height, the use of [gʲ] does not favour a certain context, since it is found to be favoured in all preceding vocalic contexts high, mid and low; it is also favoured with following non-high vowels, high back vowels, and disfavoured with following high front vowels ($R^2=0.216$ for this run).¹⁶¹

One of the assumptions based on initial observations about the distributions of the dependent variable during the coding stage with regard to preceding palatal sounds and the use of [gʲ] is that this factor would trigger a following palatal glide. The results support this prediction; only 5% of the total tokens preceded by a palatal glide (N=379) are realised with [gʲ].¹⁶²

¹⁵⁹ The palatal glide /j/ is coded as a vowel here. The R^2 number for this run (preceding and following segment) is 0.034 which is very low.

¹⁶⁰ Recall that Bybee and Easterday (2019: 281-82) found cases of fortition in weaker positions, although less common cross-linguistically.

¹⁶¹ The glide /j/ is coded as a high front vowel in this run as well.

¹⁶² Recall that tokens with following palatal glides are low compared to the rest of the factors for this group, and thus, they were recoded as front vowels.

Furthermore, the results for the use of the incoming variant [gʲ] for ‘stress’ and the other linguistic predictors (see Appendix B) that were examined during the modelling process (i.e., manner of articulation, syllabic position, and linguistic context), but which were primarily excluded from the final model due to high *vif* values, and/or unbalanced distribution of tokens for the individual factors within these predictors show that the proportion of the incoming variant is slightly more in unstressed syllables than stressed ones (30.6% and 26.4% respectively). In terms of the ‘linguistic context’, the incoming variant occurs more in a postvocalic context in V_C and V_V contexts (37.2% and 32.4% respectively) than in a postconsonantal context or when preceded by a pause (19.4%); it also occurred more in a coda position (37.7%) than an onset or a branching onset positions (28.4% and 9.5% respectively), and a singleton (28.6%) more than a geminate (14.9%).¹⁶³

With regard to ‘gemination’, which was analysed separately, the case is similar to the former predictors with regards to the correlation between the use of [gʲ] and whether the dependent variant is part of a geminate or not. Nine out of the total geminate tokens are realised with [gʲ] (N=14; total N=94) when preceded by a back sound (i.e., a back consonant or a back vowel); on the other hand, 12 out of these tokens are followed by a front vowel. However, due to the low percentage of geminates with [gʲ] in the dataset, the only conclusion we can draw from this observation is that fortition also takes place in tokens when the variant occurs as geminate, but it occurs more when the variant is realised as a singleton (28.6% out of the total for singletons compared to 14.9% of the total of geminates), when preceded by a back sound, and when followed by a front vowel.

It is noteworthy that when the internal constraints were run together, the R^2 is relatively low, $R^2=0.257$, i.e., they can only explain 25.7% of the variation in (dʒ). It is therefore

¹⁶³ The percentages in this section are out the total tokens of the respective factors within each facto group.

important at this point to emphasise that the conclusions we can draw about the statistical results pertaining to the linguistic constraints are fairly tentative and constrained by many factors, the most primary of which is the overall relatively low percentage of the fortis variant [gʲ] in the whole dataset. In addition, a more thorough and multidimensional phonological analysis is required to explore the intricacies of the different results and arguments in the literature on fortition and lenition so one would be able to better understand and comment on the relation of the different internal positional constraints analysed here and their effect on the variation at hand. What we can probably cautiously conclude with regard to the place of articulation of the adjacent segments based on the results of the statistical model presented here is that a following (high) front vowel, and a preceding palatal glide are two salient contexts for maintaining the traditional variant, and a preceding coronal segment and a following consonant (front or back) are favourable contexts for the incoming variant [gʲ].

Having shed light on the results for the internal constraints on the variation at hand, the next section presents a closer look at the results regarding the external predictors and examines the use of the incoming variant per speaker as well as the linguistic behaviour of certain informants in relation to the use of the new variant.

According to the results from Rbrul, ‘age’ is returned as the most significant external predictor with the middle age group (FW=0.608) favouring the incoming variant the most; the younger age group is best described as neutral (FW=0.507), whereas the older age group generally disfavours the use of [gʲ] (FW=0.385). It is noteworthy that as a group, the external predictors can hardly explain the variation at hand if considered for the analysis without the internal factors ($R^2=0.065$).¹⁶⁴ Table 5.4 visually presented in Figure 5.2 shows that the traditional variant is salient across the three groups. The middle age group uses [gʲ] 36% of

¹⁶⁴ Locality is returned as significant in this run; al-Batha ([gʲ]=30.4%) and al-Tharmad ([gʲ]=28.1%) favour the incoming variant, whereas al-Khadhra ([gʲ]=25%) disfavours it.

the time, whereas the younger age group uses it 27%, and the older age group 21% of the time. The gap between the middle age group and the younger age group is slightly higher (9%) than that between the younger and the older age group (6%). When the use of [gʲ] per age group is ordered from the younger to the older group, we can notice a slightly curved line with the middle age group at the peak, whereas the younger and older age groups fall at the two ends (see Figure 5.3).

Age group	% [gʲ]	% [j]	Total per age group
Younger	27.1 N=179	72.9 N=482	N=661
Middle	35.9 N=338	64.1 N=603	N=941
Older	20.5 N=184	79.5 N=713	N=897
Totals	28.1 N=701	71.9 N=1,798	Total N=2,499

Table 5.4 Distribution of [gʲ] per age group.

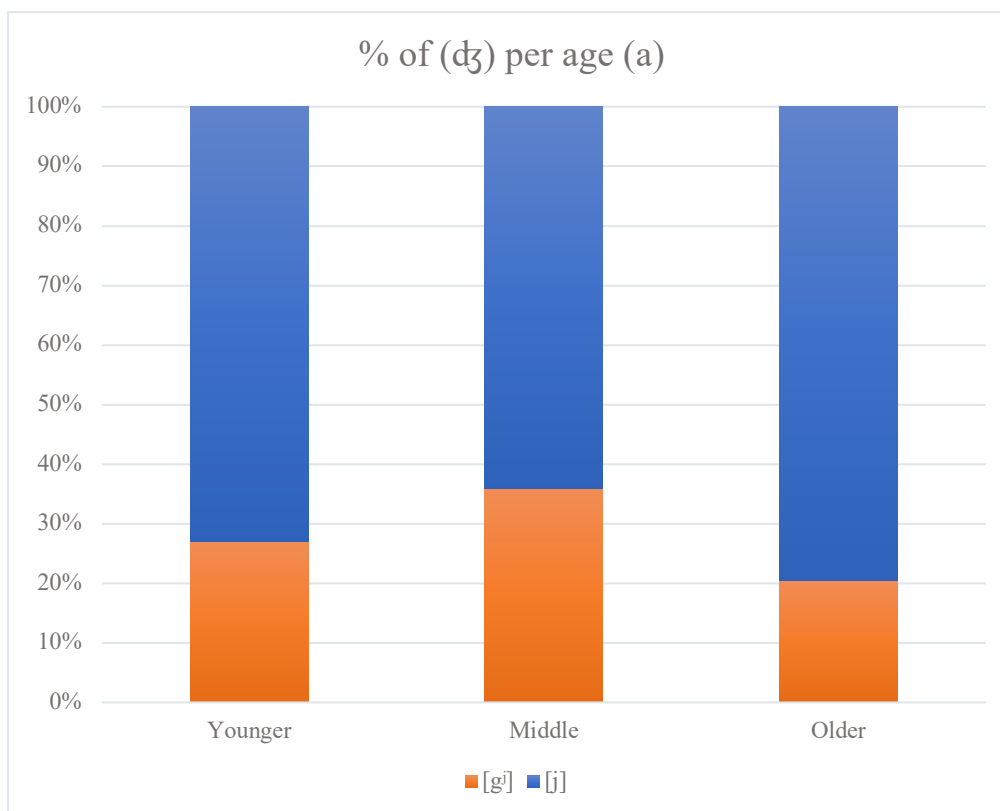


Figure 5.2 Distribution of (dʒ) variants per age group (a).

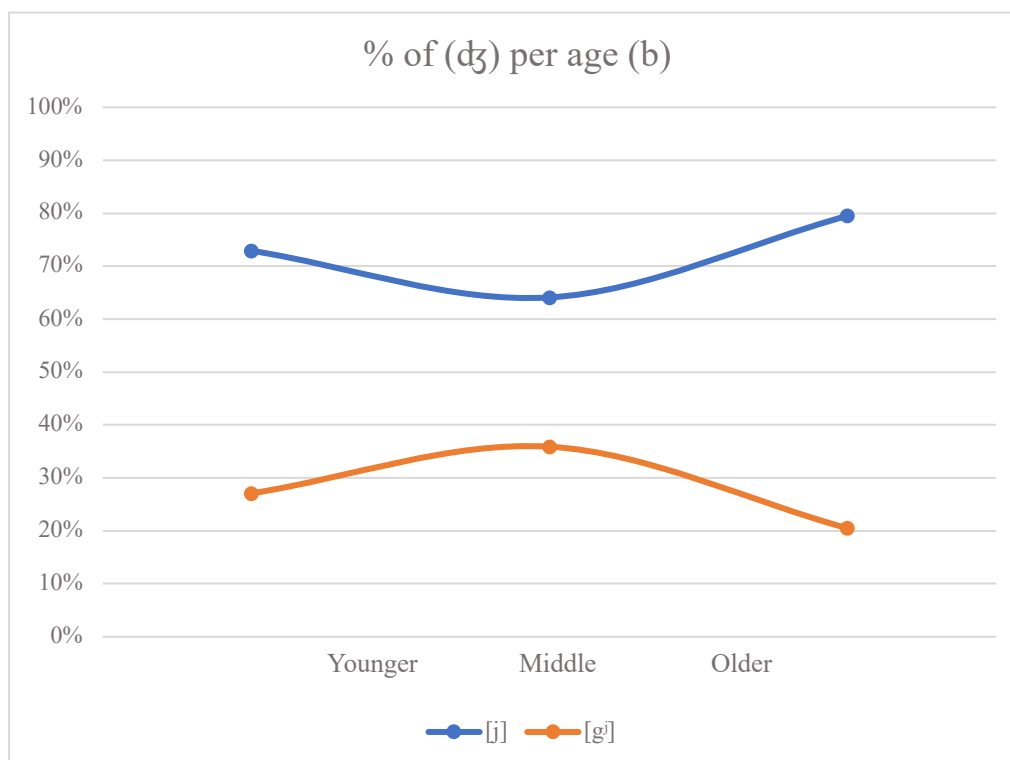


Figure 5.3 Distribution of (dʒ) variants per age group (b).

Gender comes next in terms of statistical significance to age with men favouring the use of [gʲ] with a FW of 0.578; they use it 36% of the time. Women use the incoming variant 23% of the time and with a FW of 0.422, they generally disfavour it. Although the gap between men and women is not big (12%), it is slightly higher than the gaps between the three age groups as discussed above. Furthermore, a crosstab for age and gender presented in Table 5.5 shows that the use of [gʲ] is mostly favoured by men in the middle age group (proportion=39.4%), followed by younger men who use it 36.8% of the time, and women in the middle age group who use it 33.7% of the time. Older women use the incoming variant the least (12.9%). The results also show that young men exhibit more use of [gʲ] than women in the middle age group, and that older men show more use of [gʲ] than younger women.

Gender	Age			
	Middle	Older	Younger	Total per gender
F	33.7% N=192/570	12.9% N=67/519	22.2% N=98/441	N=1,530
M	39.4 % N=146/371	31% N=117/378	36.8% N=81/220	N=969
Total per age	N=941	N=879	N=661	Total N=2,499

Table 5.5 A crosstab showing the proportion of the use of [gʲ] per age and gender.

Furthermore, Figure 5.4 shows that within each age group, men use [gʲ] more than women. It also shows that the gap between men and women in the older age group is the highest (18%). On the other hand, the gap between men and women in the middle age group is the smallest (6%), i.e., men approximate women in their use of the incoming variant in this

age group. Figure 5.5 shows that women across all three age groups use [gʲ] less than men.

The gaps between women are higher than the gaps between the men in the three age groups.

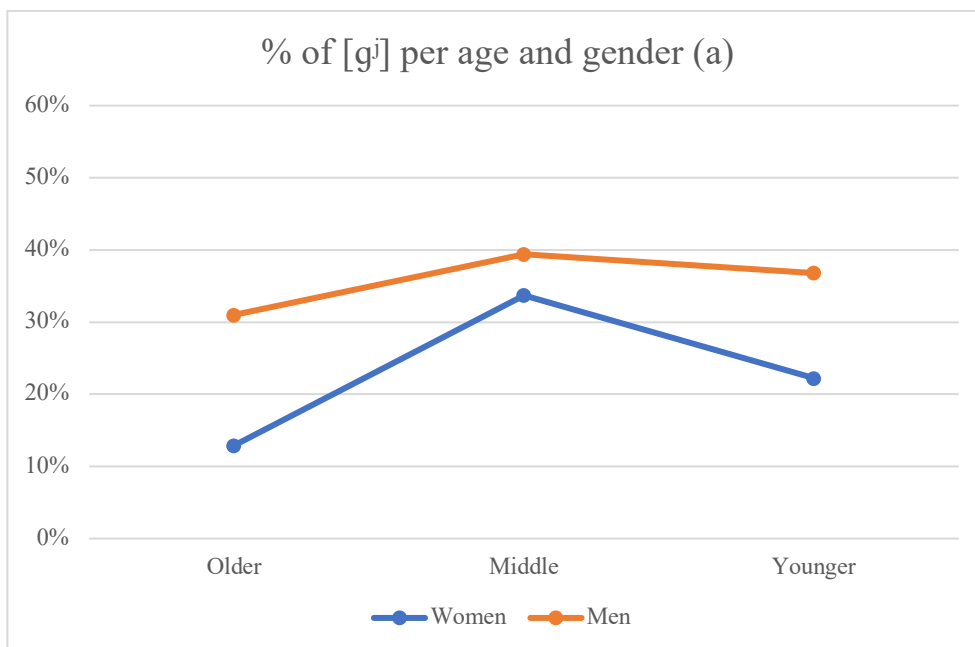


Figure 5.4 Use of [gʲ] per age and gender (a).

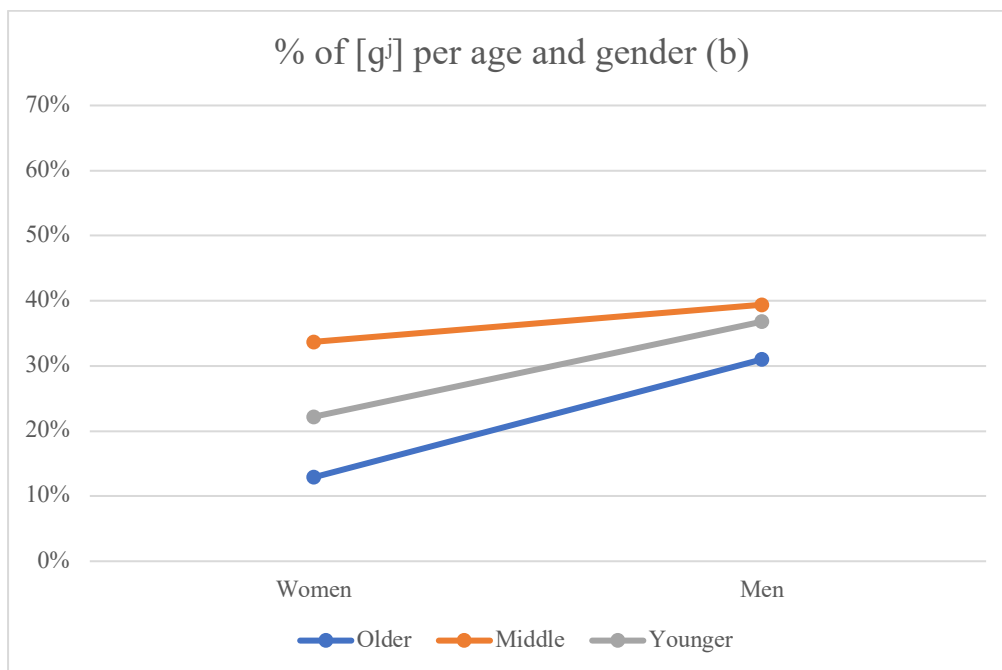


Figure 5.5 Use of [gʲ] per age and gender (b).

The last external predictor, namely locality, is selected as statistically insignificant; when ran individually, the R^2 number for locality equals zero, i.e., there is no statistical correlation between locality and the use of [gʲ] since there is no significant difference in the use of [gʲ] between the three localities; this is expected since there is a 6% gap between al-Batha, which showed the highest amount of variation (31%), and al-Khadhra which showed the least use of [gʲ] (25%). However, the correlation between locality and the other two external predictors of age and gender could be better explored via a crosstab; see Table 5.6 visually presented in Figure 5.6 below which demonstrates that, overall, the group that shows the least use of [gʲ] is older women of al-Tharmad locality (7%), whereas younger men from the same locality exhibit the most use of [gʲ] (90%). The second most users of [gʲ] in the dataset are women in the middle age group of al-Batha locality, who also are the highest users of [gʲ] among all of the women groups with a rate of (50%). In addition, men in general use [gʲ] more than women across the three localities and age groups, with the exception of women in the middle and younger groups of al-Batha locality who exhibit more use of [gʲ] than men. Also, women of al-Tharmad locality in the younger and middle age groups exhibit the same amount of variation (26%); they deviate from the older women's group who exhibit an overall low use of [gʲ]. In addition, the use of [gʲ] per gender is somewhat stable across the three age groups in the sample from al-Khadhra locality, with the men in the middle age group using [gʲ] more than the rest (38%). In this locality the difference between men and women in the older age group is lower than that of the men and women in the middle and younger age groups. Also, the gap between women in this locality is quite small compared to other groups; all of the women behave similarly to the older women in their use of [gʲ]. On the other hand, within al-Batha locality, the highest use of [gʲ] is exhibited by women and men in the middle age group (50% and 46% respectively). What is interesting about the results for this locality is that men in the younger age group exhibit less variation than women in the same age group, men in the

older age group, and men and women in the middle age group; they actually approximate the older women in their use of [gʲ] (12% and 13% respectively). It is noteworthy that the gender pattern across the three localities is more consistent in the older age group with men using [gʲ] more than women; the difference in the men's and women's use of [gʲ] is higher in al-Batha and al-Tharmad than in al-Khadhra. The behaviour of men and women in the latter locality with regards to the use of [gʲ] is quite similar. Finally, for the younger age group: the highest use of [gʲ] is exhibited by younger men from al-Tharmad, followed by women from the same locality. The difference between men of the younger age group in al-Batha locality and those in al-Tharmad locality is large (12% vs. 90%).

Locality	Gender	Age			Totals per locality
		Younger	Middle	Old	
Al-Batha	F	22% N=45/202	50% N=112/225	13% N=23/179	31% N=287/934
	M	12% N=11/89	46% N=61/132	33% N=35/107	
Al-Tharmad	F	26% N=40/154	26% N=57/217	7% N=14/190	28% N=239/848
	M	90% N=46/51	34% N=36/107	36% N=46/129	
Al-Khadhra	F	18% N=16/88	19% N=24/129	20% N=30/151	25% N=183/728
	M	32% N=26/82	38% N=51/135	25% N=36/143	
Totals per age group		28% N=184/666	36% N=341/945	20% N=184/899	28% N=709/2,510

Table 5.6 A crosstab for the use of [gʲ] per age, gender, and locality.

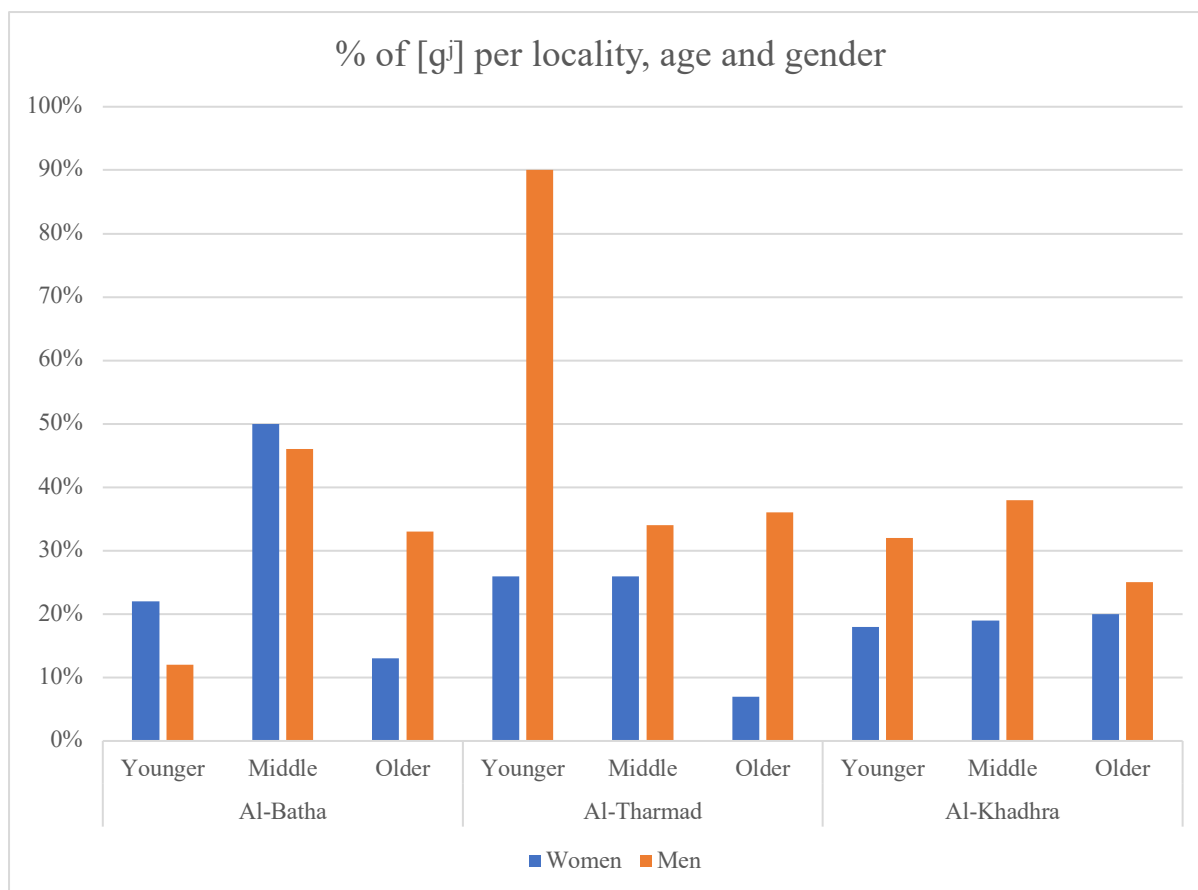


Figure 5.6 Use of [gʲ] per age, gender, and locality (N=2,510).

To sum up the results on the external predictors based on the cross tabulations and Rbrul analysis, we have seen so far that the middle age group uses the incoming variant more consistently than the other two age groups. Men generally use [gʲ] more than women. Al-Batha locality exhibits the most variation. However, some of the numbers need further discussion; in order to better understand these numbers, it helps to examine the results for the individual speakers, since these provide an important insight on certain variation patterns observed for the external predictors.

According to another Rbrul run in which ‘speaker’ is tested as an independent variable, the percentage of speakers who disfavoured fortition is 30%, i.e., twelve informants out of forty: four from al-Batha (25% of total number of informants for this locality, N=16), five from al-Tharmad (36% of the total for this locality, N=14), and three from al-Khadhra (30%

of total for this locality, N=10). Half of these informants belong to the older age group, and four of them belong to the younger age group; only two informants from the middle age group disfavour the application value. Figure 5.7 below is based on a crosstab for the use of [gʲ] per speaker; notice that the speakers with the least fortition proportions (below 10%) are all women (one from the younger, one from the middle, and two from the older age group).

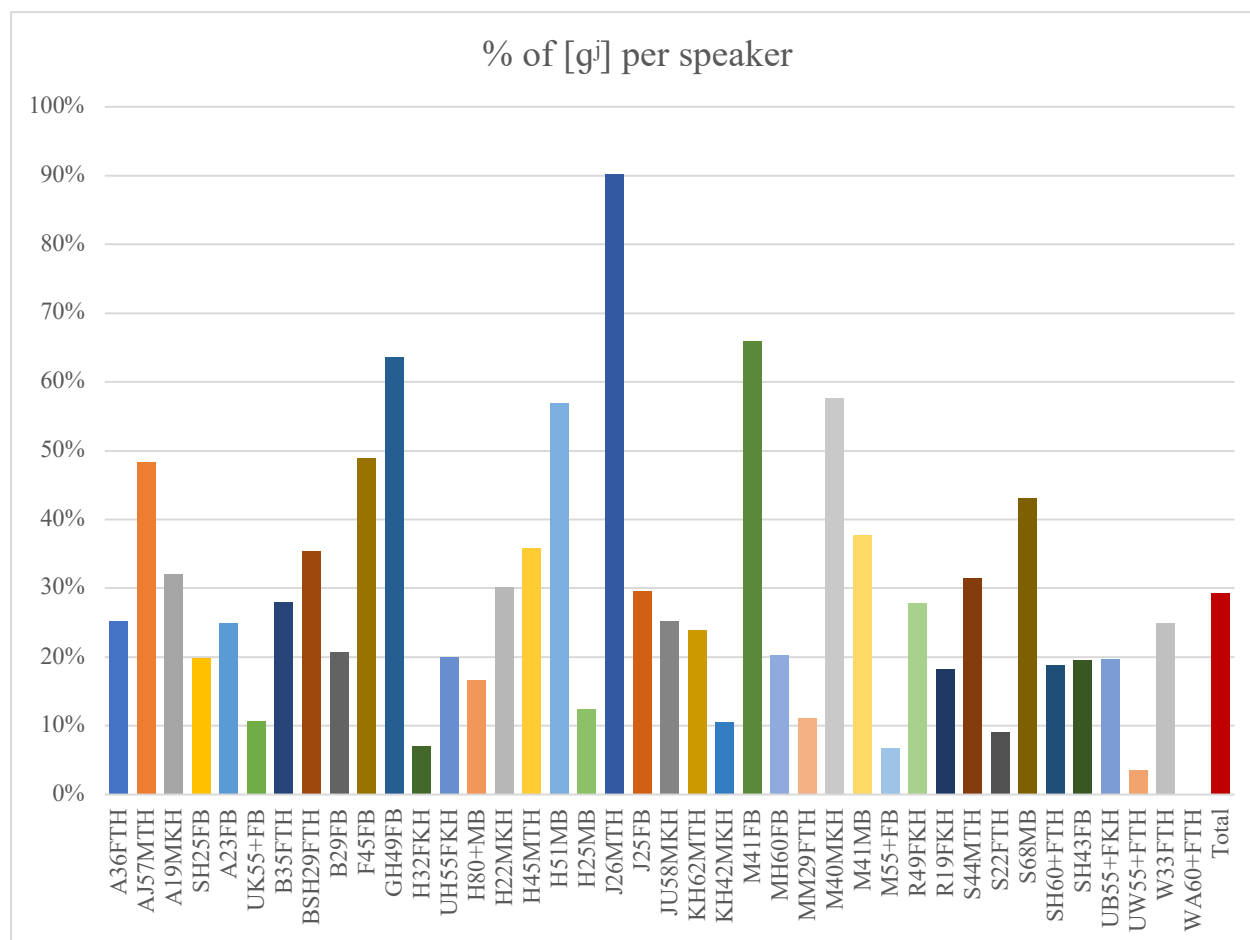


Figure 5.7 Use of [gʲ] per speaker; the speaker codes from bottom to top: the speakers' initials, exact age, gender and locality (N=2,510).

Figure 5.7 also presents the results for the most linguistically conservative speakers and the least conservative speakers, i.e., the most and the least resistant to the use of [gʲ]. There is one categorical speaker Zakiyya (WA60+FTH) who is an older woman from al-Tharmad locality. She is a housewife who also runs a small farm nearby her house. This woman and

her sister who happens to be the second most linguistically conservative speaker with 4% of fortition live in al-Bārda village, in the vicinity of al-Tharmad village itself, and which is located at the inland side of the highway crossing the study area. These women are both housewives, and they have lived in the area throughout their lives. When they were young, they lived in another small village slightly farther inland and later, their family moved to their current place where they got married to *Sa 'di* men. They barely had any kind of formal education;¹⁶⁵ the only teaching they had was the Quran lessons they attended in the area when they were children; they are functionally illiterate. The main activities these women were in charge of as young girls were taking care of cattle and chickens, fetching the water from the wells, and later on when they got married at young ages, they moved to their husband's houses; as grown women, their main job was looking after their homes, their children, and the cattle, including cows and goats. The main type of mobility these women experienced in their lives is internal and within their locality and group. In the summer, many families moved to live on their farms for the date harvest season. During the winter months, Bedouin families in particular moved farther inland to places where their animals could graze on the plantation that grew in areas affected by rainfall. This tradition has continued until recent decades. These families are fully settled now in their current places of residence; they may still own *'izba* (PL: *'zab*) in the sandy inland.¹⁶⁶ It is noteworthy that in areas surrounding al-Tharmad locality, this type of mobility was homogenous in nature; so, these families mostly stayed together and moved back and forth together. This helps us understand that these women have been exposed to limited contact with people who speak differently from them. Although the advances in the area after the 1970s created more chances for exposure to the

¹⁶⁵ The younger sister attended Grade One in school as part of the adult literacy program, but she said that she does not know how to write or read.

¹⁶⁶ The *'izba* is a piece of land normally surrounded by a fence, where the Bedouin community in the study area normally keep their camels and goats. It could be close to where they live or a bit farther in the sandy inland. It has been an important aspect of the lives of many Bedouin families in the area; part of their daily routine, work, or leisure activities.

outside world and to other dialect varieties via schooling and media, due to low mobility, the lack of formal schooling, and the nature of their demographics (older, women, al-Tharmad locality) have provided ideal conditions for these women to maintain the traditional features, which is not a new pattern in this part of the world.

On the other hand, Jum'a (J26MTH) who showed the highest proportion of [gʲ] (90%) is a 26-year-old man. He was born in al-Bārda village as well and studied in local public schools in the area; he is a high school graduate. He found a job in Muscat and at the time of the interview he commuted to work on a daily basis. The interview with this informant was shorter than most of the interviews and was conducted by a male assistant who is an acquaintance of the informant and who happens to speak a sedentary type of dialect. It is noticeable that throughout the interview, Jum'a appeared to be (linguistically) reserved and formal, probably due to the interview procedure itself (the observer's paradox; Labov, 1972), or due to the fact that he is addressing a non-Bedouin informant, which could explain the high amount of variation and accommodation in his speech. His father is a camel breeder, so the young man mainly talked about the Bedouin and camel culture. Some of his tokens can be considered as core dialectal words in common use, yet the speaker consistently uses the incoming variant with such words. In a follow up to the interview, this informant mentioned that the way he talks to his relatives and *Sa'di* friends is different than the way he talks with others outside the tribe. This may help explain the proportionately high occurrence of [gʲ] in Jum'a's data; interestingly enough, his father (AJ57MTH) is the highest user of [gʲ] in the data from the older age group for this locality.

The younger speaker with the lowest proportion of [gʲ] amongst the informants in al-Batha locality and amongst the younger age group in the whole sample is Hamdan (H25MB; [gʲ]=12%). Hamdan studied at a local school, but immediately started to work outside the local area. Hamdan is relatively conservative in his speech which he sees as a key part of his

Bedouin ‘Suwaiqi’ identity, but which I inferred is also related to Hamdan’s involvement in the local Bedouin and camel culture; this behaviour is not very surprising for Hamdan who is also the most linguistically conservative in his age group with regard to the use of the traditional null definite article (more in § 7.2.2 in Chapter Seven).¹⁶⁷

Figure 5.8 shows the results of the use of [gʲ] per speaker divided by age group. It shows that speakers within the three age groups vary in their use of the incoming variant, but it also shows more density and gradience within the middle age group in general. If we take the 30% as a point of comparison between the three age groups, 67% of the informants in the middle age group scored above or right below 30% (N=10/15) compared to the other two age groups; this means that more than half of the informants for the middle age group scored above the overall average for the entire dataset.

¹⁶⁷ The interviewer in a sedentary dialect male speaker who is an acquaintance of Hamdan.

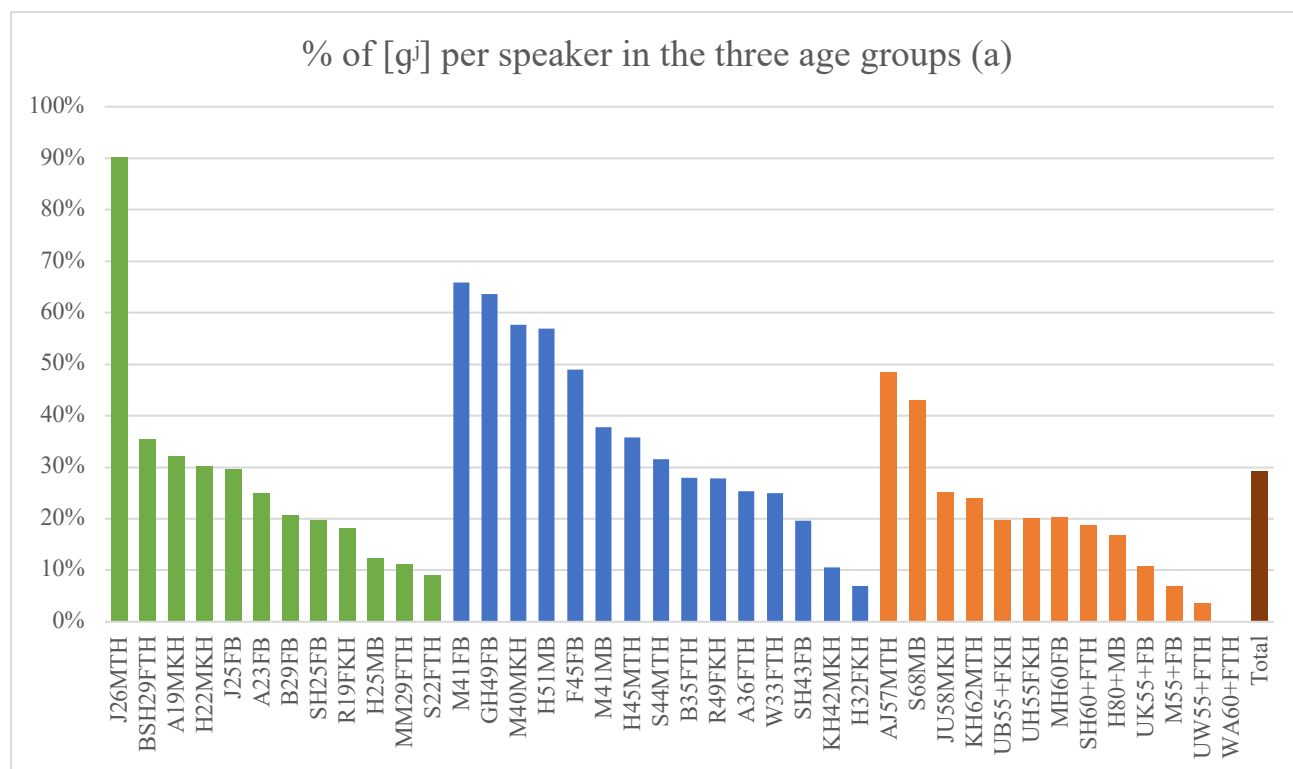


Figure 5.8 Use of [gʲ] per speaker in the three age groups (a); N=2,510.

It is also noteworthy that the younger male group in al-Tharmad locality is a one-informant cell, which means that it could have skewed Rbrul results for the external predictors in that this one informant is bringing the whole group's average up; so, Jum'a's tokens were excluded from the dataset and another run based on the Rbrul model presented in Table 5.3 was done (see Table 5.7; continued overleaf).¹⁶⁸ Rbrul results for the younger age group are a bit different; with a FW=0.46, the younger age group disfavours the application value [gʲ]; the informants in this group approximate the informants in the older age group in their use of [gʲ]; the gap between the proportion of [gʲ] use in the younger and middle age group is much higher now. These results suggest that the middle age group is generally more advanced in the use of [gʲ] than the other two age groups.¹⁶⁹

¹⁶⁸ Although the younger age group of al-Batha men is also a one-informant cell, this data is not excluded.

¹⁶⁹ Locality and stress are selected as statistically insignificant in this run as well.

Application value= [gʲ]; N=2,450; overall proportion= 26.8%				
R ² = 0.305; log likelihood= -1196.817				
Predictor	Log odd	FW	N	% of [gʲ]
Preceding environment (p<0.0001) (p=1.92e-23)				
Coronal	0.699	0.668	796	33%
Back vowel	0.542	0.632	620	32.6%
Front vowel	0.294	0.573	299	33.8%
Labial	0.269	0.567	172	22.1%
Pause	-0.131	0.467	135	17.8%
Bank consonant	-0.209	0.448	50	20%
Palatal	-1.463	0.188	378	5%
Number of syllables (p<0.0001) (p=5.53e-20)				
Polysyllabic	0.771	0.684	904	39%
Disyllabic	-0.108	0.473	1224	21.4%
Monosyllabic	-0.662	0.34	322	13%
Following environment (p<0.0001) (p=3.18e-16)				
Consonant	0.461	0.613	267	42.3%
Back vowel	0.186	0.546	1059	31.1%
Front vowel	-0.647	0.344	1124	19.1%
Age (p<0.0001) (p=3.9e-15)				
Middle	0.528	0.629	941	35.9%
Younger	-0.160	0.46	612	22.1%
Older	-0.368	0.409	897	20.5%
Gender (p<0.0001) (p=7.43e-06)				
Male	0.231	0.558	920	32.6%
Female	-0.231	0.442	1530	23.3%
Syllable weight (p<0.001) (p=0.000134)				
Light	0.268	0.567	1681	30.8%
Heavy	-0.268	0.433	769	18.2%

Syllable type ($p < 0.001$) ($p = 0.00364$)				
closed	0.18	0.545	1560	27.6%
open	-0.18	0.455	890	25.4%

Table 5.7 Results for the use of [gʲ] in the dataset; J26MTH excluded.

Figure 5.9 presents the picture per individual speakers after excluding Jum'a's tokens. The speaker with highest proportion of [gʲ] belongs to the middle age group. This informant is Moza (M41FB), a forty-one-year-old woman who works as a local educational supervisor, and who previously held a teacher's and a headteacher's positions mainly in local schools; she did her undergraduate degree in the UAE and her master's degree in Jordan. She also showed a high percentage of the affricate variant [dʒ] (28%; N=26 out of N=92 tokens altogether originally extracted from her interview).

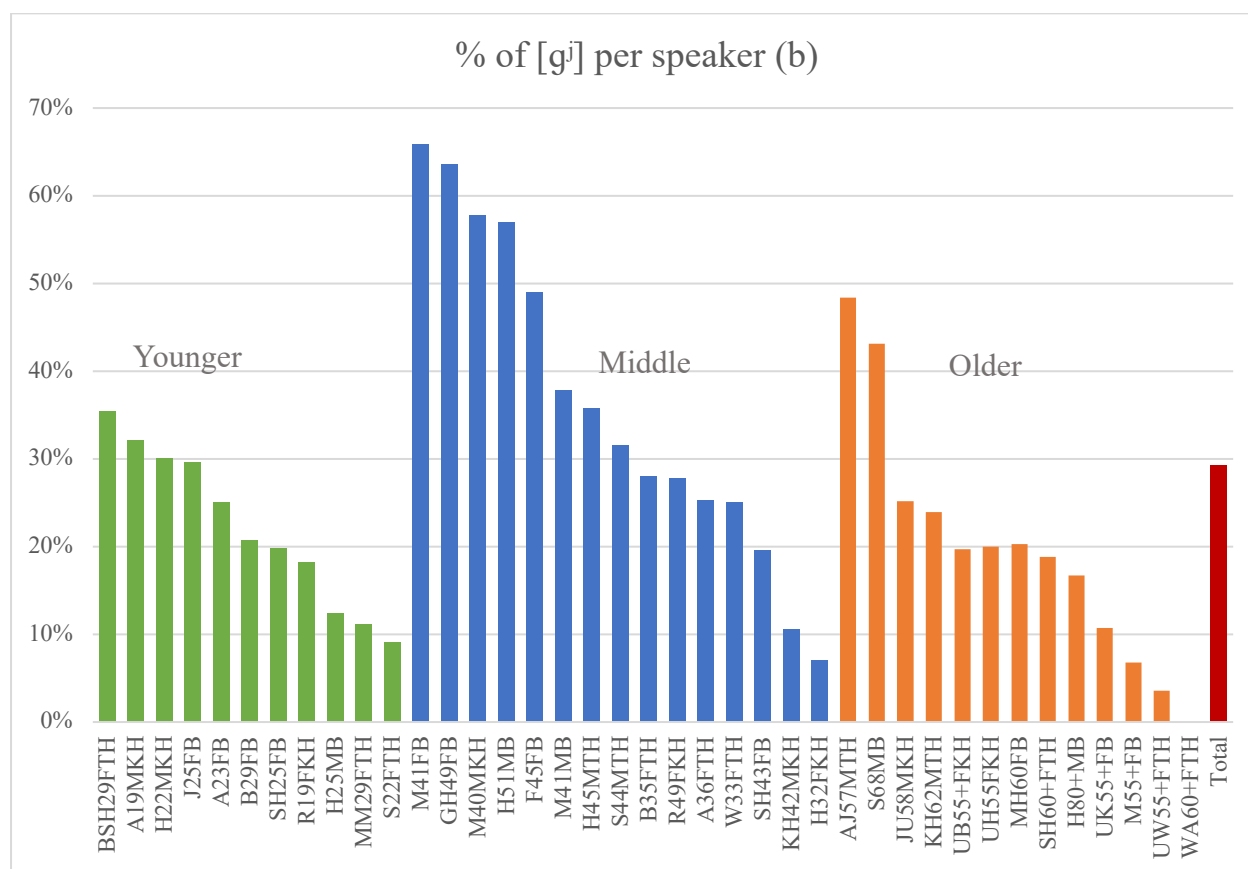


Figure 5.9 Use of [gʲ] per speaker in the three age groups (b); N=2,510; J26MTH excluded.

When reflecting on the (dʒ) variable in her speech, she said:

b-ħukm l-ʕamal [...] tiba tintʕaq sʕaħħ, w lamma tsa:fir jaʕni maθalan ana fil-urdun jaʕni mumkin ħad jʕalliq ʕalajji iða: gilt gʕi:m [...] ma: bayfahmin jaʕni maθalan gʕami:la-laʔ dʒami:la, jaʕni la:zim atkallam miθilhum kint, fa ʕwajji l-lisa:n jaʕni t-tiʔaθθar

‘Due to the nature of the job, you want to pronounce [/dʒ/] correctly, and when you travel, I mean, for instance somebody may be critical that I said *gʕim* [i.e., [gʕ]] [...] they would not understand, I mean, for instance /gʕami:la/– no, /dʒami:la/ [is the right way to say it],¹⁷⁰ I mean I had to speak like them, so, your speech must change slightly, meaning it would be affected.’ (M41FB)

The overall picture for the external predictors

Results on the external predictors show that age and gender are selected as statically significant, with the middle age group using [gʕ] more consistently than the younger and older age groups which approximate each other in their use of this variant. With regard to locality which is not selected as statistically significant in predicting the variation, the proportion of [gʕ] in al-Batha is found to be relatively higher than that of al-Tharmad and al-Khadhra. This result is compatible with the demographic description of this locality being the centre of business and administrative establishments in al-Suwaiq; it is more diverse in the demographic make-up than the other two localities as explained in § 3.2.3 in Chapter Three. In addition, results from the cross tabulations between the different predictors also reveal intricate variation patterns that are not necessarily fitting the general trend. Although the middle and young age groups are found to generally use more [gʕ], the cross tabulation between age, gender, and locality (Table 5.6) shows that older men use more [gʕ] than women

¹⁷⁰ /dʒami:la/ is a proper woman’s name; also is an adjective ‘beautiful’.

in the three age groups in al-Kadhra and al-Tharmad, but not in al-Batha were women in the middle age group use [gʲ] the most. The distribution of [gʲ] per the individual informants (Figure 5.7 and Figure 5.8), although shows that there is comparable amount of variation across the whole sample with more density and concentration in the middle age group, it reveals that, within the respective age and gender groupings, the use of [gʲ] is not necessarily consistent or predictable.

The patterns with regard to the different age groupings for men and women (Figure 5.4, Figure 5.5) can be explained in the light of the nature of contact for these groups. The older women group is the least mobile, both geographically and socially. These women are housewives. None of the older women are literate, apart from one from al-Batha, Maryam (MH60FB; [gʲ]=20%) who finished high school and currently lives with her family; three had very basic formal education as part of adult literacy programmes, but they can be considered functionally illiterate since they can barely read or write. Some of them got married within the locality, yet others got married to other *Saʻdi* men in villages a bit far from their immediate place of residence in the other two localities. It is interesting though that the fact that Maryam is literate unlike the rest of the older women does not seem to offer an explanation for the variation in the use of [gʲ] within the older women group, since three other women behaved similarly: UB55FKH ([gʲ]=20%), UH55FKH ([gʲ]=20%), and SH60+FTH ([gʲ]=20%). Leaving Maryam aside, what sets the latter three women apart from the other four in the older women group who showed less variation is that the former got married and settled outside their localities, i.e., there is concurrent change in the type of social contact that is most likely facilitated by the change in the demographics of the place they live in.

On the other hand, the older men group consists of informants who spent some time outside Oman mainly in the Gulf, with the exception of Hamid (H80+MB; [gʲ]=17%) whose

work is based in the same area where he lives.¹⁷¹ All but one of these men are retired from their ‘formal’ jobs,¹⁷² some are running private businesses, but all are based at their immediate localities. The fact that the older *Sa‘di* men are or have been more mobile within and outside the study area, and due to the nature of their social contacts which includes other non-Bedouin and expatriates on a frequent basis. This background could help better explain their higher use of [gʲ]. Saeed (S68MB; [gʲ]=43%) is probably the most geographically mobile and socially connected of the older men group due the nature of his previous job and educational background and his current status in the speech community; as a young boy, he travelled to the UAE where he got educated in a military college but only for some time and then worked at the police force for a while before he came back to Oman where he was in charge of many, some high ranking, administrative positions in different parts of Oman before he retired back to his current village within the vicinity of al-Batha locality.¹⁷³ Such background which is characterised by high geographical and social mobility could, generally-speaking, help explain the relatively higher proportion of [gʲ] for older men in general. This mobility means that the informants have a bigger chance of meeting other people who use a stop variant for the glide, be it an affricate, i.e., like the one used in some parts of the Emirates and the Gulf, or different variants of the velar stop, like the ones used in different parts in Oman; the fact that the affricate variant may have been used in places where my older male informants have worked in, although did not necessarily entail its adoption by these older men,¹⁷⁴ it probably made them ‘socially aware’ of other non-glide variants.¹⁷⁵

¹⁷¹ This informant on the other hand is nearly categorical in the use of the ‘incoming’ overt article (see § 7.2 in Chapter Seven).

¹⁷² Hamid has never been in a formal kind of job in his life due to the nature of his status as a tribal leader. His job is basically based in his place of residence where he manages the tribal affairs and runs the family business.

¹⁷³ Saeed is also nearly categorical in the use of the ‘incoming’ overt article (see § 7.2.2 in Chapter Seven).

¹⁷⁴ This is supported by my data. Only two affricate tokens were found in the older men group, precisely in the speech of a sixty-two-year-old man, who also spent some time working in the police force in the UAE.

¹⁷⁵ In light of this argument, we can further assume that the higher occurrence of [gʲ] is in response to my informants’ exposure to the affricate variant itself as well; the former can be seen as an approximation to

Again, this is not necessarily the case with older *Sa'di* women who, generally-speaking, mostly connect with their immediate families and neighbours in their immediate villages.

In addition, social and geographic mobility could readily explain the variation patterns observed with regard to the older age group. But in order to explain the overall lower proportion of [gʲ] for the older age group in comparison to the younger and middle age groups, and in order to understand the differences and similarities in the linguistic behaviour of men and women in the latter two groups, we need to go beyond demographics and work patterns as explaining factors for the variation at hand. In this case, education comes to the picture. Education is what generally sets the older age group apart from the middle and younger age groups in the sample. One of the older women commented that in addition to contact between speakers who speak different dialect types, education plays a major role in the variation in some distinctive features of the dialect, since people may become linguistically aware of the dissimilarities between their dialect and the language in school; she says that:

xalli intaj irtʰtiba:tʰ l-gʲi:ra wil-mugʲtamaʃ, al-ħi:n akθar ʃaj illi
 ʃajjar l-lahagʲa:t, ʃajjiratti d-dira:si [...] lanni l-kʲta:b maktu:b
 fi:h ʃagʲara ma: fi:h ʃjara [...]

‘leave aside the connection [with] the neighbours and the
 community; nowadays the thing that has most changed the
 dialects is education/schooling...because the book says /ʃagʲara/
 not /ʃjara/ [i.e., tree]’ (UH55KH)

All of my middle age group informants except for one have finished and studied beyond high school. At the time of the interview, eight were schoolteachers, four were working as local educational supervisors with previous teaching experience, one is a college

the latter, since there is no affricate phonemes nor allophones in the consonantal inventory of these speakers.

administrative, one is a lawyer, and one is a stay home mother. In contrast, all of my younger age group informants finished high school; four were doing their undergraduate studies at the time of the interview, six had an undergraduate degree (two were seeking jobs, three were working as teachers in school, one as a college lecturer and one was doing her master's degree at the time), and two worked right after high school. Some of the middle age group speakers went abroad for their undergraduate and/or graduate degrees, whereas some did their degrees in different places in Oman but mainly in Muscat.¹⁷⁶ The younger speakers did their graduate studies in Oman, the UAE but also abroad.¹⁷⁷ Education adds another dimension to the contact circle of the speakers in the middle and the younger groups that is not generally accessible to the older age group. The informants have been exposed to the different varieties or languages through their educational years, and then for many of them through the nature of their work.

The implications for education as a measure of contact for these speakers is apparent in the relatively higher use of [gʲ]. However, it can be argued that the difference between the younger and middle age groups is that the type of schooling they have been exposed to as young girls and boys is different in nature (see § 3.2.3 in Chapter Three). When most of the informants in my middle age group joined schools, there was a limited number of children to join elementary (and secondary) schools in their immediate localities which to begin with did not necessarily have schools, so they commuted to other few schools in the study area where students from different dialectal backgrounds attend (Bedouin, sedentary, and coastal¹⁷⁸). There were also very few Omani teachers; most of the teachers were Arabic-speaking

¹⁷⁶ UAE, Jordan, Morocco, India.

¹⁷⁷ Three were doing their undergraduate degrees in the US and the UK at the time.

¹⁷⁸ My *Sa'di* informants are contended that the communities living at the immediate adjacency of the coastal line in the study area are not 'badu' (i.e., of a Bedouin descent), but 'hal s-sāhil' (i.e., people of the coast); see § 3.3. Although I have not analysed the differences between the speech of the two groups, to my knowledge, the dialect of these families can be classified as a B-type dialect.

expatriates from the Levant, Egypt, Sudan, and North Africa.¹⁷⁹ On the other hand, the younger informants were mostly born in the 1990s at a time when an educational reform in the whole of Oman was taking place. When they joined school, there were many more elementary schools in the study area, and the student backgrounds were much more homogenous in nature. The number of local and non-local Omani teachers have increased considerably in comparison to that of expatriate teachers. We can see that although both the middle and the younger age groups have had local education at school, the nature of schooling is different. So, we can tentatively say that the findings for the two age groups are compatible with the previous assumptions that the nature of schooling would influence the variation patterns: the middle age group has been schooled in a heterogenous context and shows more variation than the younger age group which has been schooled in a more or less homogenous environment. This also could be explained in terms of the social and linguistic pressure the informants in the middle age group may have faced as Bedouin students in school due to the linguistic diversity of the school considering that their variant is quite phonemically different from the (dʒ) variants of their teachers and their peers, as opposed to the younger age group informants who probably did not have to deal with this pressure considering that their teachers are mostly local and mostly Omani, and their peers are their family and neighbours, especially within the elementary stage of their schooling. In addition, although the results for the individual speakers do not necessarily fit in within this conclusion, this can help in explaining the overall linguistic behaviour of these age groups.

Furthermore, although the middle and younger age groups show more variation than the older age group, the middle age group advances in their use of [gʲ], most men and women in the middle age group score above the 28% average for the whole sample. The informants who showed the highest use of [gʲ] are men and women, aged 40-51, and except for Jum‘a

¹⁷⁹ English was mostly taught by Indian teachers.

(J26MTH; [gʲ]=90%; al-Tharmad locality) and Mohammad (M40MKH; [gʲ]=58%; al-Khadhra locality), they are mainly from al-Batha locality. One point that should probably be highlighted here is the fact that the sample in the middle age group is not very diverse, since I have data from informants who are educated beyond high school; one woman only (R49FKH; [gʲ]=28%) in the middle age group has not continued her schooling.¹⁸⁰ This probably means that the role of education-induced contact cannot be objectively examined in this group; this is more or less also true for the younger age group in the sample.

5.5 Summary and conclusions

This chapter presented the case of (dʒ) as a sociolinguistic variable in the *Yāl Saʿad* dialect as spoken in al-Suwaiq and al-Miṣinʿa towns. In terms of the internal predictors, the preceding and following environments, number of syllables, syllabic weight, and syllabic type are selected as significant in predicting the variation. Polysyllabic words, a preceding coronal and a back vowel, and a following consonant mostly favour the incoming variant [gʲ], whereas monosyllabic words, a preceding palatal glide, and a following front vowel mostly disfavour it. In terms of the external predictors, age and gender are selected as statistically significant, with men favouring the use of [gʲ], and the middle age group mostly favouring it amongst the three age groups. In addition, even though locality is not selected as statistically significant in predicting the variation, the use of [gʲ] in al-Batha locality is found to be higher than the use of this variant in the other two, namely al-Tharmad and al-Khadhra. The crosstab between age and gender shows that the difference in the use of [gʲ] between men and women

¹⁸⁰ This informant left school after grade four; she got married and left Oman to live in the UAE while her husband worked there for 15 years before they settled back in the area. She is also the one who used the velar [g] variant of (dʒ), although I cannot quite explain such occurrences in her data.

in the older and younger group is higher than that in the middle age group, and that the women in the middle age group approximate men in their use of this variant.

The status of (dʒ) in the *Yāl Saʿad* community

This research has revealed that the glide variant of (dʒ) is quite salient. It is a salient feature of this Bedouin variety, in the same manner [j] is salient for many other Bedouin communities in Oman and the Gulf. Generally speaking, within the study area, it holds ‘local prestige’ by virtue of it being the variant of the majority, and the variant of the most dominant tribe(s). In other words, it is not associated with a certain Bedouin tribe or a certain locality, it simply is a distinctive feature of the Bedouin variety spoken there. But there are non-Bedouin tribes and families that live in the area who do not use [j], and who have other realisations for Arabic /dʒ/, mainly [gʲ] and [g]. In addition, although [j] is the traditional dialectal variant of /dʒ/ for the Bedouin coastal tribes along the Bāṭina coast, [gʲ] is also characteristic of the Bedouins’ speech in this area, since it is mainly specialised for the lexical domains (e.g., Arabic loan words) where in other Gulf speech communities, for example, the affricate [dʒ] is used. The MSA-like form [dʒ] barely occurred for these items in the whole dataset, tempting me to infer that it is not available in the phonemic inventory for most of the informants to begin with. In the few cases when some informants shifted their style (e.g., when they spoke formally or code-switched to the standard variety), they almost consistently used [gʲ] not the MSA-like form [dʒ]. So, we can infer, at least for this part of Oman, it seems that [gʲ] as a variant of (dʒ) has become some sort of a ‘local standard’. In addition, [gʲ] is analogous to the variant resulting from the conditioned *q>/g/>[gʲ] fronting in this variety, and is thus, already accessible to the speakers. This also means that [gʲ] for (dʒ) has probably co-existed with [j] for a long time in this area.

[j] ~ [gʲ] variation

There are a number of scenarios for why [j] is currently in variation with [gʲ] in this speech community. The (dʒ) as a variable in the study area is probably becoming a ‘social marker’ (Labov: 2001). When commenting on the [j] variant in their variety, the *Sa’di* informants seemed to be aware that [j] is one of the features that makes their dialect stick out and is subject to outsiders’ commentary; this is probably what renders it subject to conscious or unconscious change on the part of the speakers in favour of what they perceive as a much more common, more accessible or ‘easy to understand’ form. Some informants stated that the variation in their speech in general is induced by their attempt to ‘simplify’ their speech or to ‘accommodate’ to the interviewer, considering the fact the interviewer does not speak their variety but, in this case, a sedentary one.¹⁸¹

So, the readily available explanations for the seemingly increasing awareness of the status of [j] in this speech community are the increasing exposure to other varieties (local and non-local) facilitated by geographical and social mobility through work, education, the change in the demographics of the place and marriage patterns, some of which have been commented on in the interviews with the informants, and thus, the increase in the use of [gʲ] being primarily triggered by the informants’ need to simplify their speech for ease of communication is a possible explanation. However, we also should remember that the overall variation is relatively low which emphasises the salience of [j] as a variant of (dʒ). In other words, the variation in (dʒ) in this speech community is most likely a case of short-term accommodation or stable variation (Gardiner and Nagi, 2017), not change in progress per se, especially that the older and the younger age groups are more or less behaving similarly with

¹⁸¹ A woman in the older age group (UH55FKH) described my dialect as ‘lahgʷat s-sūg’, i.e., the dialect of the speech community living in the vicinity of the main market of al-Suwaiq; again, this dialect is sedentary in type (see § 3.3); the main realisation of etymological *j in this dialect is /gʲ/.

regard to the variation at hand; however, since this is the first quantitative investigation of (dʒ) in this area, it would be hard to confidently argue for the latter point.

The future of [j]

Again, the glide variant's local prestige is reinforced by the fact that it is the variant of the majority; the dialect spoken by the *Yāl Sa'ad* tribe is more or less representative of what is locally known as 'lahg^yat hal swēg^y' (literally: the dialect of the people of al-Suwaiq) to insiders, as well as to outsiders in Oman.¹⁸² The glide's local prestige is probably reinforced by that of its counterpart in the neighbouring Gulf states, in which the glide variant of (dʒ) entertains a high status, considering that it is part and parcel of the emerging regional Gulf variety (§ 4.4 in Chapter Four). In addition, to the Bedouin community in this area, this dialect indexes their Bedouin identity (Eckert, 2008; Eckert and Labov, 2017), and to many of the younger informants, the dialect is the only means by which they hang on to their 'Bedouinness', considering that they do not necessarily adhere to the 'traditional' Bedouin lifestyle anymore. Indeed, the existing ideology that maintaining the local speech is a key measure for how much of a 'Bedouin' a person is, is a recurrent theme expressed by several informants interviewed. These factors, at least for the time being and the near future, would insure a maintenance of a variant like [j]. One younger informant responded to my question on whether she anticipates this variant to change any time soon by saying:

ha:ði laʔ; ha:ði wa:jid 'strong', ha:ði jaʕni sʕaʕba inha tityajjar

'This one, no; this one [i.e., [j] for (dʒ)] is very strong, I mean this one is hard to change.' (R19FKH)

¹⁸² This variety is getting attention in social media; certain features of the dialect are the focus or the subject of exaggeration, including the [-jih] ~ [-jeh] 1SG.ACC/GEN suffix.

The innovative variant [dʒ]

Whatever the situation with [j] ~ [gʲ] may be, it seems that it is not new within the context of this study, mainly considering the variability in the speech of the older age group. What we need to keep an eye on for the future is the affricate variant (in spite of the very low occurrence in the current dataset (2.6%)).¹⁸³ The affricate is an innovative variant that seems to be spreading over the domains of both of other two main variants [j] and [gʲ]; this variation seems to be led by educated speakers who have done undergraduate or postgraduate degrees outside Oman probably due to the nature of contact they have been exposed to during their studies. So, one could only keep an eye on the trajectory of variation and change in (dʒ) in this dialect in the future, considering the range of possible forms for a single phoneme in a context as fast-changing and multi-dialectal as this part of Arabia.

¹⁸³ This percentage represents instances where the affricate varies with the other two variants, namely [j] and [gʲ]; i.e., instances of the variable which are part of the envelope of variation and were included in the Rbrul analysis (N=69), after the exclusion of specialised and borrowed words which are never realised with the glide variant like /dʒa:mʕa/ ‘university’ (N=118 in all of the interview data; more concentrated in the speech of three informants, one young woman and a man and a woman in the middle age group who happen to have studied outside Oman; this number is still comparatively low to other excluded categorical [gʲ] words from the Rbrul Excel sheet). In either case though, instances of the affricate are relatively low compared to the other two variants, emphasising the main conclusion that [gʲ] seems to have become a local standard, the main variant for MSA-like forms and for formal settings, and that [dʒ] is the new incoming variant. To further contextualise this last sentence, I have analysed data from an interview with an older man which was excluded from the analysis because of the high level of formality in the interviewee’s speech. This man was 55-year-old high ranking official in Oman who is a PhD holder, is settled in al-Batha locality, and was working outside the study area at the time of the interview. Listening to the first six minutes of the interview, I found a lot of direct code switching to MSA and a formal tone in general. There were 44 tokens with etymological *j; however, the only word that was realised with the affricate is /al-dʒa.miʕa/ ‘DEF-university’ which can be considered as an instance of a direct code-switching to MSA. All of the other tokens were realised with [gʲ] including the ones in parts where there is also direct code-switching to MSA. Later in the interview, the speaker recited some Quranic verses in which the word for heaven is also realised with [gʲ], not the affricate: /gʲannat-in/ (cf. /dʒannat-in/). This probably can be also used to argue for the status of the [gʲ] variant in this speech community as not just the supralocal but also the ‘standard’ form for these speakers.

Chapter Six: Literature on the definite article

This chapter presents the background for the analysis of the second linguistic variable, namely the definite article variable (DEF). Following a brief overview of definiteness in Arabic, and the Arabic definite article, the chapter presents a general description of definiteness in the dialect under study, with a particular emphasis on the null definite article.

6.1 Definiteness in Arabic

In human language, definiteness is generally a semantic concept; a notion that has to do with human cognition and the speaker's perception of referentiality of a given nominal entity. This semantic/pragmatic notion is cross-linguistic, though the specific grammar different languages employ to mark a DEF category vary. Many languages have 'overt' marking of definiteness, commonly some form of a definite article which may be a separate word like the English 'the' or an affixed element like the Arabic /l-/. Other languages lack such means for marking the DEF category and use other grammatical means for expressing it. The fact that languages express this category in various 'possible' ways in their respective grammars "underscores the susceptibility of definiteness marking to variation across time and space" (Turner, 2013: 1).

A recent treatment of definiteness in Arabic is Turner's PhD thesis (2018) which aims at a comparative examination of the grammatical variation in the definiteness systems of six Arabic varieties, namely Ḥassāniya, Moroccan, rural Sudanese, Levantine, San'āni, and Central Asian. The study also provides a preliminary model for the typological classification of the treatment of definiteness in spoken Arabic. Turner's main stance on this topic is that the traditional basic binary distribution, by which Arabic nouns with the definite article are

described as ‘definite’, whereas other nouns without this article are labelled as ‘indefinite’, is an oversimplification; the story of definiteness and the definite article is rather ‘complex’, and a simple dichotomy does not account for the ‘scaler’ nature of definiteness as a ‘multi-tiered’ semantic notion with varied ‘referential values’ (ibid: 1). Turner further argues that this scaler nature of definiteness along with the diversity in the treatment of definiteness in Arabic varieties in general can foster the view that this category is a ‘primary site’ for linguistic variation and innovation (ibid). This scaler aspect of definiteness in Arabic is discussed in Brustad’s comparative work on the syntax of spoken Arabic (2000); she holds a similar view on the inadequacy of a dichotomous classification of the definiteness states in Arabic. Brustad sees definiteness in Arabic as a continuum of semantic values with definite and indefinite at the two poles (ibid: 18). Definiteness overlaps with or employs other semantic concepts such as individuation, animacy, humanness, agency, referentiality, specificity, genericness, textual/physical prominence, qualification, quantification (Brustad, 2000; Turner, 2013; 2018 and the references therein), amplification (Beeston, 1970: 65), familiarity, and identification (Lyons, 1999: 2).¹⁸⁴

The speaker-listener/addressee along with an amalgamation of the former notions play a role when assigning the different states of definiteness in Arabic. Generally speaking, there are three definiteness states of the Arabic noun: the definite, the indefinite, and the construct ‘annexed’ state (e.g., Fischer, 1997: 159 for Classical Arabic). Within the definite state, there are three types of reference associated with prefixing the article /l-/ to a certain entity, namely, the referential definite,¹⁸⁵ the discursive definite,¹⁸⁶ and the generic¹⁸⁶ definite. The first one is

¹⁸⁴ Danon (2002), with Hebrew as a case study, for definiteness (DEF) as a formal ~ syntactic ‘feature’ which may or may not coincide with semantic definiteness, where definite DPs can be formally described independently from their semantics, due to the “definiteness agreement [in the noun-attribute strings], and the appearance of the object marker in front of definites only’ (ibid: 37).

¹⁸⁵ Known to Arab grammarians as *lām l-‘ahd* ‘article of familiarity’ (Sībawayh in Marogy, 2010: 101).

¹⁸⁶ Known to Arab grammarians as *lām l-jins* ‘the generic *lām*; i.e., *l-’* (ibid).

anaphoric because it refers to some entity previously established in the discourse;¹⁸⁷ the second refers to an entity which is definite due to its immediate relevancy to the discourse context or a common knowledge shared by the interlocutors, whereas the third is ‘generically defined’ due to its applicability to the whole category or ‘genus’ of a given entity (Sībawayh in Marogy, 2010: 101; also in Beeston, 1970: 65), e.g., Arabic generic nouns in SUB position:

- 1) r-ra:ʕi j-irʕa l-yanam
 DEF-shepherd.SGM 3-tend to.IPFV.SGM DEF-sheep.PL
 ‘Shepherds tend to sheep.’ (Adapted from Manfredi, 2017: 209).

The indefinite state is generally marked with a final /-n/ attached to the noun stems in the standard variety; on the other hand, vernacular Arabic generally has zero marking for this variety apart from dialectal nunation ‘*tanwīn*’ in some current Bedouin varieties which, nonetheless, is quite restricted in distribution (e.g., Holes, 2004 on eastern and south-eastern Arabia). The vernaculars also have innovated some indefinite markers, such as *fad* ~ *fard*, *wāhid* ~ *wahd el-*, and *ši*, to denote different shades of indefiniteness.¹⁸⁸ It is noteworthy that, in this study, I refer to nouns or noun phrases with the overt article /l-/ as simply ‘definite’. In doing so, I am not taking a stance with regard to the status of the overt article as *the* definite article, simply because the current research highlights an extra possible arrangement for this category, namely the null definite category, as we will soon see, nor am I taking a position with regard to the works that propose alternative models for (in)definiteness (e.g., Turner, 2013; Danon, 2002); this issue is beyond the scope of the current research.

¹⁸⁷ This type is described as ‘particularized’ definition by Beeston due to the reference to a particular individual entity unambiguously established by the context of discourse as opposed to the ‘generalizing’ function of the definite article (1970: 37).

¹⁸⁸ For more on these markers: Brustad, 2000 and Turner, 2018 and the references therein.

Furthermore, in addition to nouns with a prefixed definite article, there are four other forms of the definite noun mentioned in traditional Arab grammarian descriptions and in modern Arabic linguistics, namely the inherently definite proper nouns, pronominals, and demonstrative pronouns, as well as nouns made definite by annexation as in the construct state (e.g., Sībawayh in Marogy, 2010: 101-2; Holes, 2004: 186-198; Beeston, 1970: 36). In terms of definiteness, referential nouns with the prefixed definite article are lower in rank than inherently definite proper nouns and personal pronouns, but higher than the non-referential indefinite (Croft 1999, via Brustad, 2000: 22). Common nouns, then, are inherently not definite; they can be made grammatically ~ syntactically definite by:

1. Adding a pronominal suffix;
2. Adding the definite article;
3. Juxtaposing two or more nouns, as in the construct state (Ryding, 2005: 156 for MSA); the first noun becomes definite when added to another definite noun or noun phrase (ibid).

6.2 The Arabic definite article: socio-historical background

Having highlighted some concepts and arguments on definiteness in Arabic, this section elaborates on the overt definite article in Arabic, particularly in terms of the form and the morphosyntax both within Arabic and the larger Semitic context.

6.2.1 The origin of the definite article

Rubin states that based on the fact that Akkadian and Classical Ethiopic did not have a definite article, proto-Semitic ‘almost certainly’ did not have any article to begin with and that each of the descendant Semitic languages treated the issue of definiteness independently

(2005: 65). Rubin further posits that the definite article is a feature of Central Semitic, or a feature that has ‘developed independently’ in different Central Semitic languages (ibid). Arabic is part of this language group which also includes Old South Arabian, and Northwest Semitic languages amongst which descendants are Aramaic and Hebrew (See Figure 6.1). In addition, the definite article has been attested for some modern Ethiopic and Modern South Arabian varieties which do not belong to Central Semitic, but rather the West Semitic branch (ibid).

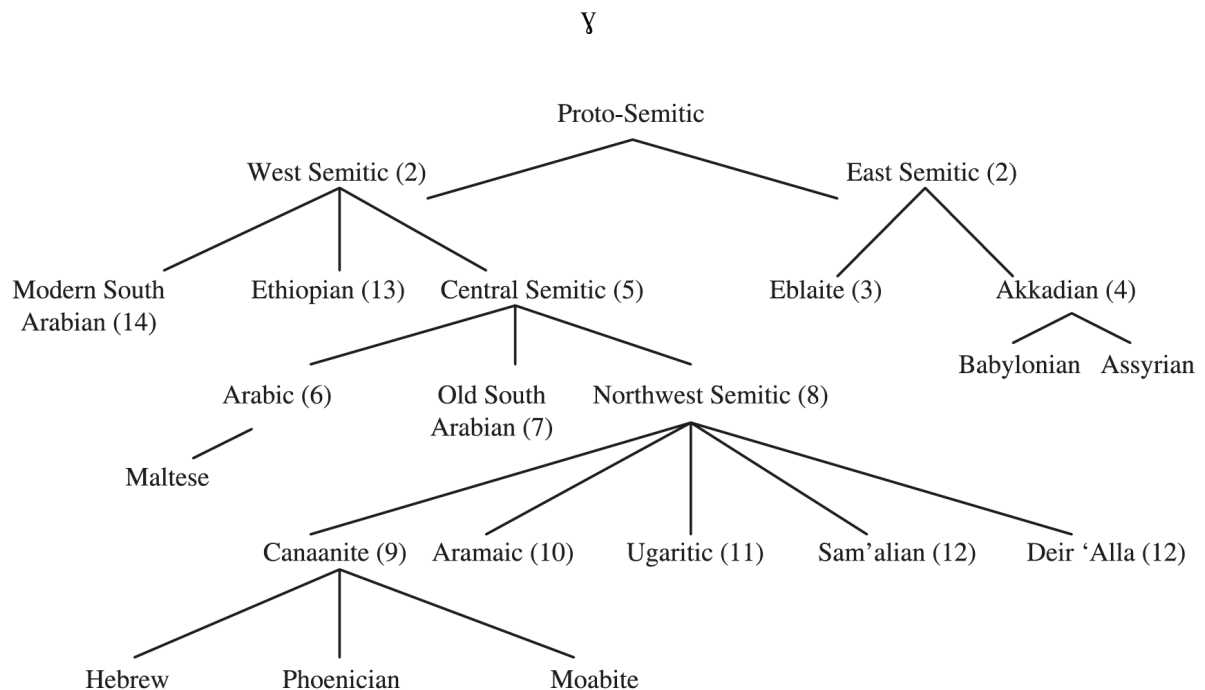


Figure 6.1 The subgrouping of the Semitic language family (Rubin, 2008: 62).

In terms of the article’s diachronic development, Rubin maintains that the origin of the Arabic definite article is the proto-form *’al- which is derived from the proto-West Semitic ‘plural’ far demonstrative *ulli- ‘that, those’ (ibid: 72-76). However, Pat-El (2009) disagrees with this proposition and argues that the definite article has not developed from an

older demonstrative. One reason is that the demonstrative-noun order is a ‘late feature’, at least later than proto-Central Semitic (ibid: 21). This means that there is another origin for the prefix /l-/ in Arabic. She argues that the article first was attached to attributive adjectives, and then it was transferred to the modified noun. One supporting argument for this is the fact that the definite article in Central Semitic allows the nominalisation of adjectives, e.g., /li-s^ʕya:r.N/ ‘the younger ones’.¹⁸⁹ In response to the argument that the Arabic definite article is a nominal prefix to begin with because it is a marker of nominal agreement, and thus the modifying adjective exhibits the same agreement marking, she argues that it is not always the case that both forms in agreement with each other are similar in form, or exhibit the same agreement marking (ibid: 26). Another supporting argument for Pat-El’s claim is that if the definite article was originally used to mark definiteness in the noun and then transferred to modifying adjectives, then why do the possessive suffixes which also mark the noun they attach to as ‘definite’ not attach to the modifying adjectives as well? (ibid: 25-26).

6.2.2 The form of the definite article

The definite article, [il-] ~ [el-] ~ [al-],¹⁹⁰ is the main means for marking nominal definiteness in Arabic; it is used to mark nominal definiteness in the overwhelming majority of the modern Arabic dialects. The earliest certain attestation of the form of the Arabic article (*a*)l- is in the word /al-ʔilāt/ ‘Alilat; a god of Arabia’ (Retsö, 2013: 435). According to Beeston, this “Arabic marker is a morpheme with positionally determined varieties: lengthening of the initial consonant where this is a lingual [and] a prefixed *l* if the initial consonant is anything else” (1970: 37). So, the article *l*- is phonologically conditioned in the sense that if it is followed by a stem-initial front consonant, more precisely a coronal, it totally assimilates to

¹⁸⁹ This example is elicited from the dialect under study.

¹⁹⁰ The *al*- variant of the definite article is typical of Bedouin Arabic (Kaye and Rosenhouse, 1997: 298).

it. Also, the phonological bulk accompanying the /l/ element of the article itself is subject to different positional constraints, e.g., whether the noun is post-pausal or contextually positioned (e.g., Rubin, 2005: 66 for the forms of the article in CA), and subject to other phonotactics like syllabification constraints (see § 5.3).

As a determiner, the Arabic definite article can be prefixed to nouns as well as attributive adjectives. A related category is improper annexation in which the head or the ‘annexing’ word is an adjective and the annexed is a noun with a prefixed /l-/, e.g., MSA: /hasan-u l-wadzh/ ‘beautiful; lit. beautiful the face’; these are attributive adjective phrases and, as a whole, are syntactically indefinite, unless the whole phrase itself modifies a definite noun and or is marked with the definite article resulting in a definite adjective.¹⁹¹

Adding to the above features, the article does not carry nominal inflection, i.e., gender or number distinctions (Pat-El, 2009: 23). Also, whereas the definite article in MSA is a bound morpheme in that, normally, it is not separated from the noun it defines by other elements, some Arabic dialects, including the one under investigation, allow [definite-cardinal-noun] order; e.g., /l-tle:t wle:d/ ‘DEF-three boy.PL’ (Lebanese Arabic; adapted from Benmamoun and Chouiri, 2013: 127).

Furthermore, in terms of the Arabic dialectological map, the *l*- reflex of the definite article is relatively widespread, but it is not the only reflex; other reflexes, such as the *m*- article do occur. The latter is a west Arabian feature characteristic of some Yemeni and close by varieties like the Tihāmi Qaḥṭāni varieties in Saudi Arabia. An example of the *m*- article as used in the latter dialect variety is /m-waqt/ ‘DEF-time’, and /m-burr/ ‘DEF-wholemeal flour’ (adapted from Al-Qahtani, 2015: 200). Vanhove (2009: 755) mentions an *am*- variant for Mukeyras and *am* ~ *um* variants for Abyan dialects in Yemen. Another related variant is

¹⁹¹ For more on this type of annexation: Ryding (2005: 253); also Goldenberg (1995: 3, as cited in Pat-El, 2009).

the Yemenite *n*- article which is speculated to be a predecessor form of the Arabic article (e.g., Retsö, 1951: 35-36); both *m*- and *n*- articles are believed to relate to close by South Arabian varieties which show a suffixed article *-hn* (ibid).

6.2.3 The syntax of the definite article

According to Huenegard (2005: 185), the definite article in Arabic and Central Semitic appears on the final member in the construct state, e.g., /be:t l-walad/ ‘house DEF-boy; i.e., the boy’s house’. It cannot appear on nouns with suffixes, e.g., /be:t-uh/ ‘house-3SGM.GEN; i.e., his house’, or on proper nouns. In addition, attributive adjectives agree in definiteness with the head noun, e.g., /l-be:t li-k^jbi:r/ ‘DEF-house DEF-big; i.e., the big house’, whereas predicative adjectives, when in conjunction with a definite noun, do not take the article, e.g., /l-be:t k^jbi:r/ ‘DEF-house big; i.e., the house is big’.¹⁹² However, agreement patterns in modern Arabic varieties do not necessarily neatly conform to the above rules; for example, agreement patterns in underlying noun-attribute strings are sometimes violated yielding forms like /ba:b f-ʃardzi/ ‘gate DEF-eastern; lit. the eastern gate’¹⁹³ and /l-ma sxun/ ‘DEF-water hot; i.e., the hot water’.¹⁹⁴ Brustad calls these ‘asymmetrically definite constructions’ (2000: 41). Such constructions are quite abundant in Moroccan Arabic and are not subject to linguistic judgment; their grammaticality is reinforced through the use of such constructions in advertisements and Moroccan Arabic language manuals (Turner, 2013: 47-48). In addition, there are many examples of proper nouns which have an overt article, e.g., toponyms, as in /l-hind/ ‘India’, and personal names, as in /al-bara:ʔ/. We have already seen that such non-

¹⁹² These examples occur in Omani Arabic as used by the author; they are sedentary in type.

¹⁹³ A toponym; Baghdadi Arabic, Holes (2018: 17).

¹⁹⁴ Moroccan Arabic (Turner, 2013: 48).

typical agreement patterns are essential to Pat-El's (2009) hypothesis on the origin of the definite argument mentioned in § 6.2.1.

Finally, in terms of position within the noun phrase (the C-structure), determiner-like elements like demonstratives, quantifiers and numerals can generally occur before or after the noun they modify, with the exception of the definite article which can only occur before, i.e., prefixed to the noun it modifies (Benmaoun and Chouiri, 2013: 126-127). In addition, in the larger functional structure of Arabic clauses, there are certain restrictions with regard to the use of the definite article, e.g., the subject position: generic subject and subjects in SVCOMP clauses, the object position: generic object, and pre-posed objects, and the relative clause; § 6.6 further highlights such constraints.

6.3 Definiteness and the definite article in the dialect of *Yāl Sa'ad*

Broadly speaking, this dialect has two main morphological forms: 1) forms with zero marking, mainly associated with nouns in the indefinite state, and 2) forms with *l-* which corresponds to the Old Arabic definite article **l-*. Although zero marking is the main means of expressing the indefiniteness, in some indefinite contexts, the dialect makes use of an adnominal linker */-in/* which primarily occurs in the speech of older informants:

- in indefinite noun-adjective constituents: e.g., */fira:ʃ-in xifi:f/* 'a light blanket'; */dagg-in ʔa:wi/* 'a good beating'; */wilad-in sʕi:r/* 'a young boy'; */marħiba:j-in ʕo:di/* 'a big hello [to you]!';
- with the quantifier */kill/*: e.g., */kill-in maʕah ho:ʃah/* 'everyone with their own goats';¹⁹⁵

¹⁹⁵ Meaning everybody goes herding their own goats.

- with the specific /wa:ħid.SGM/ ~ /wiħdi.SGF/:¹⁹⁶ e.g., /wa:ħid-in aswad/ ‘a black one’.

The dialect also makes use of /wa:ħid/ apparently for more specification, e.g., /wa:ħid jga:llah.../ ‘someone named...’.

On the other hand, the definite article [l-] ~ [il-] ~ [li-] is exclusively used in the ‘grammatically’ and/or ‘semantically’ definite domain, serving a referential ‘previous specification’,¹⁹⁷ discursive, generic use, or just attached to proper nouns. Again, this *l-* changes according to the phonology of the noun stem and the preceding environment resulting in different realisations. An epenthetic vowel may precede the article, e.g., /ir-rizze/ ‘DEF-dot.SGF’, or follow it, e.g., /li-jda:r/ ‘DEF-wall.SGM’; however, in contextual position, the article is realised as /-l-/, e.g., /rawa:ʕi l-ħa:rah/ ‘people DEF-village.SGF; i.e., the people of the village’, and /fi-l-be:t/ ‘in-DEF-house.SGM; i.e., in the house’.

In addition, as in other Arabic varieties, the article assimilates to the following stem-initial coronal /t,d,n,θ,ð,r,z,s,š,sʕ,tʕ,ðʕ/, e.g., [r-rajja:l] ‘DEF-man.SGM’, [t-timu:r] ‘DEF-date.PL’, and [ð-ðahab] ‘DEF-gold.SGM’. Assimilation of the definite article seems to be intertwined with the syllabification pattern of the noun stem it attaches to. Notice the difference between /i-s-saʕdij-ji:n/ and /li-sʕadij-ji:n/ ‘DEF-Saʕdi-PLM’ where the second shows the CXaCv,¹⁹⁸ a by-product of the *Gahawa* syndrome and Najdi re-syllabification.¹⁹⁹ Another pattern that seems to variably block assimilation and trigger a following epenthetic vowel is the superheavy CCv:C as in /li-nxi:l/ ‘DEF-palm tree.PL’.²⁰⁰

¹⁹⁶ The marked form is realised as /wiħitt-in/.

¹⁹⁷ Ryding (2005: 158).

¹⁹⁸ The X here is a guttural sound, one of /h,ʕ,x,y,ħ/.

¹⁹⁹ See § 2.2.3 and Holes (2007: 481) for the *ghawa* syndrome in Oman.

²⁰⁰ Mostly a result of short unstressed vowel deletion in such noun stems, Holes (2007: 482); but, I have not analysed these forms thoroughly.

6.4 Null definite in the dialect of *Yāl Sa'ad*

The above described zero vs. *l*- indefinite-definite dichotomy is not clear cut when this dialect is considered. Whereas the definite article is exclusive to definite contexts (semantic and/or syntactic), definiteness itself is not exclusively expressed through the use of the overt definite article since this dialect shows other means of expressing definiteness, namely through the use of what I refer to in this study as the null definite. This category is basically another type of zero marking, but while, generally speaking, the difference between +DEF and -DEF in the standard and other vernaculars is a morphological one, i.e., marked by the presence or absence of the overt article, the difference between the overt article and the null definite is a morphosyntactic one in which both types are +DEF. In other words, some NPs which can potentially show syntactic definiteness through the use of the definite article variably occur without it; such nouns or constructions fail to show the article which otherwise is expected to occur, e.g.,

- 2) n-rabbi:-hin fi Ø-be:t
 1PL-raise.IPFV-3PLF.ACC in house.SGM.DEF
 ‘We raise them [the cows] in [the] house.’

- 3) a-tʃo:f ʃa Ø-ba:b
 1SG-look.IPFV at door.SGM.DEF
 ‘I was looking at [the] door.’

- 4) Ø-bʃi:r j-ʃill ʃigid
 camel.SGM.DEF 3-carry.IPFV.SGM grudge
 ‘[The] camel can hold a grudge.’

In the above examples, the nouns in bold are grammatically definite by virtue of them being referentially, discursively, and generically definite respectively, yet they lack the overt marking. This feature is reported for few other Arabic varieties as we shall see in the next section.

6.5 Other related varieties with a null definite

Within the study area, the null definite is not an exclusive feature of the *Yāl Sa‘ad* dialect. Some other Bedouin varieties across the study area have this feature. The speakers of these dialects either live close to or co-exist with the speech community under study. In addition, one could hear some of the phrases in § 6.7 in other non-Bedouin dialects in the area. I am a speaker of a sedentary-type dialect in the same area, and I use and have heard some of the constructions with the null definite, although quite restricted in nature and distribution, and mainly with noun-attribute combinations, where the noun does not show the definite article whereas the attribute does. These contexts do not strike me as missing the definite article as much as the ones in examples (2) to (4) above. While a survey of this feature in the whole of Oman may be a step for further research, I have come across another, most likely related, Bedouin variety in Oman that has null definite constructions of the same nature as those found in the dialect under study. This observation is based on a documentary TV interview with some Bedouin women apparently from the eastern region which I have briefly analysed. The main informant in the interview is a woman who belongs to the *Maj‘ali* tribe, presumably a branch of the larger *Jnaba* tribe. Also, a *Sa‘di* informant mentioned that this feature is characteristic of the speech of some Bedouin groups in Sināw, which is located in the eastern region.

In eastern Arabia, a similar feature is reported for Bahraini Arabic. Holes (personal communication) mentioned that the null definite occurs in the direct or indirect object of a

verb of movement, in actions involving moving towards a specific destination like ‘sea’, ‘home’, ‘school’, or ‘village’, a similar case to the English phrases ‘go home’ and ‘go to school’.²⁰¹ Here are four examples of null definite nouns in Holes’s data:

5) t-ru:ḥ-i:n Ø-madrasa
 2-go.IPFV-SGF school.SGF.DEF
 ‘Do you go to school?’ (Baḥārna dialect, Muḥarraḡ)

6) n-ru:ḥ Ø-be:t
 1PL-go.IPFV house.SGM.DEF
 ‘Let’s go home.’ (all dialects)

7) a-ḡdi Ø-mana:ma
 1SG-go.IPFV Manama.DEF
 ‘I go to Manama.’ (Baḥārna dialect, Dimistān village)

8) ji-kitt-ah Ø-baḥar in-no:xaḏa
 3-tip.SGM-3SGM.ACC sea.SGM.DEF DEF-captain
 ‘The captain tips it into [the] sea.’ (‘Arab dialect, Muḥarraḡ; *ibid*)²⁰²

In addition, it appears that a similar feature occurs in the Harub dialect in Saudi Arabia. In this dialect, three variants are used to mark definiteness: overt *l-*, *m-*, and what seems to be a null *l-* (Lowry, personal communication). This community shares some lexical similarities

²⁰¹ Holes commented that while both the Bedouin and the sedentary type dialects in Bahrain exhibit this feature, it occurs in a wider range of objects and is more frequent in the Shī‘i Baḥārna sedentary type. He further added that the fact that this category also occurs in the speech community under study is unsurprising considering the similarities and historical relations linking these populations.

²⁰² Transcription is edited to be in line with the thesis; glosses added.

with the dialect under examination, such as goat shepherding terminology, e.g., /hoːʃ/ ‘goats’, /ni-sraħ/ ‘1PL-herd.IPFV’. However, we will see that the grammatical categories and syntactic functions of this construction in the dialect of *Yāl Sa`ad* seem to be broader in distribution and nature.

Furthermore, the null definite category is described in considerable detail for Moroccan Arabic (MA) (Turner, 2013; 2018). Turner (2013) is a relatively detailed synchronic and diachronic re-investigation of definiteness in natural data collected for this variety. The article *l-* does not only mark definite meanings, but is also associated with indefinite meanings, along with the existence of another category ‘prefixless’, where some nouns lack this *l-* and yet they still denote definiteness, thus, leading Turner to question the status of whether *l-* is ‘truly’ a definite article in this variety at all. Overall, his analysis calls for a re-examination of the different forms of the definite article, and the adoption of a new model where the overt article is no longer a definite article, but rather “a lexicalized component of what is now the unmarked form of most etymologically Arabic nouns” that can occur in various semantic contexts (ibid: 56). On the other hand, “the historically indefinite form *Ø has come under heavy syntactic constraints and can best be described as derived from the new unmarked form via a process of a phonologically conditioned disfixation {-/l/}” (ibid: iv). In addition, Turner’s study touches upon the sociolinguistic bit of this NULL category arguing that MA lost the article diachronically due to the impact of some external and internal predictors, e.g., a substrate influence of Berber which is an article-less variety (ibid: 51-52), or more likely a concurrent structurally motivated development of the two native varieties in the area (ibid: 89). The main relevant departing point between Moroccan Arabic and the variety under study is that null definite nouns in Turner’s data (ibid: 113), do not have an etymological /l/, i.e., they have non-Arabic origin, whereas the null definite

nouns in this dialect are predominantly Arabic. Also, in the dialect of *Yāl Sa‘ad* the article *l-* is not used as an indefinite marker as in Turner’s ‘type identifiable’ category.

The role of language contact is quite clear in the case of Central Asian Arabic, e.g., the Bukhara dialect of Uzbekistani Arabic lacks the Arabic definite, like in the non-Arabic languages in the areas where it is spoken (e.g., Ratcliffe, 2005: 152). Owens discusses that one of the features that sets such peripheral Central Asian ‘mixed’ Arabic from other modern dialectal varieties is the lack of the overt marking of definiteness, and the development of other indefinite markers (2001: 355). Central Asian varieties share this feature with Arabic pidgins and creoles where the loss of the definite article is viewed as a reductional or ‘simplification’ strategy (Tosco and Manfredi, 2013: 409). The variation in the treatment of the definite article across the ‘Arabic complex’ led Retsö to conclude that the definite article *(a)l-* “is not a pan-Arabic feature and [thus] does not constitute an isogloss distinguishing Arabic in its modern sense from everything else” (2013: 438).

Furthermore, in the wider context of Semitic, and since proto-Semitic and the major daughter languages presumably did not have a definite article, it is no wonder that many Semitic languages lack a definite article. Some varieties described in Hetzorn’s (1997) *Semitic Languages* lack a fully-fledged definite article; they compensate for that through different means, like using the far demonstrative, e.g., in Neo-Aramaic (Jastrow, 1997: 357), the near demonstrative and/or the attribute, e.g., in Western Neo-Aramaic (ibid: 339), personal pronouns, e.g., in Eastern Neo-Aramaic (ibid: 357), and the possessive suffixes, as in Harari (Wagner, 1997: 492).

A note on N+Adj NULL DEF constructions

Attributive noun phrases where the head noun lacks an overt definite marking is an old but widely spread feature found in the Arabic varieties spoken throughout eastern and south-

eastern Arabia, as well as in Iraqi Arabic, and sedentary Levantine Arabic, along with other varieties outside Arabia, like Andalusian Arabic (e.g., Holes, 2018: 17, and the references therein). In the modern varieties of Arabic, this construction tends to occur in deictic phrases involving, e.g., enumeration, location ~ position ~ direction, and colour adjectives (Holes, personal communication). Examples from Bahraini Arabic include N+Adj constructions with singular, and plural head nouns, e.g.,

9) Ø-tʿa:bik il-fa:ni
 floor.SGM.DEF DEF-second.SGM
 ‘[the] second floor; lit. floor the second.’

10) Ø-guffat l-fa:nja
 basket.SGF.DEF DEF-second.SGF
 ‘[the] other basket; lit. basket the second.’

11) Ø-ban-a:t l-əkba:r
 girl-PLF.DEF DEF-big.PL
 ‘[the] older girls; lit. girls the old.’ (Holes 2016: 213-15).²⁰³

It is noteworthy that the fact that attributive phrases with the null definite have a wider and less restricted distribution throughout Oman (and different Arabic-speaking areas) than the constructions shown in examples (1) to (3) in § 6.4 above is suggestive that the former type has a common—not specifically Omani—origin (Holes, personal communication; more on the nature of the history and distribution of this feature in Holes, 2018: 17).²⁰⁴

²⁰³ Transcription is edited to be in line with the thesis; glosses added.

²⁰⁴ More on this construction in § 7.3.3.

In addition, although genealogically unrelated but relevant to this study, English shows a zero ‘null’ article. Berezowski (2009) describes this category as ‘the oddball’ in the English article system, and the concept of a zero English article is new compared to the other well-established forms. He builds his arguments on the origin and usage of this form in relation to the incomplete grammaticalisation of the overt definite and indefinite articles in two main categories: place names, and predicate nominals denoting offices held by single individuals. On the other hand, Krug & Lucas (2018) analyse the omission of the definite article in Maltese English and British English, as an example of a norm providing variety, with regards to ‘seasons of the year’, ‘ordinal numbers’, ‘languages’, ‘proper nouns’, ‘titles’, and ‘institutions’. Their key finding is that omitting the definite article is more frequent with generic NPs than definite ones, and that Maltese English definite article omission is different in that the definite article is omitted with unique or identifiable referents which are quite salient in the context (*ibid*).

The above cited works and findings are informative and useful for the current research; however, it is essential in the treatment of the null definite in this study to understand it more from different linguistic angles: its morphophonology, morphology, and morpho-syntax, and I begin with some diagnostic tests for the NULL category as used in domains previously well-established for the overt marker *l-*.

6.6 Diagnostic tests for the NULL category

Clitic-left dislocated Object NPs

A good test for null definite in the dialect is pre-posed object constructions with resumptive object pronouns on the verb, what Arabic syntacticians call clitic-left dislocation (CLLD). In Arabic, only NPs can undergo CLLD. However, it is not only that, as we see in the example

here from Lebanese Arabic, generally, only *definite* NPs can occur in such position.

- 12) l-ʔasʔi:de ʔallaf-a ʕomar
 DEF-poem-SGF write.PFV.3SGM-3SGF.ACC Omar
 ‘The poem, Omar wrote it (F).’

- 13) *ʔasʔi:de ʔallaf-a ʕomar
 poem-SGF write.PFV.3SGM-3SGF.ACC Omar
 Intended: ‘A poem, Omar wrote it (F).’ (Lebanese Arabic; Aoun *et al.* 2009: 194)

However, in my data nouns without the definite article can be clitic-left dislocated, as can be seen in the following examples:

- 14) Ø-bakrah j-samm-u:n-hi jdaʕah...w Ø-gʕu:d j-samm-u:n-ah jidaʕ
 young camel.SGF.DEF 3-call.IPFV-PLM-3SGF.ACC *jdaʕah* CONJ young camel.SGM.DEF 3-
 call.IPFV-3PLM-3SGM.ACC *jidaʕ*
 ‘[The] young female camel, they call it *ydaʕah*, and [the] young male, they call it *yidaʕ*.’

- 15) Ø-bʕi:r fallit-o:-h tʕilg
 camel.SGM.DEF unleash.PFV-3PLM-3SGM.ACC loose
 ‘[The] camel, they let it loose.’

- 16) awwal Ø-ghawa n-ħu:θ-θi
 before coffee.SGF.DEF 1PL-roast.IPFV-3SGF.ACC
 ‘In the past, [the] coffee, we would roast it (F).’

Subjects in SVCOMP

For sentences with grammatically indefinite subjects in MSA and Bedouin-type dialects, the norm is VSCOMP, e.g.,

- 17) *ʃa:f-ha* *riʒʒa:l*
 see.PFV.3SGM-3SGF.ACC man.SGM.INDEF
 ‘A man saw her.’ (adapted from Holes, 2004: 259)

Indefinite subjects normally cannot occur in SVCOMP (ibid); however, in this dialect, a null definite subject in an SVCOMP clause is a common occurrence.

- 18) *Ø-bʃi:r* *ma* *ðʕuwa* *be:t* *ra:ʕ-i:h...*
 camel.SGM.DEF NEG return.PFV.3SGM house.SGM owner-3SGM.GEN
 ‘[The] camel did not return to its owner’s house.’²⁰⁵

- 19) *kull ʃaj fi:-ha manfaʕah w Ø-bgurah t-ra:-ha t-ʕammir*
 every thing.SGM in-3SGF.GEN benefit.SGF CONJ cow.SGF.DEF 2-see.SGM-3SGF.ACC 3-live
 long.IPFV.SGF
 ‘Everything in it.F is beneficial, and [the] cow, you see, lives long.’

- 20) *jo:m ij-ji j-barriki Ø-bʃi:r, Ø-bʃi:r j-θu:r*
 when 3-come.IPFV.SGM 3-kneel down.IPFV.SGM camel.SGM.DEF camel.SGM.DEF 3-
 stand.IPFV.SGM
 ‘When he attempts to make [the] camel kneel down, [the] camel stands up.’

The above nouns also happen to be null generic subjects which are by default expected to show the overt article (see § 6.1).

²⁰⁵ Part of a narrative about a camel which did not return to his owner’s place; it turned out it was stolen.

Definite relative clause

The Arabic relative pronoun which introduces a definite relative ‘adjectival’ clause modifies a definite noun or noun phrase (Holes, 2004: 282). On the other hand, the indefinite relative clause, in which the relative pronoun is omitted, would modify an indefinite antecedent (ibid). I found one example only of a noun with a null definite and a modifying relative pronoun in the speech of a woman from al-Tharmad locality:

- 21) ja:n-ni Ø-mah-a:t illa gʻirri:b-a:t
 come.PFV.3SGM-1PL.ACC mother.PL.DEF REL close-PLF
 ‘...[the] female relatives who are closely related came to visit.’

Demonstrative-noun

When a noun follows the demonstrative constituting a single unit, the noun has to be definite (Pat-El, 2009: 23). I found few instances of demonstrative-noun constructions with the null definite, e.g.,

- 22) ga:l-l-ah ja: rajja:l haðe: Ø-bʻi:r kʻa:n ani a-ʻarif Ø-bo:ʃ he: Ø-bʻi:r ʃidu
 say.PVF.3SGM-DAT-3SGM VOC man DEM.SGM camel.SGM.DEF be-PFV.3SGM 1SG.NOM
 1SG-know.IPFV camel.PL.DEF DEM.SGM camel.SGM.DEF enemy.SGM
 ‘He said, ‘Oh man, this camel, if I really know camels, this camel is aggressive.’

The above evidence based on the syntactic distribution of the NULL category strongly suggests that such constructions are underlyingly definite. The next section concerns the linguistic distribution of the null definite as found in the interview data.

6.7 The linguistic distribution of the NULL category in the dialect

This section's objective is to provide a brief description of the NULL category in terms of its distribution across different linguistic levels. Starting with the phonological environment, the NULL category can occur with both: coronal-initial stems, e.g., /Ø-rja:l/ 'rial.SGM.DEF', and /Ø-ʃti/ 'winter.SGM.DEF', and non-coronal-initial stems, e.g., /Ø-ba:b/ 'door.SGM.DEF', and /Ø-gru:sʰ/ 'pancake.PL.DEF'. It can also occur with nouns that start with a voiced segment, e.g., /Ø-gimar/ 'moon.SGM.DEF', and /Ø-baħar/ 'sea.SGM.DEF', and with a voiceless segment, e.g., /Ø-kialme/ 'word.SGF.DEF', and /Ø-ʃams/ 'sun.SGF.DEF'. In addition, it can occur with different initial-syllable structures, including initial CC clusters, where the cluster is a result of: a) short unstressed vowel deletion, e.g., /Ø-gʃu:d/ 'young camel.SGM.DEF', b) the *ghawa* syndrome, e.g., /Ø-nxaʔa/ 'palm tree.SGF.DEF', or c) Najdi re-syllabification, e.g., /Ø-rguba/ 'neck.SGF.DEF'.

One interesting aspect of the phonology of this category concerns the preceding helping ~ linking vowel in some null definite constructions as used by older speakers, mainly the construct and prepositional phrases. This vowel is sometimes accentuated, e.g., /fo:go Ø-bo:ʃ/ 'on camel.PL.DEF; i.e., on camels' backs', /ʃani Ø-bo:ʃ/ 'about camel.PL.DEF', /hali Ø-mikja:n/ 'people place.SGM.DEF', and /ajja:mu Ø-ge:ðʰ/ 'day.PL summer.SGM.DEF'. For now, it seems that this is some sort of phonological modification to compensate for the omitted definite article, but acoustic analysis of the nature of these vowels is needed to understand the nature of this feature.

Furthermore, in terms of derivational morphology, nouns with NULL could be primitive, e.g., /Ø-ba:b/ 'door.SGM.DEF', derivative, e.g., /Ø-xajja:tʰ/ 'tailor.SGM.DEF'; they can be common, or proper. They can also be concrete, e.g., /Ø-θo:b/ 'dress.SGM.DEF', or

abstract, e.g., /Ø-bada:wa/ ‘bedouinness.DEF’. The null definite also occurs with different inflections of the noun (see Chapter Seven, § 7.3).

The types of phrases in which the null category occurs are also broad in range. In addition to occurring in independent NPs, the NULL category occurs in construct phrases, e.g.,

23) ziman(i) Ø-bo:f
 time camel.PL.DEF
 literally: ‘camels’ time’;

24) aḏa:n Ø-myarib
 call dusk.SG.DEF
 literally: ‘the *maḡrib* prayer’s call’;

25) su:g Ø-swe:gi
 market Suwaiq.DEF
 literally: ‘al-Suwaiq market’.

In addition, a predominant environment for the NULL category is noun-adjective phrases, where the modifying adjective, including numeral adjectives, has an overt article but not the modified head noun, e.g.,

26) Ø-ḡi:d l-watʿani
 celebration.SG.DEF DEF-national
 ‘the National Day’;

27) Ø-be:si s-so:di
 baisa.SGF.DEF DEF-black.SGF
 ‘literally: the black baisa’,²⁰⁶

28) ħiri:m li-ḳiba:r
 woman.PL.DEF DEF-OLD.PL
 ‘older women’ (see § 6.5).

Sometimes, both the head noun and the attribute do not show an overt article although the NP is underlyingly syntactically definite, e.g.,

29) Ø-jo:m Ø-ta:siʕ
 day.SGM.DEF ninth.SGM.DEF
 ‘the Ninth Day’,²⁰⁷

30) Ø-sʕaff Ø-awwal
 class.SGM.DEF first.SGM.DEF
 ‘the first grade’.²⁰⁸

Other phrases can also occur with the null category, e.g., demonstrative phrases, and prepositional phrases, as the following examples show:²⁰⁹

²⁰⁶ A type of traditional currency.

²⁰⁷ The day before religious celebrations in the ninth and the twelfth months of the Hijri calendar.

²⁰⁸ One can find /sʕaff awwal/ and /sʕaff l-awwal/, but never /sʕ-sʕaff awwal/, like in Moroccan Arabic, where the semantically indefinite noun shows the article, but the attribute does not (Brustad, 2000; Turner, 2013)

²⁰⁹ There are no quantitative phrases with NULL in the interview data, partly suggesting that this maybe a high-ranking constraint, but most likely due to the nature of quantification in the dialect; it seems to be a similar case to what Turner (2013) describes for the ‘type-identifiable’ indefinite nouns as a primary domain of quantification in MA. Preliminary analysis of such phrases in the dialect under study suggest the quantifier phrases with overtly marked nouns maybe borrowed wholesale from MSA.

31) da:xil hað-i:kj Ø-mikirfeh
 inside DEM-SGF stomach.SGF.DEF
 ‘inside that [animal’s] stomach’;

32) ʕa Ø-ba:b
 at door.SGM.DEF
 ‘at [the] door’.

Finally, the null category also occurs in different functional positions, including, but not restricted to:

Oblique objects, e.g.,

33) ni-rgʕaʕ ila: Ø-bla:d
 1PL-return.IPFV to town.SG.DEF
 ‘We return to [the] town’;

Subjects (agent, topic), including subjects of relative, and subordinate clauses, e.g.,

34) Ø-hiri:m li-kʕba:r j-libs-in ðahab
 woman.PL.DEF DEF-old.PL 3-wear.IPFV-PLF gold.SGM
 ‘Older women put on gold [jewellery]’;

Objects, including direct, indirect and pre-posed objects, e.g.,

35) la:kʕinna ma: n-ħib Ø-bgar
 CONJ NEG 1PL-like.IPFV cow.PL.DEF
 ‘...but [we] don’t like [the] cows’.

Adverbial adjuncts, e.g.,

36) le:lat Ø-sa:biʃ j-razz-il-ni awtʃi Ø-be:t jdu:ʃ

night.SGF seventh.DEF 3-nail.IPFV.PASS.SGM-DAT-1PL in front of house.SGM.DEF
trunk.PL

‘On [the] seventh night, tree trunks are nailed [to the ground] in front of [the] house
for us’.²¹⁰

Another significant point about the distribution of the NULL category is that, mainly in the speech of ‘linguistically conservative’ older speakers, some nouns could be realised with the null definite even when they have already been introduced in the discourse. One example is words for camels in the data, which are rather interesting. Generic camel names, e.g., /bʃi:r/ ‘camel.SGM’, /gʃu:d/ ‘young camel.SGM’, /bakrah/ ‘young camel.SGF’, and the collective /bo:ʃ/ ‘camel.PL’ are very likely to trigger the null category generally amongst these speakers. In general, for the older group, the null category is prone to occur with ‘core’ dialectal nouns associated with the older days, the lifestyle, and culture of the past. The list includes names for traditional practices, occasions, food, jewellery, household items, camel terminology, among others. More discussion on this part is in § 7.3.3 in Chapter Seven.

6.8 Sociolinguistic studies on the definite article

Quantitative variationist sociolinguistic research on the NULL category are limited if not scarce; I have come across two works only, namely, Tagliamonte and Roeder (2009), and Al-Qahtani (2015); the rest are descriptive or theoretical in nature (e.g., Brustad, 2000, Turner,

²¹⁰ In order to make traditional swings with them.

2013; and Turner, 2018 for Arabic, Krug and Lucas, 2018; Berezowski, 1997, and Berzowski 2009 for English).

6.8.1 Tagliamonte and Roeder on Yorkshire English

Tagliamonte and Roeder (2009) investigated variation in the English definite article, namely the definite article reduction (DAR) in a sample of fifty speakers from the city of York, northeast Yorkshire, England, using historical, dialectological and contemporary variationist approaches and employing three levels of analysis: diachronic, distributional, and multi-variate quantitative, with the aim of furthering the understanding of variation in the English definite article. One of the traditional DAR variants listed is the zero article which is quite similar to the null definite article in this study:

37) at \emptyset *seaside*.

38) into \emptyset *Fox Pub* for a pint.

39) going to \emptyset *lighthouse*.

40) \emptyset first *week*, it was great, but \emptyset second *week*, I waited to come home.

41) in \emptyset *house*. (ibid, citing W. Jones, 1952: 91)

The zero article in Tagliamonte and Roeder's study is reported as a characteristic specific to the dialect of southeast Yorkshire, and that is an old form of the English article (ibid: 437). The researchers also contend that the zero form is not related to the other DAR variants (five main variants including glottalised, and vocalic forms), and that both categories followed independent paths in their development which is evident in the linguistic tradition of the region (ibid: 439). With regard to the definiteness paradigm in the dialect, the researchers suggest a definiteness continuum with the full form 'the' and the zero form ' \emptyset ' at the poles (ibid: 442).

For the linguistic aspect of this variable, ‘multidimensional’ linguistic constraints including phonological, grammatical, and discourse pragmatic ones are examined using instrumental-auditory (Praat for duration and intensity), distributional, and multivariate (GoldVarb x) analyses. The results show a rich variability in the use of the different variants of the definite article. In terms of the social predictors, younger men are found to use more of the DAR variants, suggesting that these variants are being recycled as a marker of the local vernacular.

6.8.2 Alqahtani on Tihāmi Qaḥṭāni Arabic

Alqahtani (2015) also conducted a quantitative variationist study on the phonological variation in the definite article as used by informants from al-Jawwa village in the highlands, and al-Farša village in the lowlands of the ‘Asīr region, southwest Saudi Arabia. Her study highlights the effect of the geographical context on the nature of variation and the trajectory of change of the use of the definite article. The two variants of the definite article, namely the incoming *l*-article and the traditional *m*-article, as in /*l*-gahwa/~/*m*-gahwa/ ‘the coffee’, are examined. The results from multivariate analysis run using Rbrul show significant effect of the linguistic, social, and spatial predictors on the variation in the use of the definite article; a change in progress towards the incoming ‘koine’ form *l*- is specific to the lowlands speech community where it is led by younger women, and facilitated by the speakers’ ‘ambitions’, ‘attitudes’ towards the local vernacular, ‘tribal identity’, and socio-economically driven ‘geographical mobility’. In terms of the linguistic context, Alqahtani found that the *l*-article is never used when the noun stem starts with the ‘archaic’ fricative /*ħ*^s/, as for the rest of the factors, vowels are found to mostly favour the innovative *l*- followed by labials, back sounds, /*l*/, coronals, and /*z*/ respectively. The main points of departure between Alqahtani’s and my study are: a) the definite article as a variable in Alqahtani’s study is phonological, not morpho-syntactic, b) in terms of physical location, the two localities she studied are blocked

from each other by a mountain barrier, and in terms of accessibility, one locality is ‘more isolated’ and is connected to the ‘less isolated’ one by a narrow road across the mountain (ibid: 125-127); the localities in this study are equally accessible, there are no physical barriers blocking access to any of them, and there is a double highway system running in the middle of the study area.

Having provided an overview of the relevant literature and a background description for the NULL category in the dialect under study, the following chapter proceeds to present the quantitative analysis of the DEF variable. It outlines the main levels of analysis for the linguistic and external constraints along with describing the coding and modelling procedures for the multivariate analysis. Results from the statistical analyses are independently presented for individual sections of the analysis.

Chapter Seven: Data analysis for the DEF variable

This chapter presents a quantitative analysis of the DEF variable in this dialect. Starting with defining the envelope of variation, the chapter moves on to presenting the descriptive statistics and multivariate analysis of the external predictors, followed by a multivariate analysis of the linguistic constraints. The third main chunk of the chapter is a discussion of the findings of the different layers of analysis.

7.1 Quantitative analysis

7.1.1. Levels of analysis

The analysis of the DEF variable is multi-layered in order to effectively provide a holistic view of the variation in DEF in this dialect, but more importantly to help us understand the internal mechanisms by which the null definite in this dialect is being lost, and the external factors that may have fostered this process. There are three main layers of analysis, summarised in Figure 7.1:

- The first level of analysis targets the social and spatial aspects of this variation. The whole dataset is first described in terms of the distribution of the dependent variable within the three external predictors of age, gender, and locality. Then, a multivariate analysis of a subset of the data is carried out for all of the informants to explore the effect of these predictors on the variation at hand.
- The second level is concerned with the linguistic aspect of this variation. A balanced dataset extracted from interviews with the two informants (a man and a woman) with the highest proportion of null tokens per gender. The tokens are coded for a list of linguistic constraints, and a multivariate analysis is carried out.

- The third level of analysis involves a descriptive examination of some linguistic constraints for the null category only in the whole dataset.

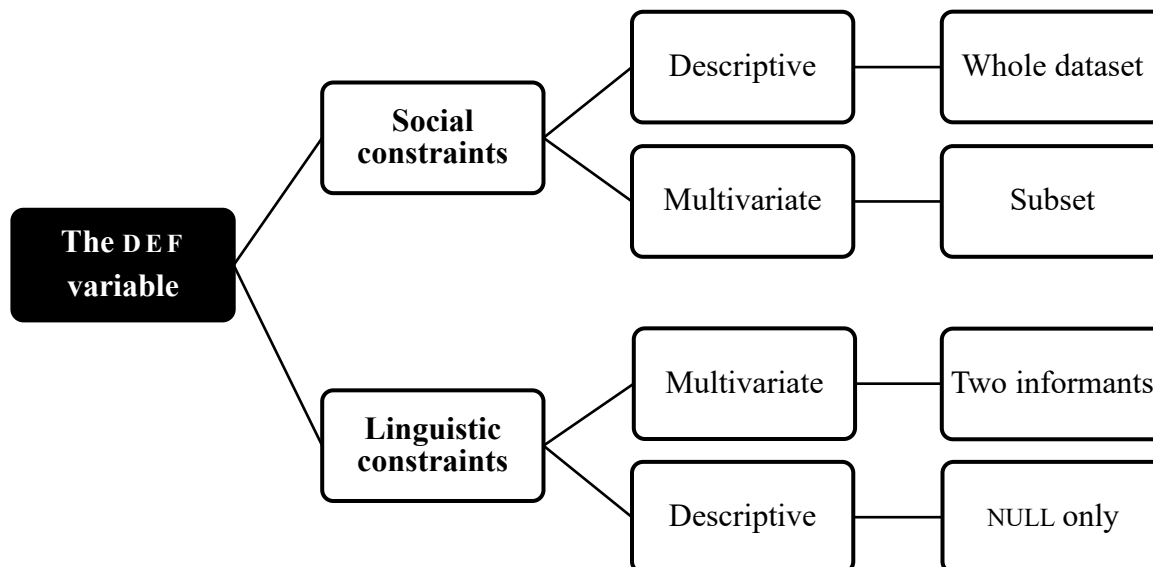


Figure 7.1 Stages of the quantitative analysis of the DEF variable.

7.1.2. Preparing the data for analysis

This section covers the process of preparing the data, mainly how and on what basis the tokens are extracted, and also defines the envelope of variation.

7.1.2.1 Token extraction

For coding the DEF variable, I extracted all of the tokens of the definite article *l-*, including nouns that can *potentially* take the definite article (i.e., nouns with the null definite). Overall, I depended on my intuition, the judgment of a native speaker, and the distributional properties of the definite article in Arabic in general and this dialect in particular as guidelines for coding the NULL category. Dialects in this study area are very much mutually intelligible, and thus, I can rely on my intuition as a native speaker of another local variety of

Omani Arabic, and my judgement on which constituents strike me as missing the overt article, i.e., sound ‘semantically or contextually inappropriate’ because of the lack of the overt marking.²¹¹

Furthermore, the coding of the DEF variable was for the most part impressionistic; however, for some of the tokens, where the assimilated definite article forms a geminate with the following coronal, it was sometimes not very clear whether there was indeed a geminate or not. In this case, the segment in question was measured in milliseconds using E-LAN; based on its length, it was decided whether it is a geminate or not. To do so, the segment’s length was compared to the length of a true geminate produced by the same speaker, i.e., a geminate which results from the doubling of a radical consonant. The length of a plain consonant was also measured for the purpose of comparison. If the segment in question was closer in length to the true geminate, then the token was coded as having an overt article. If its length approximated the length of a plain consonant, then the token was coded as having a null article. For instance, in an interview with a 62-year-old man from al-Tharmad locality, in the phrase /ajja:m ʃ-fti/ ‘days of winter’, the fricative is approximately 132 milliseconds long, compared to 169 milliseconds for a proper geminate produced by the same speaker in /mʃaʃʃi/ ‘a man whose job is to feed camels’. A plain fricative, on the other hand, is approximately 75 milliseconds, as in /la ʃwajji/ ‘a little’. Since the length of the fricative is closer to that of the true geminate, the above token was coded as having an overt article. In addition, I faced a problem with few tokens where the token is preceded by a word-final /l/; it was difficult to decide on whether there was an overt article or not, as in /ma:l l-.../ ‘GEN DEF-...’, thus, such tokens were excluded.²¹²

²¹¹ Turner (2013: 95).

²¹² Potter faced the same issue with an s- or a z-initial word following verb *to be* and decided to exclude them as it is impossible to decide whether the segment belongs to the verb *to be* or the following word (2018: 143).

7.1.2.2 The envelope of variation and excluded tokens

The envelope of variation includes instances of the definite article, OVERT and NULL, where the article is or can potentially be attached to a bare noun. It covers tokens expressing the referential, discursive, and generic definite(s), see § 6.1 in Chapter Six. In addition, in the instances when the noun is modified by an adjective, the whole NP is considered as a token, but the article attached to the modifying adjective is not considered as an instance of the DEF variable. Instances of noun-adjective phrases where the overt article in the modifying adjective is omitted as well, such as in phrases like /Ø-jo:m Ø-ta:siʕ/ ‘[the] ninth day’ and /Ø-sʕaff Ø-awwal/ ‘[the] first grade’ are included. Adjectives which function as substantives, i.e., nominalised adjectives, are included. In addition, the relative marker /illi/ < **ʿallaḍi* ‘that’ along with its variants is not considered as instances of the DEF variable.

The following categories are excluded for a different set of reasons:

- Some words like /l-ḥi:n/ ‘now; literally: DEF-now’ and /l-jo:m/ ‘today; literally: DEF-day’ are always realised with the overt article; as far as this dialect is concerned, these words never appear as /ḥi:n/ or /jo:m/ with the meaning of ‘now’ and ‘today’.²¹³ In Arabic dialects, these tokens function as time adverbials where the categoricity of such words is probably due to a lexically conditioned use of the definite article; thus, such instances are excluded from the analysis.
- There is a whole set of adverbial and functional phrases that are borrowed from MSA; these expressions almost always appear with an overt article, and thus, are excluded, e.g., /l-muhum/ ‘most importantly’, /bin-nisba/ ‘with regard to’, /biðʕ-ðʕabtʕ/ ‘exactly’, /ʕala sabi:l l-miḥa:l/ ‘for instance’, /bil-ʕaks/ ‘on the contrary’, /bið-ða:t/ ‘especially’. Also excluded are proportion phrases like /bil-mija/ ‘percent’, /bin-nisʕsʕ/ ‘in half/

²¹³ One may hear *kill ḥi:n* ‘every time’, but never *ḥi:n* on its own and also meaning ‘now’.

fifty-fifty’.

- The excluded list also includes the name of Allah and other formulaic expressions involving this noun, e.g., /subħa:nalla:h/ ‘Glory be to Allah’, /ma:fa:ħah/ ‘God willed’, and /infa:ħah/ ‘God willing’. The only exception to this category is /l-ħimdilla:h/ ‘Praise be to Allah’, because it is in variation with /ħimdilla:h/.
- All tokens included in recited traditional poems and songs are excluded from the coding process.
- Instances of code-switching to English or MSA are excluded straight away. Fixed expressions and phrases borrowed wholesale or verbatim from MSA are not included in the analysis. Older borrowings from other languages like Persian and Hindi are included. Borrowings from English, on the other hand, are dealt with differently:

⇒ Older borrowings that were introduced to the study area in the pre- and early 1970s, which have been fully integrated in the vernacular are included, e.g., /dre:wil/ ‘driver’; /karjal/ ‘carrier; a pick-up truck’s cargo bed’, /swi:ki/ ‘ignition key of a car’, and /le:ta:t/ ‘lights/light bulbs’.

⇒ Also included are the ‘not so recent’ borrowings during the post-TV pre-internet era, mostly used by the informants in the middle age group and the immediately younger informants; these include words that have equivalents in the standard, yet the speakers generally prefer to use the English equivalents, e.g., /t-tilfizju:n/ ‘the TV’, /t-te:lafu:n/ ‘the telephone’, /l-kambju:tar/ ‘the computer’, /mo:de:l/ ‘model’, /l-iliktro:nijja:t/ ‘the electronics’, and /l-internet/ ‘the internet’. In the dataset, some of these are variable; they can occur with NULL; their etymology does not have /l-; they got incorporated in the dialect and by analogy to other dialectal forms, their definiteness is marked with the article.

⇒ Newer borrowings, which are primarily produced by the younger informants in the sample are excluded, e.g., /l-masiq^j ~ l-masid^z/ ‘the text message’, /l-wa:jfa:j/ ‘the Wi-Fi’, /l-wa:jirles/ ‘the Wireless’, /s-so:ʃal mi:dja/ ‘the social media’, /l-ju:tju:b/ ‘the YouTube’, /s-sna:b ʃa:t/ ‘the Snapchat’, and /l-isntigra:m/ ‘the Instagram’.

- Tokens of the variable included in the metalinguistic comments made by the informants on the topic of the null definite are excluded. Similarly, tokens that appear in discussions about the feature itself are not included.

7.2 Data analysis: external predictors

This level of analysis includes two sub-sections:

- A descriptive analysis of the distribution of OVERT and NULL tokens across the three external predictors of age, gender, and locality in the whole dataset.
- A multivariate analysis of these predictors in a subset of the data for all of the informants.

7.2.1 Distributional statistics

For the first stage, an Excel sheet that has all of the tokens extracted from the interviews of the forty informants was prepared, and tokens of the DEF variable were coded for the dependent variants and the external constraints. Out of the total number of tokens (N=16,197), 543 were realised with the NULL variant; this is just 3% of the total, and the remaining 97% (N=15,636) is the proportion of the OVERT variant in the dataset. There are

266 different noun phrases coded as NULL, and these phrases carry 184 different bare nouns; see Appendix C for a whole list.

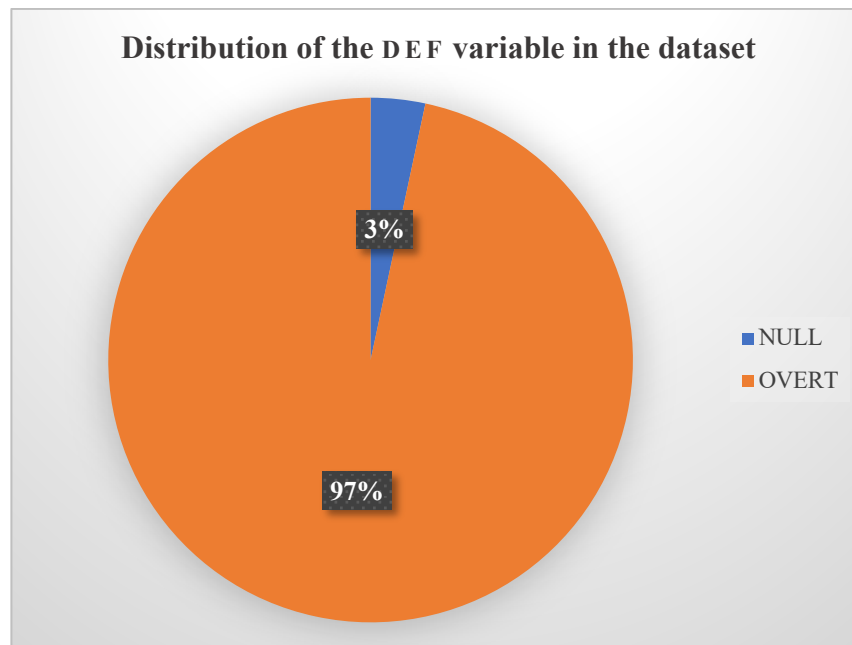


Figure 7.2 Distribution of the dependent variants in the dataset.

The rest of the section will focus on the distribution of the NULL variant, since generally speaking the proportion of the OVERT variant is considerably higher across the external predictors. Also, since the variation is limited due to the low proportions of NULL compared to OVERT, which although enough to show trends, the conclusions and generalisations that we can draw about the variation in DEF are best considered cautiously. However, a closer look at the NULL variant may present a clearer picture on where the different predictors stand in relation to this variation. The pattern showed in Figure 7.3 generally reflects the situation across the external categories as well: the percentages for OVERT approximating the 100%, whereas NULL is very restricted. Nonetheless, some trends can still be deduced.

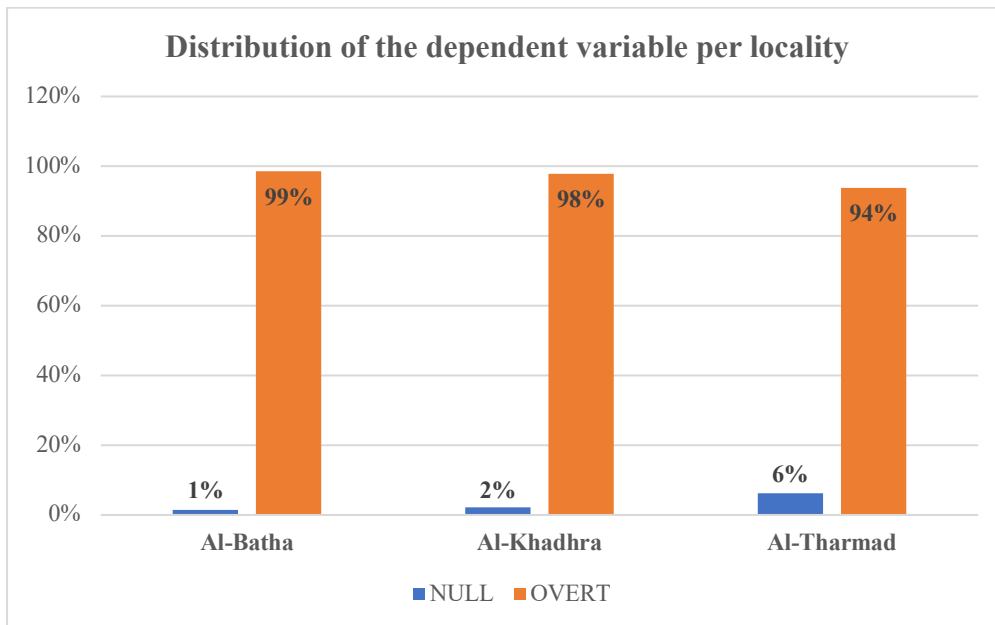


Figure 7.3 Distribution of the dependent variable per locality.

Starting with locality, al-Tharmad shows the most variation, with the NULL variant comprising 6% (N=367) out of the total for this locality (N=5,931), followed by al-Khadhra (2%; N=80 out of N=3,701), and al-Batha (1%; N=96 out of N=6,547); see Figure 7.3.

In terms of age, the middle and younger age groups show similar patterns with 98% (N=7,191) and 97% (N=3,728) of the OVERT variant compared to 2% (N=126) and 3% (N=128) of NULL out of the totals for each group (N=7,317 and N=3,856 respectively); on the other hand, the older age group is slightly lagging behind with 94% (N=4,717) of the OVERT variant out of the total for this group (N=5,006), as shown in Figure 7.4. The older age group's proportions mirror those found in al-Tharmad locality with both groups showing 6% of the NULL variant out of their totals.

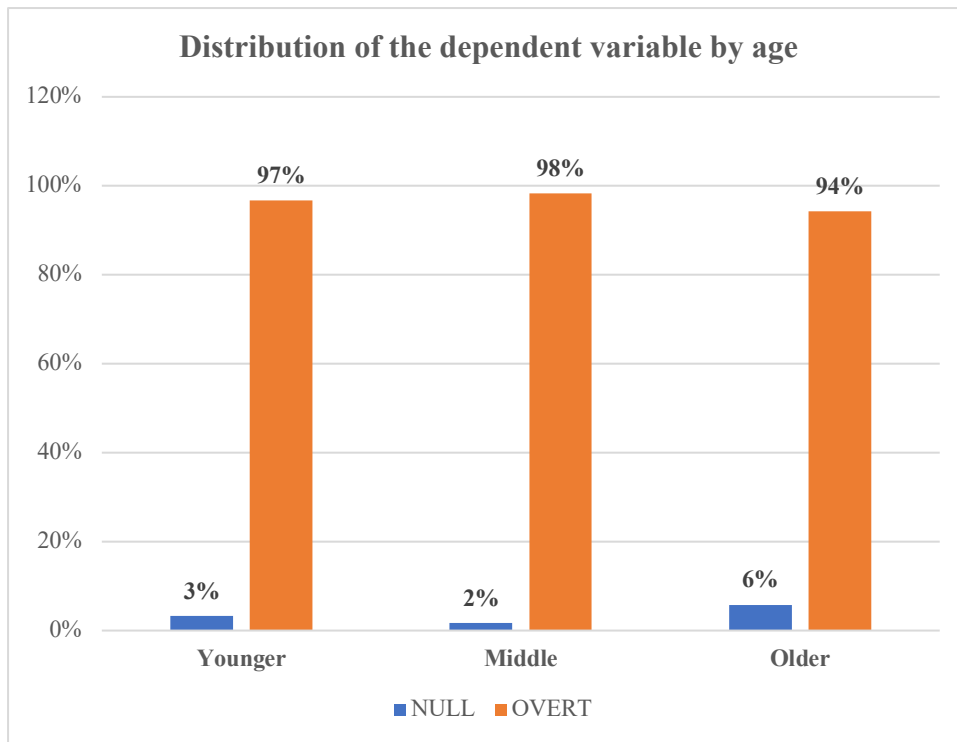


Figure 7.4 Distribution of the dependent variable by age.

On the other hand, there is very slight difference within the gender group; men and women behave quite similarly with the men producing 3% NULL (N=221 out of N=7,044), and women producing 4% for the same variant (N=322 out of N=9135), as shown Figure 7.5.

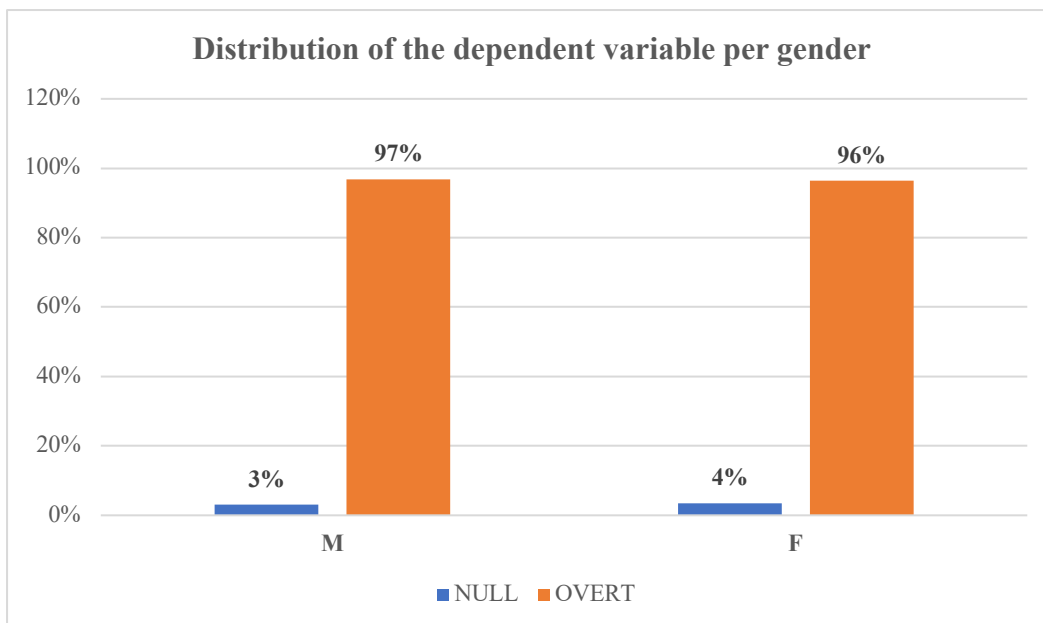


Figure 7.5 Distribution of the dependent variable per gender.

Age/ gender	Younger males	Younger females	Middle males	Middle females	Older males	Older females	Totals
NULL	3% N=33	4% N=95	2% N=70	1% N=56	4% N=118	8% N=171	N= 543
OVERT	97% N=3350	96% N=2578	98% N=3029	99% N=4162	96% N=2644	92% N=2073	N= 15636
Totals	N=1183	N=2673	N=3099	N=4218	N=2762	N=2244	N= 16179
	N=3856		N=7317		N=5006		

Table 7.1 Distribution of the dependent variable per age and gender.

A cross-tabulation of age and gender shows that women and men in the middle age group show the least proportions of NULL (1% and 2% out of the total for each group respectively); on the other hand, older women show the highest proportion of the same

variant (8%); the rest of the groups show, more or less, similar patterns as shown in Table 7.1.

Locality	Gender	Age						Totals
		Younger		Middle		Older		
		NULL	OVERT	NULL	OVERT	NULL	OVERT	
Al-Batha	M	4%	96%	2%	98%	1%	99%	N=2487
		N=20	N=425	N=15	N=845	N=11	N=1171	
Al-Batha	F	3%	97%	0%	100%	1%	99%	N=4060
		N=32	N=1162	N=8	N=1984	N=10	N=864)	
Al-Khadhra	M	2%	98%	1%	99%	2%	98%	N=2012
		N=12	N=499	N=7	N=1006	N=11	N=477	
Al-Khadhra	F	1%	99%	2%	98%	6%	94%	N=1689
		N=6	N=470	N=14	N=700	N=30	N=469	
Al-Tharmad	M	0%	100%	4%	96%	9%	91%	N=2545
		N=1	N=226	N=48	N=1178	N=96	N=996	
Al-Tharmad	F	6%	94%	2%	98%	15%	85%	N=3386
		N=57	N=946	N=34	N=1478	N=131	N=740	
Totals		N=128	N=3728	N=126	N=7191	N=289	N=4717	N=16,197
		N=3856		N=7317		N=5006		

Table 7.2 Distribution of the dependent variable per age, gender, and locality.

Furthermore, the cross-tabulation between the three predictors this time shown in Table 7.2, generally shows that older women have higher proportions of NULL, with the ones from al-Tharmad coming first with 15% of this variant out of the total for their group. Older women from al-Batha, however, show the least proportion of NULL with 1% only out of the total for their group. It is also older men from al-Tharmad who show the highest proportion

of NULL amongst men (9% out of the total for their group for this locality), the rest of the men groups more or less behave similarly with regards to favouring NULL, apart from younger men from al-Batha and men in the middle age group from al-Khadhra locality. It is noteworthy that the group that shows the highest proportion of NULL after the older men and women from al-Tharmad locality is younger women from the same locality with 6% of NULL out of the total for this group in this locality. On the other hand, generally speaking, women in al-Batha locality overall display the least variation, whereas women from al-Tharmad show the most. The two near categorical groups are: women in the middle age group from al-Batha, and the younger man from al-Tharmad locality.

7.2.2 Multivariate analysis

For the second stage, another Excel sheet with a total of 7,906 tokens was prepared, but this time for the purpose of running a logistic regression for the external predictors for the whole sample. A fairly balanced number of tokens from each of the forty informants was extracted, with an average of just above 197 tokens per speaker, and a token count ranging from as few as N=114 to as many as N=406 per speaker for cells with fewer numbers of informants, but mostly the range is within 150-200 per speaker. The results from this stage are shown in Table 7.3. However, before moving to the Rbrul results, the distribution of the dependent variable per speaker is shown below to highlight the linguistic behaviour of some outliers who showed different patterns to the rest of the informants in their groups.

Distribution of the dependent variable per speaker and some individualised explanations

Figure 7.6 shows distribution of the dependent variable per speaker; this table is based on the Excel file specifically prepared for the multivariate analysis to achieve a fair distribution of tokens per speaker:²¹⁴

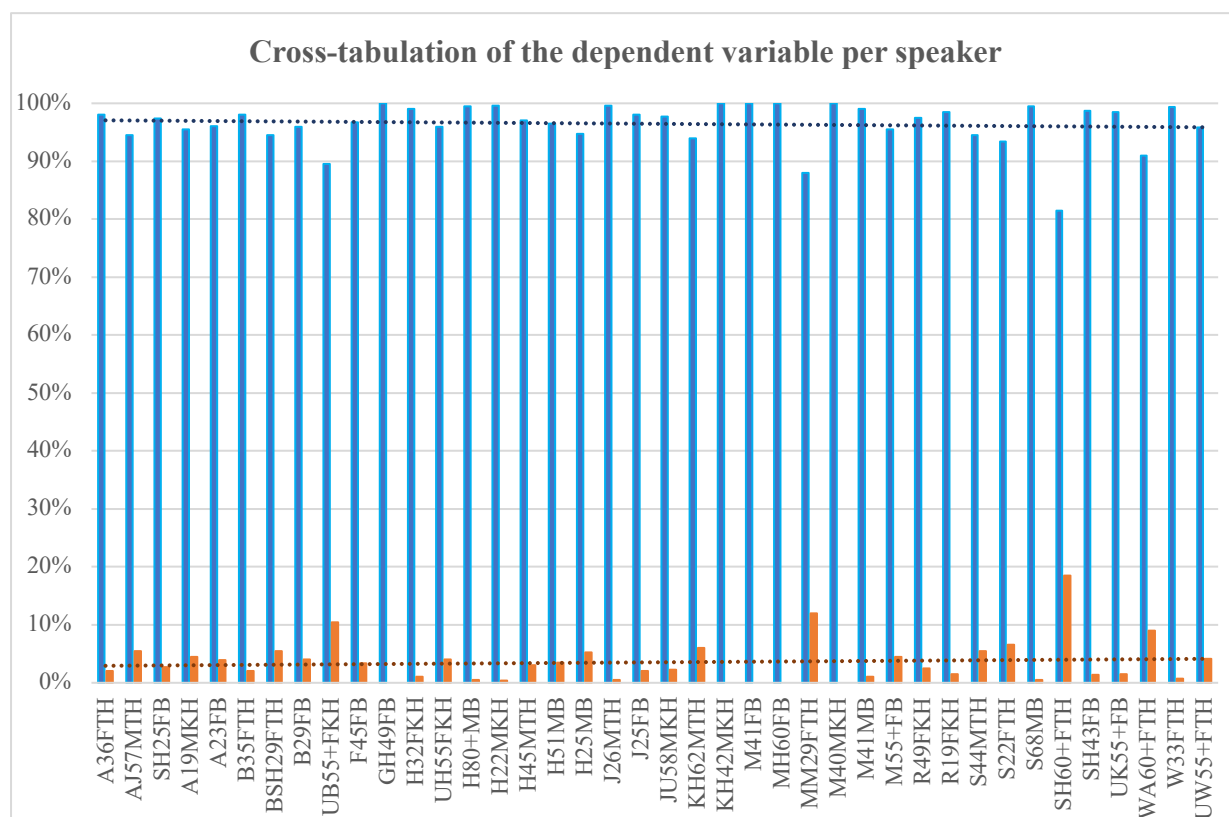


Figure 7.6 Distribution of the dependent variable per speaker.

The pattern shown in Figure 7.6, clearly shows that most speakers fall close to the 100% for OVERT and to the 0% for NULL. In fact, in terms of the distribution of OVERT, 85% of the speakers (N=34) produced 95% or more of this variant; the remaining 15% is the six informants who scored below 95%. Out of the thirty-four informants who scored above 95%,

²¹⁴ The speaker codes are presented as the informant's initial, real age, gender, locality: B for al-Batha, KH for al-Khadhra, and TH for al-Tharmad.

there are nine participants who use the OVERT variant categorically and six near- categorical users with 99% of the OVERT variant. In addition, the lowest proportion of the OVERT variant reported is 82%. Some speakers' results are discussed below.

Salima (SH60+FTH)

Salima is an older woman from al-Miladda village.²¹⁵ She estimated her age as 'maybe 60'.

With a proportion of 82% of the OVERT variant, she is the most linguistically conservative amongst the forty informants. This result came as no surprise since she may be considered a representative of the 'native' speaker of this dialect; she was raised as a child in a *Sa'di*-only village. Describing the place that she grew up in, she says:

killhum ja:l saʕad; ħin ji:ra:nni, killni ja:l saʕad; w killni ahal, w
ma: ʕo:di baʕad l-ħa:rah; jmikʕin θima:njat bju:t

'All of them are *Yāl Sa'ad*, our neighbours, all of us are *Yāl Sa'ad*; and we are all one family, and the village is not big; there were probably eight houses.' (SH60+FTH)

This informant had not done any formal education; apart from attending Quran lessons when she was a child; she is functionally illiterate. She has lived locally her entire life and maintained local social contacts. When Salima was sixteen years old, she got married to a *Sa'di* man from al-Khadhra locality. She still visits her childhood house occasionally, and her daughters-in-law are from the same area she grew up in. This could explain the high proportion of NULL in her speech. This informant is interesting because she is quite assertive about the idea that *Sa'dis* from all over the *mishāb* of *Yāl Sa'ad* speak the same; they also have the same traditions, and the same names for items related to their traditional lifestyle. When I commented that other informants did not mention some of the traditional jewellery she was talking about, implying that al-Batha's dialect maybe different from the other

²¹⁵ In this research, part of al-Tharmad locality.

localities, she got a bit defensive and disregarded this idea. Interestingly enough, this seems to be the stance adopted by older female informants in general.

Hamid (H80+MB) and Saeed (S86MB)

Another interesting observation about the older age group is that the two oldest informants, Hamid and Saeed, are near categorical users of OVERT. Although quite unexpected, the results for both men could be understood with regard to the nature of their circle of contacts and their status in the community. The first, Hamid is an important tribal figure. He lives in al-Batha locality and has not had any formal education. His family has been in contact with Bedouin and non-Bedouin speech communities, through the family business mainly farming, the immediate neighbourhood, friends, outside connections, and to a lesser extent marriage. Hamid constantly meets people from within and outside the study area due to his current position. On the other hand, Saeed has occupied some relatively high-ranking governmental offices and moved around for a while until his retirement many years ago. Before that, he spent some time in the Gulf where he worked and got some form of formal education. Saeed's interview was interesting because throughout the interview, which lasted more than an hour, Saeed was constantly using overt marking for the definite article. Yet, once he started talking about the past lifestyle of the Bedouin family, and the Bedouin-camel connection, he suddenly started using the null definite; the word /Ø-boːʃ/ 'camel.PL.DEF' in his interview, as also the case with other older informants, is constantly realised with NULL.

Salim (S44MTH)

In the middle age group, we see the most linguistically advanced speakers from both gender groups. The majority of the informants in this group have had at least an undergraduate degree, and most of them work in jobs involved in the educational sector, either as current or previous schoolteachers. One informant who behaves quite differently from the rest of his

group is Salim from al-Tharmad locality, (NULL=6%). Salim worked as a schoolteacher for a while, but now works in the Ministry of Education as a local supervisor. Just looking at the numbers may give the impression that Salim uses the traditional form much more than the rest in his group; however, a quick analysis of his NULL tokens reveals that he uses null in a more restricted sense, mostly with local place names, and phrases related to schools and traditional events (more in § 7.3).

Hamdan (H25MB)

Data from a young man from al-Batha locality, Hamdan, is also noteworthy; 5% of his tokens are NULL, the highest amongst young men. Hamdan finished high school, worked for the military for a while, and at the time of the interview, he was commuting to Dubai where he worked as a private driver. His father has been a camel trainer for a long time and Hamdan and his brother used to help him with his job. Hamdan is active in organising/attending camel events in and around the study area. The reason Hamdan's linguistic behaviour is intriguing is that his speech is more conservative in general, but more because of the way he relates to the Bedouin identity in terms of the dialect and traditional lifestyle, and to the camel culture as well. He used the word /bo:j/ with NULL twice in /ziman(i) Ø-bo:j/ 'time camel.PL.DEF' and /hal Ø-bo:j/ 'people camel.PL.DEF', and these two are the only instances of such usage outside the older group. Hamdan's linguistic behaviour could be understood with regard to his involvement with his father's career throughout his upbringing and to his current personal interests. Although he was 25-year-old at the time, he expressed his admiration of older times and described himself as /wilad zima:n/ 'lit. the son of the past'.

Furthermore, while these observations tell us more about the motivations of some informants to linguistically behave in a certain way, i.e., are important in that they can show

the relationship between social factors and the use of the NULL category in a more fine-grained way, what is needed, however, is a multivariate analysis for the three external predictors to understand where the latter groups stand with regards to the DEF variable, and this is what the next section is about.

7.1.2.3 Rbrul results in relation to the external constraints

Table 7.3 shows the results from the multivariate regression analysis for the social constraints on the DEF variable. The application value in all of the modelling stages is OVERT. Rbrul returned all of the three external predictors as significant; locality came first, followed by age and gender respectively.

Application value=OVERT; N=7,906; overall proportion= 96.5%; R²=0.119; log likelihood= -1146.365				
External predictors	Log odd	FW	Tokens	OVERT %
Locality (p<0.0001) (p=3.04e-13)				
Al-Batha	0.375	0.593	2,893	97.8%
Al-Khadhra	0.247	0.561	2,473	97.5%
Al-Tharmad	-0.623	0.349	2,540	94%
Age (p<0.0001) (p=1.25e-09)				
Middle	0.609	0.648	2,650	98.2%
Young	-0.198	0.451	2,597	96.1%
Old	-0.412	0.398	2,659	95.1%
Gender (p<0.005) (p=0.00144)				
Male	0.205	0.551	3,491	97.3%
Female	-0.205	0.449	4,415	95.8%

Table 7.3 Rbrul results of the DEF variable: external predictors.

The statistical results confirm those reported in § 7.2.1. Overall, the OVERT variant is highly favoured with an overall proportion of 96.5%, compared to that of the NULL variant (3.5%). It is noteworthy that when considering the three external predictors together, i.e., running them in the same model, their effect is low ($R^2=0.119$) probably due to the steep proportion of the OVERT variant compared to that of the NULL. In terms of locality, al-Batha favours the application value the most with a factor weight of 0.593, followed by al-Khadhra (FW=0.561); al-Tharmad, on the other hand, disfavours the application value with a factor weight of 0.349; it shows the most variation amongst all other groups with a 94% proportion of the OVERT variant. Even though al-Tharmad disfavours the application value, the proportion of the OVERT variant is quite high; again, this is true of all the groups in the model. Turning to gender, men favour and show slightly higher proportion of the application value OVERT (FW=0.551) than that of women who slightly disfavour the application value with a factor weight of 0.449. Furthermore, the middle age group shows the least variation amongst all other groups with a 98.2% overall proportion of OVERT and a factor weight of 0.648. The younger and the older age groups disfavour the application value with FWs of 0.451 and 0.398 respectively. The type of NULL variants for the latter groups are quite revealing and can guide us to more careful conclusions about the behaviour of these categories. It is right the younger age group approximates the older age group in terms of the proportion of the NULL variants. However, the younger age group varies in that the function and the type of the noun that is realised with the NULL variant are different. More discussion on the type of nouns will follow, but we can generally say that the types of tokens for the younger age group fall within the traditional practices, scholastic terminology, toponyms, and adverbials (more on this in § 7.3.3).

7.3 Data analysis: Internal constraints

I have not coded for the linguistic constraints for the selected subset of the data analysed above, primarily due to the paucity of the NULL tokens which most likely would result in a big number of zero cells in the different values for the linguistic groups. However, when I was coding the whole dataset, I had generally observed some patterns or trends with regard to the distribution of the dependent variants in the dataset. For instance, nouns preceded by demonstratives tend to have a prefixed overt article. In addition, nouns would likely show an overt article if preceded by some function words, e.g., /mi-/ ~ /min/ ‘from’, /maʕ/ ‘with’, /ʕind/ ‘with/by/at’, /ʕala/ ‘on’, the analytic genitive and the conjunction /w/. Therefore, a quantitative analysis of the linguistic nature of this variation was deemed to be necessary, and this section concerns the analysis designed to address this aspect of the DEF variable in the dialect under study.

For this step in the analysis, tokens of the DEF variable are extracted from the interviews of a sixty-eight-year-old man and a woman aged 60+ from al-Tharmad locality, namely Khamis (KH62MTH) and Salima (SH60+FTH);²¹⁶ these two speakers are chosen because they showed the highest frequency of the null definite tokens per the two gender groups, thus the null definite is still functional and relatively much-less restricted in their grammars. These informants are within the same age group, and they originate from the same locality.

A total of 800 tokens were extracted and coded for the dependent variable (OVERT or NULL) and five linguistic constraints. Table 7.4 shows the proportions/counts for the two informants; the male informant precedes the female informant in the realisation of the OVERT variant. Whereas the extracted tokens for the woman run throughout the entire interview (total

²¹⁶ The same woman in § 7.2.2.

N=394), because the man's interview is quite long, less than two thirds of the total number in the entire interview (total N= 697) were extracted; nonetheless, the amount of variation depicted in the sample extracted from his interview more or less reflects the variation in the entire interview; the total percentage of NULL for the male informant in the entire interview is about 11% (N= 80).

Speaker code	NULL	OVERT	Totals
Khamis	10% (N= 41)	90% (N= 365)	51% (N= 406)
Salima	19% (N=74)	81% (N= 320)	49% (N= 394)
Totals	14% (N=115)	86% (N= 685)	N= 800

Table 7.4 Distribution of the dependent variants for Salim and Khamis.

The following section presents the coding procedures, the modelling process, and the results for this part of the analysis.

7.3.1 Coding procedure for the linguistic constraints

Along with making use of some of the existing literature on definiteness in Arabic and the Arabic definite article discussed in § 6.1 and 6.2 in Chapter Six, the choice of the linguistic factor groups and individual factors is largely based by previously established hypotheses on the nature of the distribution of the dependent variants, as observed throughout the interviews and then during the token extraction process; e.g., a hypothesis that local place names, items relating to the speech community's traditional lifestyle, and camel terminology tend to be realised with the NULL variant, especially within the older age group, whereas the younger and

middle age groups would show more concentration of NULL tokens around words relating to schooling, and traditional occasions. Figure 7.7 summarises the five main linguistic constraints coded for in this study.

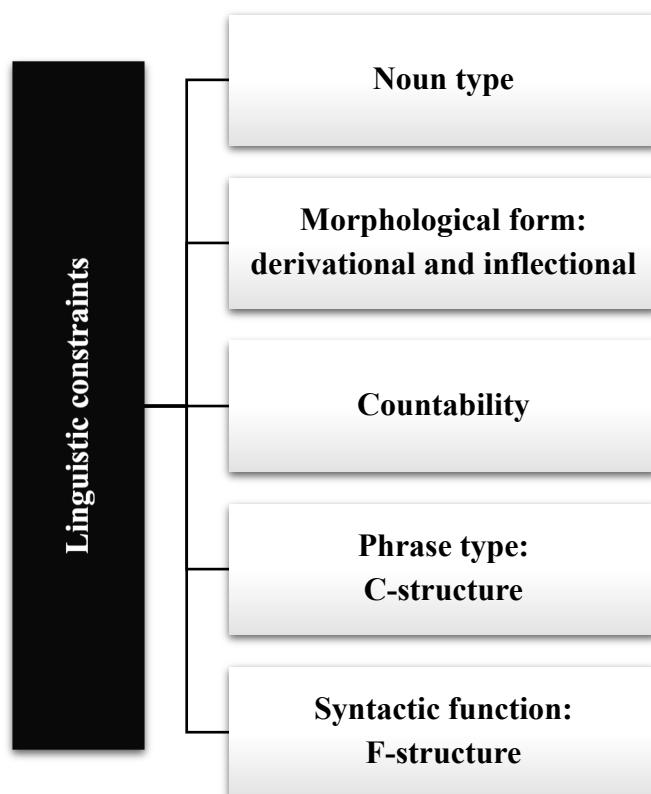


Figure 7.7 Linguistic constraints on the DEF variable in this study.

Type of the bare noun

This factor group is coded in three ways:

- common vs. proper.
- abstract vs. concrete.
- a customised recoding of the common vs. proper noun group, which incorporated animacy, humanness, concreteness, and ‘culturalness’ of the nouns, yielded eighteen

factors summarised in Table 7.5.

Type of bare noun	Examples
Personal name	/s ^s -s ^s alt ^s / ‘a personal name’; /l-jahja:ʔi/ ‘a surname’
Weekday	/jo:m l-ximi:s/ ‘Thursday’
Nationality-ethnicity	/l-banga:lijji/ ‘the Bangladeshi [workers]’
Toponyms	/l-ima:ra:t/ ‘the UAE’
Formulaic phrases	/l-ħimdilla:h/ ‘Praise to Allah’
Place adverbial	/l-gʲamb/ ‘the side’
Time adverbial	/ʒas ^s ir/ ‘[the] afternoon’
Season	/ʃ-ʃiti/ ‘the winter’
Numeral: cardinal and ordinal	/θ-θinte:n/ ‘the two [camels.F]’; /θ-θa:nji/ ‘the second.F’
Human	/l-wa:lid/ ‘the father’
Camel	/bo:ʃ/ ‘[the] camels’
Other animals	/l-yanam/ ‘the sheep’
Food related noun	/l-ħili:b/ ‘the milk’
Topographical feature	/fi baħar/ ‘in [the] sea’
Domestic	/l-lak ^l da:ni/ ‘a type of jewellery’; /fi be:t/ ‘in [the] house’
Other inanimate	/l-maʃrað ^s / ‘the exhibition’, /l-mistaʃfi/ ‘the hospital’
Traditional practices (occasion-event)	/jo:m ta:siʃ/ ‘the ninth day [before the Eid celebration]’
Other abstract nouns	/t-taʃli:m/ ‘the education’

Table 7.5 Coding for the type of noun factor group.

Morphological nominal form (shape) of the bare noun

Derivation:

All the nouns included in this analysis are used substantively, but in terms of their derivation, their nominal morphological forms could be classified in terms of three main categories associated with Arabic nouns, namely: the verbal abstract (also known as the *maʃdar* or the *nomina verbi*), primitives, and derivatives. The first category includes verbal nouns, e.g., /r-rakið^s/ ‘a traditional camel show; lit. the running’, and /t^s-t^si:bi:x/ ‘DEF-cooking’. Primitive

nouns, e.g., /l-ba:b/ ‘DEF-door’, and /l-bidi/ ‘DEF-well’, are coded as ‘simple’. Simple nouns in this analysis also include borrowings such as /d-dangiʊ/ ‘DEF-chickpeas’, and /d-dira:jiʃ/ ‘DEF-windows’, along with forms of *mašdar* other than the *nomina verbi*; these are substantives and can be either in the singular or the plural form, e.g., /s-se:l/ ‘DEF-rain’. On the other hand, derivatives include nouns derived from other parts of speech such as: 1) the participles, e.g., /l-wa:lid/ ‘DEF-father’, and /l-makʰbu:s/ ‘DEF-traditional rice meal’, 2) place names, e.g., /l-mistaʃfi/ ‘DEF-hospital’, and /l-mida:ris/ ‘DEF-schools’, 3) nouns of intensity/profession, e.g., /d-dabba:ʃ/ ‘DEF-tanner/leatherworker’, and /l-jima:lah/ ‘DEF-camel riding’,²¹⁷ 4) nouns of instrument, e.g., /li-msaʒiqʷil/ ‘DEF-recorder’, along with 5) nominalised adjectives, e.g., /l-fa:ri/ ‘DEF-poor’, 6) diminutives, e.g., /li-jzajra/ ‘a toponym; lit. the small island’, and /l-ge:ma:t/ ‘a type of fried dough; lit. the mini bites’, 7) unit nouns, e.g., /li-nxaʃah/ ‘DEF-palm tree’, and /li-bgurah/ ‘DEF-cow’. Also, included in the derivatives are: 8) the *m*-verbal noun, ‘*l-mašdar l-mīmiyy*’, e.g., /mna:tʰaħ/ ‘a traditional bullfighting contest’, and 9) the nouns of instance, e.g., /tʰ-tʰaʃa/ ‘DEF-going outdoors’.

Inflection:

The morphological form is also coded in terms of the grammatical number based on the shape of the noun as: 1) singular, e.g., /r-rajja:l/ ‘DEF-man’, and /l-baħar/ ‘DEF-sea’, 2) broken plural, e.g., /d-dikʰa:kʰi:n/ ‘DEF-shops.PL’, and /ʃ-ʃita:wi/ ‘DEF-winter.PL’, 3) sound plural, e.g., /l-ja:r-a:t/ ‘DEF-neighbours-PLF’, and /l-lawwilijj-i:n/ ‘DEF-ancestor-PLM; lit. the first ones’, 4) dual, e.g., /θ-θin-te:n/ ‘DEF-two-DUALF’,²¹⁸ 5) generic singular, e.g., /l-ħili:b/ ‘DEF-milk’, and /l-ʃisal/ ‘DEF-honey’, and 6) generic plural /t-tamir/ ‘DEF-dates’, and /s-simakʰ/ ‘DEF-fish’;

²¹⁷ /l-jima:lah/ is a name of profession used to describe the work of porters who carry goods on camel backs.

²¹⁸ This is the only example of the dual in the data.

collective nouns which do not have a singular that is derived from the same root consonants, e.g., /Ø-bo:f/ ‘DEF-camels’, are coded as singular.

Countability

The tokens are also coded in terms of whether they are count, e.g., /r-rajja:l/ ‘DEF-man’, and /l-be:t/ ‘DEF-house’, or non-count, e.g., /l-ya:z/ ‘DEF-gas’ and /l-kahraba/ ‘DEF-electricity’.

Nouns that may have count and non-count usages, e.g., /l-γidi/ ‘DEF-lunch’, and /f-ʃiti/ ‘DEF-winter’, are coded according to the context but mostly found to be non-count.

Type of phrase (constituent structure)

The tokens are also coded in terms of the type of phrase as: 1) independent NP, e.g., /l-wa:ild.../ ‘DEF-father...’, 2) construct phrase, e.g., /hali Ø-mikʔa:n/ ‘people place.DEF’, 3) demonstrative NP, e.g., /haðe: s-se:h/ ‘DEM DEF-sandy inland’, 4) NP with a modifying adjective, e.g., /Ø-hiri:m li-kʔba:r/ ‘women.DEF DEF-old’. In order to avoid the overlap with the syntactic function factor group, prepositional phrases are not included as a factor in the type of phrase.

Position of phrase (functional structure)

In addition, the tokens are coded in terms of their position in the sentence, as in Table 7.6 (continued overleaf):

Syntactic function	Example
<i>Subject (agent, topic)</i>	/...awwal l-ʕi:d s ^s -s ^s iγi:r, n-na:s be:-ðibhu:n ha:ʃf-i wiħd-ih.../ ‘...in the past, at <i>Eid al-ḥiṭr</i> , the people would slaughter one goat [only]...’
<i>Subject of a relative clause</i>	/... n-na:s illa tibaaha.../ ‘the people you want’

<i>Subject of a subordinate clause</i>	/...lak ⁱ anna l-jima:ʕa mina:xa:t jsawwu:n.../ ‘but the group [invited guests] set campsites’ ²¹⁹
<i>Subject of a passive verb</i>	/... t-tamir jinðʕad , ja:klu:n minn-i n-na:s.../ ‘the dates are harvested, people eat them’.
<i>Delayed subject</i>	/...l-ħi:n mitnawʕa:t l-asa:mi / ‘nowadays, the names [of camels] are varied’.
<i>Oblique object</i>	/... manfiʕ-at l-bugra tqa:rib ila n-naxlah / ‘the benefit of the cow approximates that of the palm tree’; /..... firwa he: l-majlas .../ ‘like this guestroom’; /...fri:-ha min d-dikⁱka:n .../ ‘buy it from the shop’.
<i>Adjunct</i>	/...awwal l-ʕi:d s^s-sⁱyi:r n-na:s be:-ðibhu:n ha:jf-i wiħd-ih.../ ‘...in the past, at <i>Eid al-ŕitr</i> , the people would slaughter one goat [only]...’.
<i>Direct object</i>	/awwal na:s faq:ra, tiq ⁱ lid haði: li-ħba:l ma:l n-nixi:l.../ ‘in the past there were poor people who braid ropes [made] of palm tree [fibres]’; /twarr ha:ði:kⁱ r-raʕbu:bi min he:n tku:n/ ‘that seashell is thrown as far as it can land’.
<i>Indirect object</i>	/...jgarri:nni s-suwar wa:ħad wa:ħad/ ‘[the <i>mʕallim</i>] ²²⁰ makes us recite the [Quran] verses, one by one’.
<i>Proposed object (CLLD)</i>	/... l-xammu:d tsawwi:la bahbu:ħ fil-arðʕ.../ ‘the hole [for the <i>wallam</i> game] ²²¹ , you make for it a gap/pocket in the ground’.
<i>Other functions:</i>	
<i>Predicate</i> ²²²	/...he: l-lawilijji:n , ha:ði ħija:ttum.../ ‘these are [our] ancestors and this is their life’; [...ha:ji brikat l-ħurma l-wuħdeh] ‘this is the blessing [benefit] of one woman’.
<i>Negated NP</i>	/... ab 0-0o:r .../ ‘not the ox’; /... ma: 0-xajja:t^t l-ʕarabi l-ħi:n/ ‘not [the] <i>ʕarabi</i> Tailor [name of a brand] of nowadays’.
<i>Analytic ~ periphrastic genitive</i>	/...0-bo:ʕ haðe:la ma:l r-rakið^s maʕzu:la:t.../ ‘... [the] camels for the camel running event are isolated [from the rest of the camels in the place where they are kept]...’.
<i>Coordinated conjunct</i>	/...jo:b be:n z-zu:lijj-i w- li-jda:r / ‘...roughly between the rug and the wall...’.

Table 7.6 Coding for the syntactic function factor group.

²¹⁹ Where they can rest with their camels during a traditional event.

²²⁰ A person who teaches the recitation of Quran to children.

²²¹ The *wallam* is a traditional game where the players throw a small stick made out of palm leaves which is about 15cm long in the air by using another but longer piece of stick (like a baseball bat). They place the smaller stick on a small hole made in the ground called *xammūd* and then they put the longer stick underneath it to carry it and hit it simultaneously.

²²² The predicate category here is analogous to the ‘classifactory’ and ‘identifactory’ predicate types only (Beeston 1970: 66); the other two, namely the ‘locatory’, and the statement of event, are included in the oblique object, adjunct, and object factors.

Table 7.7 provides a snapshot of how a token is coded in the final Excel sheet.

Token	overt/null	common/proper	concrete/abstract	type of noun	morph 1	morph 2	count/non-count	phrase type	function	larger context
ba:b	n	c	c	domestic	simp.	s	c	np	oblique obj.	...atʃo:f ʕa ba:b.
l-ima:ra:t	o	p	c	toponym	der.	pl	c	np	oblique obj.	...iʃtiyaʔit fil-ima:ra:t
r-rakið ^s	o	c	a	traditional occasion	verbal	s	n	np	analytic genitive	...bo:f haðe:la ma:l r-rakið ^s maʕzu:la:t...
xajja:t ^f l-ʕarabi	n	c	c	human	der.	s	c	n- adj	negated NP	...ma: xajja:t ^f l- ʕarabi l- hi:n...
ajja:am ʃ-fti	o	c	a	season	simp.	s	n	comp	adjunct	...baʕad ajja:am ʃ-fti, nqarrib hni:h...

Table 7.7 Sample of the token coding.

A note on coding for the DEF variable

The coding of the linguistic constraints itself is guided by a number of general references on the Arabic language, such as Wright's Arabic grammar (3rd edition), Ryding (2005), Holes (2004) on modern Arabic varieties, Holes's (2001) glossary of east Arabia, the *Encyclopaedia of Arabic Language and Linguistics*, *ǧāmiʿ d-durūs l-ʿarabiyya*²²³ by l-Ġalāyīnī (1912-1913,

²²³ <http://islamport.com/d/3/lqh/1/77/1071.html>

investigated by Shabbara 2010 edition), along with Arabic-Arabic dictionaries like *mu‘ğam l-luġa l-‘arabiyya l-mu‘āšira* and the online dictionary *almaany.com*²²⁴ which gathers definitions from other Arabic-Arabic dictionaries, like *‘al-mu‘ğam l-wasīf*, and *lisān l-‘arab*.

In addition, generally speaking, the tokens are coded based on their ‘citation forms’ or ‘base forms’ (Ryding, 2005: 119). So, for instance, to decide on the morphological form of the word /bi:ba:n/ ‘door.PL’ for the first category that concerns whether the noun is derivative or verbal or simple, the decision on which shape this noun has is based on the base singular ‘unmarked’ noun /ba:b/ ‘door’.

7.3.2 Modelling process

This section sums up the modelling that led to the final model on which the results section is based. The application value in all of the modelling stages is OVERT. At the beginning of the analysis, I ran all the ‘raw’ unconflicted factor groups in Rbrul together to get a feel of which factor groups are most important and try to identify sources of interaction/overlap. This was followed by a series of individual runs, and cross-tabulations carried out for each of the linguistic constraints independently first, and then for the interaction between them, in order to obtain a fine-grained look at the data, in terms of how the dependent variable is distributed and the main sources of interaction that may lead to overlap and thus skew the analysis.

Results from the cross-tabulations for the independent linguistic constraints confirmed initial observations on sources of overlap in the data and shed light on some more. For example, the morphological form factor group overlaps with the morphological number and countability factor groups, since the tokens in the ‘verbal noun’ factor are also coded as ‘singular’ and ‘noncount’. This factor also overlaps with the ‘traditional practices’ factor in the type of noun

²²⁴ <https://www.almaany.com/ar/dict/ar-ar/>

category. There was considerable overlap caused by the factors with low Ns in general, e.g., between the type of noun category and some of the other factor groups, particularly with the syntactic function.

Based on the cross-tabulations and initial runs, some conflating was done in order to better balance the token distribution and to decrease the overlap/interaction between the different factor groups. Certain factors/factor groups were left out at initial stages, but then reintroduced as different combinations at later stages. The regrouping decisions were guided by 1) the proportion of the application value, in this case OVERT, for the individual factors, 2) the similarity between the factors, and 3) the log odd values and factor weights for different values in order to avoid blurring the effect of any factor as much as possible. For instance, the morphological form factor group had an unbalanced token distribution ('derivative': N=170; 'verbal noun': N=42; 'simple': N=588) and caused overlap; the overlap again is caused by the 'verbal noun' factor, so this factor was excluded earlier on in the modelling process and then was re-introduced at the final stages, conflated with the 'simple' category as 'other form', since both favour the application value. The morphological number factor group is also conflated first as 'plural', 'singular', 'generic plural', and 'generic singular', where the 'plural' category included the 'dual', the 'broken plural', and the 'sound plural' factors. The former four factors were not convenient for the modelling also due to the imbalance in token number ('singular': N=576; 'plural': N=142; 'generic singular': N=47, and 'generic plural': N= 35), and due to the overlap with other factors, particularly the 'count' factor; so, later on, the 'singular generic' and the 'plural generic' were conflated into one factor 'generic'. One last configuration for the morphological number was the 'singular generic' with the 'singular' and the 'plural generic' with the 'plural'. In any case, this category was not chosen to be significant in most of the runs and models, and thus, was excluded in some of the models; however, it was reintroduced in the final model as 'generic', 'singular' and 'plural'.

In addition, the type of phrase, which is found significant in all of the models, was conflated at early stages to reduce the gap between the factors as well; the ‘demonstrative NP’ is grouped with the ‘Noun-adjective’ factor as ‘modified NP’. Another factor group chosen as significant, or rather most significant, by the different runs is the type of noun category. There were initially 18 values in this group (see Table 7.5); some of which had low token numbers, as low as $N=4$, and $N=6$. This required changes in the grouping, which happened at three different levels during the analysis. In the beginning the low token factors, namely ‘personal name’, ‘nationality/ethnicity’, ‘special category’, ‘weekday’, ‘numeral’, were grouped together as ‘other noun types’. Then, ‘season’, ‘time adverbial’, and ‘place adverbial’ were grouped together as ‘adverbial’; ‘camel’ and ‘other animal’ were grouped together and later conflated with ‘domestic’. Also, ‘topographical feature’ was grouped together with ‘toponyms’. At this stage, there were nine values. Later on, the ‘food item’ was conflated with the ‘other noun types’; this group eventually also included ‘other abstract nouns’, and ‘other inanimate nouns’ leaving us with six values altogether. The third and last major change to the type of noun category was combining ‘traditional practices’ with ‘domestic’, and although the R^2 number is slightly higher for the same model when these two are combined, eventually the decision taken was to leave them separate.

The syntactic function went through similar processes: ‘subject of a passive verb’, and ‘subject in a relative clause’ were conflated with ‘subject’; ‘subject of a subordinate clause’, and ‘delayed subject’ were conflated together as ‘other subjects’; ‘indirect object’ and ‘pre-posed object’ were grouped together as ‘other objects’; the last two object combinations were conflated as ‘object’, and at a later stage in the analysis, all of the subject functions were conflated in one group, ‘subject’, thus amounting to five values altogether for this factor group. The final model chosen for the multivariate analysis is summarised in Table 7.8 below:

Linguistic constraint	Factors
Noun Type	traditional practices, topography, human, domestic, adverbial, other noun types
Morphological form	derivative, other form
Countability	count, non-count
Phrase Type	NP, construct phrase, modified NP
Syntactic function	subject, object, oblique object, adjunct, other functions
Morphological number	singular, plural, generic

Table 7.8 The final model for the multivariate analysis.

7.3.3 Results and discussion

Table 7.9 (continued overleaf) summarises the results of a step-up/step down analysis in

Rbrul for the DEF variable:

Application value=OVERT; N=800; overall proportion= 85.6%; input probability= 0.837 R ² = 0.278; log likelihood= -276.693				
Linguistic predictors	Log odd	FW	Tokens	OVERT %
Noun Type (p< 0.0001) (p=4.34e-09)				
Other noun types	1.200	0.769	326	94.5%
Human	0.343	0.585	98	86.7%
Adverbial	0.330	0.582	77	85.7%
Domestic	-0.189	0.453	145	81.4%
Traditional practices	-0.815	0.307	76	68.4%
Topography	-0.870	0.295	78	71.8%

Phrase Type (p< 0.0001) (p=6.27e-06)				
Construct phrase	0.713	0.671	113	88.5%
NP	0.305	0.576	580	88.3%
Modified N	-1.018	0.265	107	68.2%
Countability (p< 0.005) (p=0.00218)				
Non-count	0.402	0.599	322	91.9%
Count	-0.402	0.401	478	81.4%
Syntactic Function (p< 0.05) (p=0.0488)				
Objects	0.518	0.627	136	91.2%
Subjects	0.360	0.589	229	90.4%
Oblique objects	-0.087	0.478	241	84.6%
Adjuncts	-0.199	0.45	111	78.4%
Other functions	-0.592	0.356	83	75.9%

Table 7.9 Results of the DEF variable: linguistic constraints.

With an overall probability of 85.6%, the application value OVERT is highly favoured in the data of the two informants; about 14% of the tokens are NULL. In terms of the internal constraints, all but morphological number and morphological form are found to be significant in predicting variation for the DEF variable. The most significant constraint is the type of noun, followed by the type of phrase, countability, and lastly the syntactic function. The first two factor groups are very salient throughout the modelling stage; they were almost always found to be significant. All of the noun types coded for are found to favour overt, apart from ‘domestic’, ‘traditional practices’, and ‘toponyms’; the latter one disfavours the application value at 0.295, which is relatively low compared to the factor weight of the ‘other’ category (FW=0.769) which favours OVERT the most and exhibits the highest proportion of OVERT (94.5%) among all other factors in the model. With regard to phrase type, the ‘construct

phrase’ is found to favour the application value the most at 0.671, followed by the independent ‘NP’ which also favours the application value at 0.576; the last factor, ‘modified N’, with 68.2% of OVERT, disfavours OVERT at 0.265, which is the lowest factor weight within the values for the whole model. In addition, countability is selected to be significant with count nouns favouring OVERT (FW=0.599) and non-count nouns disavouring it (FW=0.401). The last predictor, syntactic function, is found to be the least significant with a p -value=0.0488. Objects and subjects both favour OVERT, whereas the rest disfavour it; the ‘oblique object’, although is found to disfavour OVERT, it is almost neutral at FW=0.478.

Furthermore, although not chosen as statistically significant in the final model, counts for the morphological number are displayed in Table 7.10 below reveal that the amount of variation in the ‘singular’ category (17%) is almost twice as that of the ‘plural’ (8%) and ‘generic’ (9%) categories.

Morphological number	NULL	OVERT	Total
Generic	9% N=7	91% N=75	10% N=82
Plural	8% N=12	92% N=130	18% N=142
Singular	17% N=96	83% N=480	72% N=576
Total	14% N=115	86% N=685	N=800

Table 7.10 Distribution of the dependent variants per the morphological number.

Also, Table 7.11 shows that there is more variation in the ‘derivative’ set with 78.8% OVERT than in the conflated ‘other form’ set, with 87.5% OVERT.

Morphological form	NULL	OVERT	Total
Derivative	21.2% N=36	78.8% N=134	21.3% N=170
Other form	12.5% N=79	87.5% N=551	78.8% N=630
Total	14% N=115	86% N=685	N=800

Table 7.11 Distribution of the dependent variants per the morphological form.

In addition, since a model with both external and internal predictors for the forty informants was not carried out, it is not possible to compare these two sets of predictors against each other in one model to test for which is more significant in terms of the constraints on the variation. The two informants chosen for the modelling of the linguistic constraints belong to the same age and locality group; only gender is variable. I carried out another run with the same model shown in Table 7.12.

Application value=OVERT; N=800; overall proportion= 85.6%; input probability= 0.821 R2= 0.317; log likelihood= -271.221				
Linguistic predictors	Log odd	FW	Tokens	OVERT %
Noun Type (p< 0.0001) (p=2.16e-10)				
Other	1.213	0.771	326	94.5%
Human	0.610	0.648	98	86.7%
Adverbial	0.205	0.551	77	85.7%
Domestic	-0.227	0.444	145	81.4%
Topography	-0.881	0.293	78	71.8%
Traditional occasion/event	-0.920	0.285	76	68.4%
Phrase Type (p< 0.0001) (p=4.11e-07)				
Construct phrase	0.658	0.659	113	88.5%
NP	0.449	0.61	580	88.3%
Modified N	-1.107	0.248	107	68.2%
Gender (p< 0.0001) (p=4.33e-05)				
Male	0.466	0.614	406	89.9%
Female	-0.466	0.386	394	81.2%
Countability (p< 0.005) (p=0.00296)				
Non-count	0.407	0.6	478	91.9%
Count	-0.407	0.4	322	81.4%
Morphological Form (p< 0.05) (p=0.0428)				
Other form	0.284	0.571	630	87.5%
Derivative	-0.284	0.429	170	78.8%

Table 7.12 Results of the DEF variable: gender included.

Gender is found to be significant; it came third in terms of significance after the type of noun and phrase type, at a p -value < 0.0001 . The male informant favours OVERT (FW=0.614),

whereas the female informant disfavours it at a factor weight of 0.386. The morphological form is also selected as significant this time at a p -value <0.05 , where the ‘other form’ favours OVERT at $FW=0.571$ and the ‘derivative’ disfavours it at $FW=0.429$. On the other hand, the syntactic function is not selected as significant in this model.

One last set of tables to include here are those of the noun categories ‘common’ vs. ‘proper’ and ‘abstract’ vs. ‘concrete’; although these two are not included in the multivariate analysis,²²⁵ counts for these two categories are presented to get an idea about the distribution of the dependent variable per these two major classifications of the noun, and also because it seemed a necessary step in order to be able to see the bigger picture of the effect of the noun type in general on the variation at hand, since a more refined categorisation and modelling of this factor group may blur the more basic distinction.

Type of bare noun	NULL	OVERT	Totals
Common	14% (N=103)	86% (N=618)	90% (N=721)
Proper	15% (N=12)	85% (N=67)	10% (N=79)
Totals	14% (N=115)	86% (N=685)	N= 800

Table 7.13 The distribution of the DEF variable per the type of bare noun category: ‘common’ vs. ‘proper’.

Table 7.13 shows, interestingly enough, that the common nouns exhibit more or less the same pattern of variation as that of the proper nouns with 86% and 85% of OVERT compared to 14% and 15% of NULL respectively. This is also, quite unexpectedly, a similar case to the ‘concrete’ vs. ‘abstract’ set (Table 7.14), where the overall proportions of the dependent variants within the two values are comparable, although ‘abstract’ is anticipated to show more OVERT than

²²⁵ Quite expectedly, according to a previous hypothesis about the nature of the distribution of the NULL variant that bigger categorisation would mask the distinction within these groups.

‘concrete’, since in Arabic in general, abstract nouns are reported to normally take the definite article (Brustad, 2000: 23).

Type of bare noun	NULL	OVERT	Totals
Abstract	16% (N=41)	84% (N=219)	33% (N=260)
Concrete	14% (N=74)	86% (N=466)	68% (N=540)
Totals	14% (N=115)	86% (N=685)	800

Table 7.14 The distribution of the DEF variable per the type of bare noun category: ‘abstract’ vs. ‘concrete’.

Having reported the statistics from the multivariate analysis, in the next section I take a closer look at the distribution of the dependent variable across the linguistic categories, and I discuss the different trends noticed here and in the whole dataset in general in an attempt to come up with some interpretation on what is going on with the DEF variable in this speech community. At this point, it is quite useful to look at the individual counts, and the interaction between the different linguistic constraints to help thinking in terms of which factors could be seen as essential in constraining the use of the different variants. In terms of the function of the noun phrase, the ‘object’ and ‘subject’ categories have the least variation with 91% and 90% of OVERT respectively. What Table 7.15 probably reveals is that the object and subject are highly constrained syntactic positions that require an overt article; this echoes the observation about topics and generic nouns in Arabic being a realm of the overt definite article. The ‘oblique object’ factor is heading in the same direction of marking the definite article overtly. On the other hand, adjuncts show a higher degree of variability, stressing that this position is not as syntactically salient as the former ones and thus the pressure to show definiteness overtly is not as high.

Syntactic function	NULL	OVERT	total
adjunct	22% N=24	78% N=87	14% N=111
oblique object	15% N=37	85% N=204	30% N=241
object	9% N=12	91% N=124	17% N=136
other functions	24% N=20	76% N=63	10% N=83
subjects	10% N=22	90% N=207	29% N=229
total	N=115	N=685	N=800

Table 7.15 Distribution of the DEF variable per syntactic function.

Because of the varied nature of the ‘other function’ factor, we cannot make a claim as to its status in relation to the variation at hand, but a closer look at the tokens within this group reveals that tokens which function as ‘predicates’ (N=28) are mostly realised with NULL, whereas the ‘analytic genitive’ construction (N=11) is categorically realised with OVERT; this takes us to the type of phrase factor group, where the ‘construct phrase’ triggers the use of the overt article in that 88% (N=100) of the total (N=113) is realised with OVERT, which is understood in the light of the nature of the Arabic construct state where the overt definite article is a salient feature of the ‘annexed’ common noun (see § 6.1 in Chapter Six). The ‘simplex NP’ behaves like the former factor in that it is moving towards the OVERT. On the other hand, tokens in the ‘modified NP’ show greater variability. Setting the ‘demonstrative-noun’ combination aside, the numbers in the ‘noun-adjective’ strings are quite interesting though not unexpected considering the literature on this type of phrase. This is probably the only category, besides the ‘predicate’ function, which has more NULL (57%, N=26) than OVERT (43%, N=20) tokens.

Another interesting distributional point relates to the morphological form. It is established that non-count generics in languages such as French and Spanish require a definite article, but the opposite is true for English (Danon, 2002: 44), although, statistically speaking, I have only analysed genericity in terms of the traditional treatment of morphological number in Arabic, i.e., including the ‘generic singular’ vs. the ‘generic plural’ as morphological number categories, it seems that the ‘generic singular’ is a high-ranking constraint, since all of the tokens (N=47) in this factor, like ‘honey’ and ‘sugar’, are categorically realised with OVERT. Another seemingly high-ranking constraint is the morphological form of the noun, where ‘verbal abstracts’ (N=42) as well are categorically realised with OVERT; this also applies to other tokens coded as ‘other abstract’ in the type of noun category which highly favour the overt article in line with the literature on this type of nouns. Both values ‘singular generic’ and ‘verbal abstract’ are ‘non-count’, and the fact that they are categorically realised as OVERT adds to the salience of the ‘non-count’ nouns in the overt marking of the definite article; this is reflected in the results from the multivariate analysis as well where ‘non-count’ nouns show less variability than ‘count’ ones, although the latter category is catching up with 81% of the tokens being overtly marked. I leave commenting on the type of noun category for now, since it seems best to be analysed in terms of the larger dataset for the whole sample.

The above observations are based on the data from the two speakers with the highest NULL tokens, and for which the NULL category could be considered as still functional. We have seen some categories that prefer overt marking of the definite article and others which have higher proportions of NULL. However, for a more complete analysis of NULL as a definiteness feature in the speech community, we need to look at the nature of the linguistic distribution of the NULL category across the external predictors in the whole dataset (N=543),

but more importantly across the three age groups, which the following section attempts to tackle.

NULL in the whole dataset

Table 7.16 (continued overleaf) shows the distribution of NULL for the whole dataset per age group and type of noun categories. The focus on the noun type will be clearer in the discussion, but one motivation is that in the above multivariate modelling and analysis, this factor group has been very salient, almost always chosen to be the most significant of all linguistic constraints.

Noun type ²²⁶	Middle	Old	Young	Total
Adverbial noun	9% (N=11)	8% (N=22)	7% (N=9)	8% (N=42)
Camel	0%	18% (N=51)	2% (N=2)	10% (N=53)
Domestic	7% (N=9)	14% (N=42)	2% (N=2)	10% (N=52)
Food item	2% (N=3)	3% (N=8)	1% (N=1)	2% (N=12)
Human	1% (N=1)	7% (N=21)	1% (N=1)	4% (N=23)
Other proper nouns	5% (N=6)	1% (N=1)	2% (N=2)	2% (N=10)
Other abstract nouns	6% (N=7)	4% (N=11)	0%	3% (N=18)
Other animals	0%	4% (N=11)	0%	2% (N=11)
Other inanimate nouns	3% (N=4)	9% (N=25)	6% (N=8)	7% (N=37)
Scholastic term	28% (N=35)	2% (N=7)	35% (N=45)	16% (N=87)

²²⁶ Percentages for the noun types within the different age groups are out of the total of each group individually, but the percentages for the totals both on the right and the bottom of the table are out of the total (N=543).

Season	0%	2% (N=7)	1% (N=1)	1% (N=8)
Special category	3% (N=4)	1% (N=2)	0%	1% (N=6)
Topographical feature	0%	5% (N=14)	1% (N=1)	3% (N=15)
Toponym	15% (N=19)	10% (N=29)	23% (N=29)	14% (N=77)
Traditional practices	21% (N=27)	13% (N=38)	21% (N=27)	17% (N=92)
Total	23% (N=126)	53% (N=289)	24% (N=128)	(N=543)

Table 7.16 Distribution of NULL for the whole dataset per age group and type of noun.

Setting the age group aside for now, what Table 7.16 displays is a clustering of NULL tokens in noun phrases that relate to schooling, traditional practices, toponyms, camel, domestic lifestyle, and to a lesser extent adverbials. A closer look at the type of phrases in these categories (see Appendix D) reveals that ‘scholastic term’ and ‘traditional practices’ predominantly have a ‘noun-adjective’ pattern, a pattern associated with the NULL or zero category in the literature (e.g., Turner, 2013; Krug and Lucas, 2018); and as it happens, the adjective is also predominantly a modifying numeral. When this is the case, either only the modified noun lacks the overt marking or both the numeral and the noun (*cf.* Turner, 2013: 48). In other noun-adjective strings where the adjective is mainly attributive, the adjective would always show the overt marking, whereas the noun does not. In general, most of the above noun types do mainly occur as independent ‘simplic’ NPs.

With regard to the age groups, the three groups more or less pattern similarly with regard to toponyms, and traditional practices, although toponyms constitute the highest concentration of NULL tokens in the younger group’s data (23%). However, the older age group departs quite considerably from the other two in terms of what seems a more conservative usage of the null definite, which occurs in simplex NPs with nouns that relate to

the traditional lifestyle and camel culture (18% for ‘camel’, and 14% for ‘domestic’). The middle age group’s NULL tokens are mainly nouns that relate to schooling (28%). So, what can be concluded so far is that noun-adjective phrases with NULL, as well as toponyms and traditional practices, are generally more tolerated, and that speakers in the older age group also use this category within contexts not used by the other age group.

Why NULL: possible explanations²²⁷

A possible analysis of noun-modifier constituents, adverbials, and toponyms

With many noun-adjective constituents, like /ʕi:d s^s-sⁱʕi:r/ ‘the small Eid’, /ʕi:d li-k^bi:r/ ‘the big Eid’ and /ħiri:m li-k^ba:r/ ‘older women’, it seems that they have been re-analysed as compounds or constructs, rather than noun-modifier constituents.²²⁸ For the other group of tokens with noun-adjective phrases where both elements lack overt marking and where the overt marking is only shown on the adjectives, especially with terms related to schooling like /s^saff l-awwal/ ~ /s^saff awwal/ ‘first grade’ as produced by the younger and middle groups’ speakers in particular, the use of the definite article is probably best viewed as ‘optional’.

Another possibility as to why phrases with an ordinal modifying noun appear with no article on both the numeral and the noun may be that the speaker views such constructions as

²²⁷ I would like to acknowledge two of my colleagues, Dr Uri Horesh and Sharaf Yassin, for their valuable contribution to the discussion on the possible scenarios for the NULL category.

²²⁸ One wonders whether /ʕi:d s^s-sⁱʕi:r/ ‘small Eid’ and /ʕi:d li-k^bi:r/ ‘big Eid’ have lost the marking on the head noun by analogy to the borrowed standard construct phrase equivalents /ʕi:d l-fit^r/ and /ʕi:d l-að^ħa/ respectively. This sounds a likely explanation which can be supported by the data from Bahraini Arabic (Holes 2016: 214), in which a modified NULL feminine head noun is phonologically marked with an overt *-t* feminine ending, e.g., /Ø-sin-t l-dʒa:j-a/ ‘year.SGF.DEF DEF-come.PTCP-SGF; literally: the coming year’, in the same way an annexation relationship between a feminine head noun and another ‘definite’ noun is marked, e.g., /madars-at l-ban-at/ ‘school-SGF DEF-girl-PLF; literally: the girls’ school’ (ibid: 215; the transcription is mine). This *-t* is also found for similar constructions in Procházka’s data on Harran-Urfa in the Northern Fertile Crescent (2018: 286). It is noteworthy though that, in the dialect under study, the *-t* marking of a modified NULL feminine head noun does not occur, e.g., /Ø-ħis^ss^s-a r-ra:b^ʕ-a/ ‘period DEF-SGF DEF-fourth-SGF; i.e., the fourth period’ and /Ø-be:s-e s-o:d-i/ ‘baisa.DEF-SGF DEF-black-SGF; i.e., the black baisa’.

‘inherently definite’ and thus ‘redundant’ (*cf.* Krug and Lucas, 2018: 267 for English modifying ordinal-noun phrases).²²⁹ Optionality of OVERT or NULL marking also seems to be characteristic of adverbials where the use of the NULL category is comparable for the three age groups (see Table 7.16). This is a similar case to the treatment of nouns for seasons in English, e.g., ‘in summer’ vs. ‘in the summer’, which occur ‘in free variation’ or with the zero article when used generically (*ibid.*: 266).

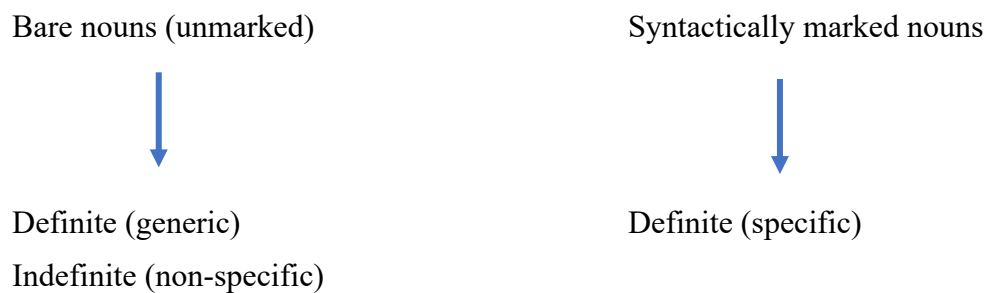
In addition, for toponyms in general, the loss of the article could be correlated to the fact that these are proper nouns which are inherently definite and for which the article would be redundant; nonetheless, this is not the whole story. In Oman, toponyms must have undergone standardisation at some point when many local place names previously used without the article were assigned the prefix *l-*. The informants probably prefer using these ‘vernacularised’ or ‘traditional’ local place names over the new forms, and thus, the variation and the relative high frequency of NULL with local place names.

These observations can so far lead us to conclude that the overt marking for such nouns is also an optional category which has not yet grammaticalised in this dialect; this is a quite plausible direction to take at this point considering that NULL in other languages like English is also not fully grammaticalised in some noun types but not in others (e.g., Krug and Lucas, 2018; Berezowski, 2009). On the other hand, the results of the linguistic multivariate analysis summarised in Table 7.9 shows that the overt marking is preferred with other types of proper nouns, e.g., personal names, and weekdays; this observation could also be deduced from the results for this category in the whole dataset, since there are very few tokens (2%, N=10/543) of other proper nouns with NULL, see Table 7.16.

²²⁹ A similar stance is adopted by Procházka (2018: 268) for the existence of such construction in some varieties of the Northern Fertile Crescent, “An important factor in explaining the ‘success’ of the development described here is the complete redundancy of the article attached to the noun—a fact which certainly facilitated its omission.” (*ibid.*).

A possible analysis of generic bare nouns with NULL

Many instances of the words for animals in the speech of the older group actually refer to ‘kind’, i.e., bare nouns functioning as generic nouns, not nouns referring to specific or individual entities. Doron (2003) argues that Hebrew and Brazilian Portuguese exhibit a similar feature.²³⁰ Arabic, however, requires generic nouns, although semantically indefinite, to show grammatical definiteness through the use of the definite article (Marogy, 2010: 101).²³¹ We could simply argue that in this dialect, generic bare nouns or NPs with generic bare nouns could occur as simply indefinite or as definite with a generic reading, and syntactically marked nouns to be definite:



While a good number of NULL bare nouns, in the older group’s data, have a generic reading, the above account while it provides an explanation for NULL generic bare nouns, it is very partial, because: a) it ignores the fact that in real use, NULL generic bare nouns are in variation with generic bare nouns marked with the definite article, and b) it fails to account for NULL non-generic bare nouns like ‘house’ and ‘door’, and c) it fails to account for generic

²³⁰ Doron’s (2003) argument is for languages that do not have the definite article, but also for Hebrew which has it.

²³¹ “Although Generic nouns in English are pragmatically definite but syntactically indefinite, it is a language particular fact about Arabic that generic nouns are nearly always accompanied by a the prefixed *lām al-jins* (the generic *lām*)”.

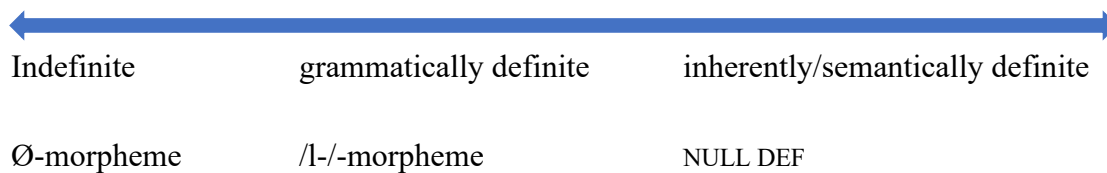
bare nouns which do have specific reference, i.e., referential definite(s), but which are realised with NULL. So, what could possibly be going on with generic and non-generic common nouns with NULL in this dialect?

One, probably main, extralinguistic motivation is the salience of such nouns in terms of their ‘culturalness’ and their association to the speech community’s ‘traditional lifestyle’. Such nouns as /Ø-bo:f/ ‘camel.PL.DEF’ and /Ø-be:t/ ‘house.SGM.DEF’ could be viewed by the speakers of the dialect, primarily the ones in the older age group, as cognitively and textually salient and or inherently definite, in the same way proper nouns are, and thus, the speakers may consider the use of the overt marker as redundant. The next two sub-sections elaborate on this argument.

Definiteness hierarchy

What could have happened diachronically is that the speakers, due to the cultural ‘coreness’ of nouns like /Ø-bo:f/ ‘camel.PL.DEF’ and /Ø-be:t/ ‘house.SGM.DEF’, have re-defined or modified their definiteness continuum. Unlike the standard and other varieties of Arabic which show a dichotomy in their marking of the DEF category: definite (marked through the use of the prefix *l-* with the above common nouns) and indefinite (marked through the use of nunation in the standard and Ø-morpheme in the dialects), this dialect has three levels of marking definiteness (the former two in addition to the null definite morpheme Ø). The hierarchy within the continuum of definiteness itself in this dialect requires a thorough analysis. However, one likely possibility is that nouns with NULL are inherently definite in the speakers’ cognition, and thus, are more definite than other nouns; that is to say, they probably are as definite as proper nouns which do not necessarily have to mark the article overtly, or simply are viewed as proper nouns and by analogy have developed to become inherently

definite. So, in the same manner that a personal name such as /aħmad/ ‘Ahmad.DEF’ is higher definite than a common noun such as /l-walad/ ‘DEF-boy.SGM’, and thus the use of an overt article could be found redundant, for the older informants at least, a word like /Ø-bo:j/ ‘camel.PL.DEF’ is a higher definite than another common noun like /t-taʕli:m/ ‘DEF-education’, and thus, they use NULL to mark such hierarchy. Based on this hypothesis, a basic version of the definiteness continuum for the dialect would look like the following:



We can equally hypothesise that in this dialect, the underlying form of the semantically definite nouns is /l-/+N, but the article does not surface in the realisation of some nouns like highly culturally specific nouns and local place names. A culturally instilled cognitive status of NULL could also be the reason it can occur in textually prominent positions in the discourse, such as a subject position in an SVCOMP clause as shown in § 6.6 in Chapter Six.

Incomplete grammaticalisation (+DEF)

We can imagine a situation where a process of grammaticalisation of NULL, /l-/>NULL, started but was reversed due to contact. Within this view, we can argue that the $N_{\emptyset}+Adj_{DEF}$, and $N_{\emptyset}+Adj_{\emptyset}$ patterns are relic, a result from different stages in the development of the definite article in some Arabic varieties and other Semitic languages which have been reported to exhibit such patterns (Pat-El, 2009). The fact that these forms are still productive and are in variation with $N_{DEF}+Adj_{DEF}$ suggests that the grammaticalisation of the article is not complete

yet for this category. This may also apply to other categories where there is an optionality of the usage of the article.

Overall, there are two available scenarios for the grammaticalisation of +DEF in this dialect. The first is that diachronically, the NULL is the historical form of the definite state, and that the overt article was introduced due to long-term contact with other dialects with the overt *l-* and, as a result, null definite constructions became *localised* synchronically. Most people may have lost the intuition as to when to use it, and NULL has become a ‘marked’ feature in the region. This scenario is appealing, but not quite plausible because it is hard, if not impossible, to ascertain that the dialect has NULL as the primary definiteness marking, and as a variety of Arabic, it is hard to argue that it didn’t have an overt definite marker to begin with. This may be true of an ancestor language or a major donor language like proto-Semitic or proto-Central Semitic, but for a dialect that started as an Arabic variety, a descendant of some sort of a proto-Arabic or Old Arabic which presumably developed an overt marker earlier on, it would seem a far-fetched possibility.

The second scenario is that the NULL category is an archaic form which has always been quite restricted in distribution, and which may have started as an innovation, when some lexical items lost the definite article, due to familiarity, commonality, or even due to some sort of a substrate influence. However, synchronically, the definite article is re-introduced to these items due to levelling and contact. This speculation is less extreme and could be backed by the fact that the overt marker in the dialect has been well-established, considering its current distribution, and that wholesale borrowings of morpho-syntactic elements are not common cross-linguistically (see for example Gardani, 2018). In addition, the suggestion that the NULL category is an archaic form restricted in distribution finds support from outside the area. Procházka reports that in the Bedouin dialects in of the Northern Fertile Crescent, the null definite $N_{\emptyset} + \text{Adj}_{\text{DEF}}$ construction occurs in free variation with the overt $N_{\text{DEF}+} + \text{Adj}_{\text{DEF}}$

‘standard pattern’. However, in the sedentary varieties of the area, $N_{\emptyset}+Adj_{DEF}$ is the default system for *all* definite attributive phrases (2018: 267-8), which can be suggestive of the oldness of the NULL category, and that it is an ‘inherited’ Arabic feature; its existence in this area is ‘archaic’ and ‘pre-diasporic’ (ibid: 262). In fact, it can be argued that this ‘peripheral’ construction in the history of Arabic in general is extremely old, maybe even proto-Arabic, since it was attested for Old Arabic as well as Middle Arabic (ibid: 268, and the references therein).²³²

Regardless of the trajectory the NULL category has historically followed, what is certain is that, in this dialect, there may have been a stage where certain common nouns were “prefix-resistant” to use Turner’s terms (2013: 50); synchronically, these nouns are, generally speaking, in a state of variation, though it is not very clear to what extent. In addition, regardless of the historical process which led to the loss of the overt marking in the first place, that is if it has been lost at all, we can conclude that, synchronically, the overt article is on the way of becoming fully grammaticalised in a number of common noun types. In place names and other constructions where it seems to be more of an optional category, e.g., adverbials, and nouns and adjectives in noun-adjective strings, it has yet to reach this stage.

Why NULL is receding

Notwithstanding the position of NULL diachronically, synchronically, as a sociolinguistic variant, it is clear from the results so far that NULL is being levelled out. Contact disguised in the form of geographical and socioeconomic mobility, education, and population

²³² The null definite attributive construction was found in a Greek papyrus, dated AD 530, which includes the phrase *bayth al-axbar* ‘the very big first floor unit’ (Al-Jallad *et al*, 2013: 32, via Procházka, 2018: 286). Procházka also argues for the possibility of the null definite attributive phrases in some parts of the area being innovations rather than well-established archaisms (ibid).

demographics is a readily available explanation as to why this feature is being lost in this speech community. But importantly, this is likely the case due to this category being a marked feature, both linguistically, as we have seen above, and socially. Within the study area, NULL is generally viewed as a minority feature of certain groups or localities within the *Yāl Sa‘ad* tribe itself, particularly in al-Tharmad locality and the villages close to it in al-Suwaiq and al-Miṣin‘a; these groups are often described as /wa:jjid badu/ ‘very Bedouin’. A young woman from al-Khadhra locality commented on this point by saying:

...θ-θarmad akθar na:ss aḥissum mħa:fð^si:n ... ma: jgu:lu:n alif
w la:m t-taṣri:f, ma: ṣindhūm taqri:ban

‘[people] in al-Tharmad are the most conservative people...they do not say the *alif* and *lām* of definition [i.e., the definite /l-/], they do not have it probably.’ (R19FKH)

The informants also link the NULL category to the speech of older women. In fact, a younger woman from al-Tharmad locality itself commented that older women are the ones who preserve the /l-/ most often:

l-ṣija:jiz k̄i: jgu:lin... il-mħa:fð^sa:t ṣali:hin jadda:tni, al-ħi:n tara
ṣa l-ixtīla:t^t ṣwajji qallat ṣwajji, t^t-t^tawwurat, tiyajjirat

‘The older women use this [i.e., NULL] ...the ones keeping it are our grandmothers; right now, actually, because of mixing [with others], it [the dialect] is reduced; it has evolved; it has changed.’ (B29FTH)

The observations made by the informants quoted above are in line with the results of the current study which found that both al-Tharmad locality and older women exhibit a higher proportion and a wider variety of the NULL tokens. Older women and al-Tharmad seem to play the role of the ‘conservative’ groups or the ‘custodians’ of the traditional features of the

local dialect, such as the NULL category, and thus, reasserting the ‘Bedouin-ness’ of the speech community, which otherwise would be jeopardised. In addition, another significant point to raise here is the way the informants perceive the NULL category itself in order to understand their profiling of the former groups as ‘the ones dropping the /l-/’; what these informants clearly refer to is the type of NULL nouns where the overt marking is obligatory due to the syntax or the context, e.g., generic nouns in subject position, not to the noun types where the results have shown that the use of the overt marking or the null definite is optional; in fact, they themselves produce such NULL constructions unconsciously. Also, the fact that the former type of NULL is more salient for al-Tharmad locality calls for a refined intergenerational analysis of this feature for this locality in particular to test for the actual distributional range of the NULL category in the speech of younger speakers.

7.4 Summary

To sum up, the null definite is a traditional, possibly archaic, but now rare feature in the dialect under study. The results show that this feature occurs rather infrequently compared to the overt variant /l-/. For constructions and noun types where traditionally, the NULL category would be used, as judged by the behaviour of the older informants, the overt article is becoming the ‘new’ norm. The null definite is a more restricted category for the middle and younger age groups who generally prefer the overt variant. In terms of the external predictors for this study, men, both the young and middle age groups, and both al-Batha and al-Khadhra localities generally favour the OVERT variant. However, the social aspect of the variation with regard to the DEF variable is not straightforward. A mixture of overlapping external factors is at play; the external predictors of age, gender, and locality incorporate the type of education,

mobility, and work experience, all of which can be looked at as ‘proxies’ for contact and geographical mobility (Al-Wer, 2013).

In terms of the linguistic constraints, subjects, objects, construct phrases, independent noun phrases, and non-count nouns favour OVERT; within the noun type, domestic items, , toponyms, topographical features, and traditional practices all favour NULL; on the other hand, nouns referring to human referents, adverbials, food items, abstract notions, other than traditional practices, and ‘other’ inanimate entities favour OVERT. Furthermore, the results so far have emphasised that the cultural aspect of certain noun types plays a role in the use of the null definite. Generally speaking, common nouns with the null definite in the speech of older speakers are mainly ‘core’, ‘local’ or ‘vernacularised’ dialectal items that correlate with the Bedouin culture and traditional lifestyle. They are ‘salient’, and the speakers probably associate with them at a higher degree, making them ‘highly’ or ‘inherently’ definite’, a domain usually associated with ‘proper’ and ‘textually prominent’ nouns. In addition, there seems to be some ‘optionality’ in the choice of OVERT or NULL for most of the speakers when it comes to certain noun and phrase types, e.g., place names, and noun-adjective phrases that are associated with ‘traditional practices’ and ‘schooling’.

We can explain why some archaic features still survive in this speech community, despite the fact that it is located in transitional coastal towns like al-Suwaiq and al-Miṣin ‘a and is viewed as socially ‘marked’. For many decades if not centuries, the Bedouin populations in such towns have been closed to outsiders because of their tribal systems, except for the allies they agree to share their territory with. This, along with endogamous conservative marriage patterns, and a resilience to conform to and to intergenerationally transfer the Bedouin culture and identity, provide the perfect atmosphere for such linguistic features to survive. The reasons why the NULL category is probably being lost now, or more accurately being replaced by the overt definite article could be generally understood with

regard to changes to the above factors due to the socioeconomical development in the study area over the past few decades which brought about more accessibility, and facilitated mobility, and outside contact.

Conclusion

This study investigated sociolinguistic variation in the dialect of the Bedouin *Yāl Sa‘ad* tribe in northern Oman, namely in the neighbouring towns of al-Suwaiq and al-Miṣīn‘a. The thesis provides a brief sketch of Oman and the Bāṭīna region where the study area is situated. It outlined the relevant socio-linguistic profile of the study area; it presented an overview of the descriptive literature relevant to the Bedouin dialect under investigation and remarks on the phonology, morphology, morphonology, and morphosyntax of the dialect as spoken by the Bedouin community in the study area. However, the main focus of the thesis was investigating sociolinguistic variation in the speech of the *Yāl Sa‘ad* tribe through examining two variables, namely the (dʒ) variable (Chapters Four and Five) and the variation in the use of the definite article DEF (Chapters Six and Seven).

The first variable (dʒ) has two main variants, namely, the traditional glide [j] and a local variant [gʲ]. Results from the multivariate analysis show that the variation is constrained by a number of intra-linguistic and extralinguistic constraints. The traditional variant is quite salient; it occurred 71.9% of the time, whereas the stop variant occurred 28.1% in the dataset of N=2,499 ($R^2=0.31$; application value= [gʲ]). Linguistically, [gʲ] is mostly favoured in the context of a ‘preceding coronal’ (FW=0.67) or a ‘back vowel’ (FW=0.635), a ‘following consonant’ (FW=0.608), and in ‘polysyllabic’ words (FW=0.679). It is mostly disfavoured in the context of a ‘preceding palatal’ (FW=0.178), a ‘following front vowel’ (FW=0.346) and in ‘monosyllabic’ words (FW=0.342). This variant is also favoured in ‘light’ and ‘closed’ syllables. Stress is selected as insignificant in predicting the variation at hand. In terms of the social predictors, [gʲ] is mostly favoured by the ‘middle age group’ (FW=0.608), and ‘male’ informants (FW=0.578). The ‘older age group’ mostly disfavours the use of [gʲ] (FW=0.385), whereas the ‘younger age group’ informants can be best described as neutral (FW=0.507). It is

noteworthy that locality is selected as insignificant in predicting the variation in (dʒ), but descriptive statistics for locality revealed that ‘al-Batha’ shows slightly more variation than other two (31%, compared to 28% for ‘al-Tharmad’, and 25% for ‘al-Khadhra’.

The second variable, DEF, has two variants: OVERT and NULL, where syntactically definite substantives realised with the first variant show overt definiteness marking through the use of the definite article *-l*. In the dialect under study, these forms are in variation with the syntactically definite substantives that lack such marking. The analysis of this variable was done in two stages. The first stage targeted the social aspect of the variation. Descriptive statistics for the three external predictors, age, gender, and locality were carried out to the whole dataset (N=16,197), and a multivariate analysis was carried to a subset of the data to see the effect of these variables on the variation at hand (N=7,906). Results from the distributional statistics show that the overall proportion of NULL is 3% only in comparison to 97% for the OVERT variant. Al-Tharmad locality showed the most variation with 6% NULL tokens, whereas al-Batha showed the least variation with 1% NULL tokens. With regard to age, the older age group were the highest users of NULL (6%); the middle and the younger age groups behaved similarly with regards to the use of this variant (3% and 2% respectively). The multivariate analysis of a subset of the larger dataset (Application value=OVERT; $R^2=0.119$; N=7,906) generally reflected the distributional patterns found in the whole dataset, with the middle age group mostly favouring the use of OVERT (FW=0.648), followed by al-Batha (FW=0.593). On the other hand, al-Tharmad mostly disfavoured the use of OVERT (FW=0.349), followed by the older age group (FW=0.398). Men generally favoured it (FW=0.551), while women did not (FW=0.449).

The second stage of analysis involved an examination of the relevant linguistic constraints with regard to this variable in a dataset from two informants, a woman (NULL=19%) and a man (NULL=10%), who used the NULL category the most. They both come

from A-Tharmad locality and belong to the older age group. An Rbrul step-wise analysis was run for five types of linguistic constraints, namely, the noun type, the morphological form both derivational and inflectional, countability, phrase type, and syntactic function. Results from the analysis (Application value=OVERT; $R^2=0.278$; $N=800$) showed that the noun type is the most significant constraint followed by phrase type, then countability and syntactic function. The morphological form, both inflectional and derivational, is found to be statistically insignificant. The OVERT variant is mostly favoured in the ‘other noun type’ category, the ‘construct phrase’, ‘non-count’ nouns, and ‘objects’. On the other hand, ‘modified nouns’, and nouns that relate to ‘topography’, and ‘traditional practices’ mostly disfavour OVERT, which is a similar case to that of Bahraini Arabic (Holes, 2016: 213-15)

The results for the variability in (dʒ) are explained in relation to the status of the glide as a prestigious variant in the study area due to its association with the dialect of the majority ‘lahgʷat swēgʷ’ and the tribal leaders in the *Mishāb* of *Yāl Saʿad*; the prestige of [j] as variant of the regional Gulf variety that is gaining power in the region stemming from the power of the social groups that use it seems to be feeding into the local prestige of [j]. The behaviour of men and women in the different age groups could be understood in the light of the nature of contact these groups have been exposed to through education, geographic, and social mobility. On the other hand, the results of the DEF variable can be interpreted in the light of the nature of contact as well, but also in the light of the demographics of the different localities. Al-Tharmad, for instance, represents what could be described a less heterogenous context, and older women especially those from al-Tharmad represent the least mobile speakers; linguistic conservatism would be expected to be characteristic of such groups and we have seen this with regard to the speech community and the variables investigated. On the other hand, men, and the middle age group represent the most mobile groups in the speech community due to the nature of their education, and/or work. In addition, we have a locality

like al-Batha with its heterogenous demographic composition. Such types of social ~ demographic characteristics make the latter three groups the perfect candidates for adopting or accommodating to the incoming forms.

Some challenges and setbacks

The main setbacks faced in this research are mostly methodological in nature, e.g.,

- The time of the main fieldwork was limited; also, the access to the community was relatively restricted as explained in Chapter Three. This affected the kind of informants that I could have access to; for example, most of the informants in the middle age group are educated, and work mainly in the educational sector, apart from one woman who did not study beyond the primary school and does not have a job. This can be understood in light of the socio-economic development in Oman and in the study area in particular; most younger speakers have been formally schooled and are doing or have done an undergraduate degree and/or currently in the workforce. It would have been more informative as to the status of the two variables to investigate in a wider range the behaviour of low-contact speakers in the middle and young age groups.
- The other setback is also methodological in nature; there are few one-speaker cells mainly in the younger age group in comparison to the others. So, the results from the multivariate analysis may have been skewed to certain social characteristics of the individual speakers rather than the group itself.

Implications for further research

Time never seems to be on the side of the researcher. The scope of the study is limited by the time constraint; however, some main issues emerged for further exploration.

- It would be intriguing to see how the observed [j]→[dʒ] and [gʲ]→[dʒ] variation (§ 5.5) which looks like a case of a sociolinguistic merger will progress in this speech community considering the fast-changing nature of the study area and Oman as well (Al-Wer, personal communication).
- The study is initially focused on uncovering sociolinguistic variation in an area composed of two main localities. On one side, this scope is enlarged when a third locality, al-Khadhra, is included due to the difficulties faced in recruiting informants. On the other, the research has revealed that the null definite article feature is mostly characteristic of a certain cluster, namely al-Tharmad. It also has revealed that certain forms are characteristic of ‘older speech’, and that the younger speakers show a different treatment of this variable. This finding calls for an inter-generational variationist investigation in order to give an informed opinion on the status of this feature as a sociolinguistic variable with greater certainty, and to make valid judgments with regard to the type of change taking place in this area, that is if ‘linguistic change’ is the case indeed.

In addition, another aspect of the sociolinguistic scene in al-Suwaiq highlighted in § 1.2 is the rapid and abrupt change in the demographics and socioeconomics of the town which is already heterogenous and transitional in nature. An investigation of the effect of these factors on the general dialectal scene in this area would be the way to go. A dialectal map for the study area would be a first step that would also offer a valuable insight to the larger B/H dialectological scenario in Oman.

References

- Abdel-Jawad, H. (1986). The emergence of an urban dialect in the Jordanian urban centres. *International Journal of the Sociology of Language*, 61(1), 53–63.
- Al-Abri, K. (2002). *The Phonological Phenomena in Al-Hamra Dialect: A Descriptive Analysis*. MA thesis. Sultan Qaboos University, Oman.
- Alaodini, H. (2019). *A Sociolinguistic study of the Dawāsir Dialect in Dammam, Eastern Arabia: fortition of /j/ and unrounding of /a/*. PhD thesis. University of Essex, UK.
- Al-Aghbari, K. (2001). *The Broken Plurals in the Muscat Dialect of Omani Arabic*. MA thesis. University of Victoria, Canada.
- Al-Ṭā'ī, A. (2008). *Tārīx 'Umān s-Siyāsiy*. Kuwait: Ar-Rubē'an Publishing. Accessed [Dember, 2016] at: <https://archive.org/details/Omanpoliticalhistory/page/n245/mode/2up?view=theater>
- Al-Amadidhi, D. (1985). *Lexical and sociolinguistic variation in Qatari Arabic*. PhD thesis. University of Edinburgh, UK.
- Al-Āni, A. (1999). *'Umān fī l-'Uṣūri l-'Islāmiyya l-'Ulā*. London: Dāru l-Ḥikma. Accessed [January, 2017], at: <http://www.omanw.com/books.asp?Page=5&cat2=17&cat=>
- Al-'Azīzi, A. (2014). *Juġrāfiyyat 'Umān fī Kutubi r-Raḥḥāla l-'Arab*. Muscat: Ministry of Heritage and Culture.
- Al-Balushi, R. (2016). Omani Arabic: more than a dialect. *Macrolinguistics*, 4(4), 80-125.
- Al-Essa, A. (2019). Phonological and Morphological Variants. In E. Al-Wer & U. Horesh (Eds.), *The Routledge Handbook of Arabic Sociolinguistics* (pp. 151-168). Oxon-New York: Routledge.

- Al-Ġalāyīnī, M. (1912). *Ġāmi' d-Durūs l-'Arabiyya* (Collection of Arabic lessons, First Edition, 2010). Beirut: Risāla Publishers. Accessed at: https://books-library.online/files/books-library.online_noo59d2dbbc5847be47b63bb4-6396.pdf
- El-Gindi, A. A. (1983). *Al-Lahajātu l-'Arabiyya fi t-Turāt* (Heritage of the Arabic Dialects, Vol. II). Ad-Dār l-'Arabiyya lil-kitāb.
- Al-Hawamdeh, A. (2016). *A sociolinguistic investigation of two Hōrāni features in Sūf, Jordan*. PhD thesis. University of Essex, UK.
- Al-Muhannadi, M. (1991). *A sociolinguistic study of women's speech in Qatar*. PhD thesis. University of Essex, UK.
- Al-Nassir, A. A. (1993). *Sibawayh the phonologist: a critical study of the phonetic and phonologic theory of Sibawayh as presented in his treatise "Al Kitab"*. London: Kegan Paul International.
- Al-Sa'di, F. (2015). *At-Ta'rīx s-Siyāsīy wa l-'Ilmiy li Wilāyatay s-Suwayq wa l-Miṣin'a* (Political and scholastic history of the towns of al-Suwaiq and al-Miṣin'a, Vol. I). Muscat: Dākirat 'Umān.
- Al-Salhi, M. (2015). *The phonetic and morphological structure in Alhoqain in light of the Arabic language*. MA thesis. Sultan Qaboos University, Oman.
- Al-Sālmī, A. (1928). *Tuḥfat l-'A'yān fi Ansābi Ahli 'Umān*. Accessed [January, 2017] at: https://archive.org/details/t7fET_al23yan_bsyrET_ahl_oman01
- Alshawi, W. (2019). *A sociolinguistic study of a southern Iraqi dialect: fortition of the variants [j] and [ʃ]*. PhD thesis. University of Essex, UK.

- Al-Shehri, A. (1993). *Urbanisation and Linguistic Variation and Change: a sociolinguistic study of the impact of urbanisation on the linguistic behaviour of urbanised rural immigrants in Hijaz, Saudi Arabia*. PhD thesis, University of Essex, UK.
- Al-Shibli, H. (2010). *As-Suwayq ft Dākirat t-Tārīx*. Al-Suwaiq: Maktabat Ash-Shikēli, Maktabt Manaru l-‘ilm.
- Al-Siyābi, S. (1964). *Is ‘āf l- ‘A ‘yān ft ‘Ansāb ‘Ahli ‘Umān*. Accessed [October, 2016] at: <http://www.omanw.com/books.asp?Page=5&cat2=17&cat=>.
- Alqahtani, K. (2015). *A sociolinguistic study of the Tihami Qahtani dialect in Asir, Southern Arabia*. PhD thesis, University of Essex, UK.
- ‘Āšūr, S. (Ed.). (1980). *Tārīḥ Ahli ‘Umān* (The History of the People of Oman). Muscat: Ministry of Culture and Heritage. Accessed [November, 2016], at: <http://www.omanw.com/books.asp?Page=5&cat2=17&cat=>
- Al-Qouz, M. (2009). *Dialect contact, acquisition and change among Manama youth, Bahrain*. PhD thesis. University of Essex, UK.
- Al-Wer, E. (1991). *Phonological Variation in the Speech of Women from Three Urban Areas in Jordan*. PhD thesis. University of Essex, UK.
- Al-Wer, E. (1999). Why Do Different Variables Behave Differently? Data from Arabic. In Y. Suleiman (Ed.), *Language and Society in the Middle East and North Africa: Studies in Variation and Identity* (pp. 38-57). London: Routledge.
- Al-Wer, E. (2013). Sociolinguistics. In J. Owens (Ed.), *The Oxford Handbook of Arabic Linguistics*. Oxford: Oxford University Press (pp. 241-263). Doi: 10.1093/oxfordhb/9780199764136.001.0001

- Ambu Saidi, S. (2019). *A Sociolinguistic Study of Dialect Contact and Change in Omani Arabic*. PhD thesis. Newcastle University, UK.
- Aoun, J., Benmamoun, A. & Choueiri, L. (2010). *The Syntax of Arabic*. Cambridge: Cambridge University Press.
- Badger, J. P. (Ed.). (2010). *History of the Imams and Seyyids of Oman* [Ibn Ruzaiq, 1873]. London: Printed for Hakluyt Society.
- Bani-Yasin, R. & Owens, J. (1987). The phonology of a northern Jordanian Arabic dialect. *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 137(2), 297-331.
- Bauer, L. (2008). Lenition Revisited. *Journal of Linguistics*, 44(3), 05-624. Accessed at: <https://www.jstor.org/stable/40058066>
- Beeston, A. F. L. (1970). *The Arabic Language Today*. London: Hutchinson University Library.
- Berezowski, L. (2009). *The Myths of Zero Article*. London, New York: Continuum Logo.
- Berezowski, L. (1997). Iconic Motivation for the Definite Article in English Geographical Proper Names. *Studia Anglica Posnaniensia*, 32, 127-144.
- Bettega, S. (2016). *Syntactic Issues in Modern Omani Arabic*. PhD thesis. Università degli Studi di Torino, Italy.
- Bhat, D. N. S. (1978). A general study of palatalization. In Greenberg, J. H. (Ed.), *Greenberg Universals of human language* (pp. 47-92). Stanford: Stanford University Press.
- Blanc, H. (1969). The Fronting of Semitic g and the qāl-gāl Dialect Split in Arabic. *Proceedings of the International Conference on Semitic Studies*, Jerusalem 19-23 July 1965, 7-37. Jerusalem: The Israel Academy of Sciences and Humanities.

- Brockett, A. (1985). *The Spoken Arabic of Khābūra on the Bātina of Oman*. Journal of Semitic Studies, Monograph No. 7. Manchester University Press.
- Brustad, K. (2000). *The Syntax of Spoken Arabic: A Comparative Study of Moroccan, Egyptian, Syrian, and Kuwaiti Dialects*. Washington, D.C: Georgetown University Press.
- Bybee, J. & Easterday, S. (2019). Consonant strengthening: A crosslinguistic survey and articulatory proposal. *Linguistic Typology*, 23(2), 263-302. Doi:[10.1515/lingty-2019-0015](https://doi.org/10.1515/lingty-2019-0015)
- Cantineau, J. (1966). *Cours de phonétique Arabe*. Translated to Arabic by Sāliḥ Al-Qarmādi. Tunis: University of Tunis.
- Caspari, C. P., Wright, W., Smith, W. R. & Goeje, M. J. (1967). *A grammar of the Arabic language*. London: Cambridge University Press.
- Census Administration. (2010). *General Census of Population, Housing, & Establishments 2010*. Accessed at: https://www.ncsi.gov.om/Elibrary/LibraryContentDoc/bar_Census%20Final%20Result%202010_388bd9c6-a938-467d-8c92-f6950cc1785f.pdf, last accessed [09-12-2016].
- Christopher, L. (1999). *Definiteness*. Cambridge: Cambridge University Press.
- Choueiri, L. (2006). Determiners. In K. Versteegh *et al* (Eds.), *Encyclopedia of Arabic Language and Linguistics* (pp. 579-583). Leiden, Boston: Brill.
- Crowley, T. (1997). *An introduction to historical linguistics*. New Zealand: Oxford University Press.
- Danon, G. (2002). *Case and Formal Definiteness: The Licensing of Definite and Indefinite Noun Phrases in Hebrew*. PhD thesis. Tel Aviv University, Tel Aviv.

- Davey, R. J. (2016). *Coastal Dhofārī Arabic: A Sketch Grammar*. Published PhD Dissertation. University of Manchester: Manchester. [Published by Brill in 2016].
- De Carvalho, J. B., Scheer, T. & Ségéral, P. (2008). *Lenition and fortition* (Studies in Generative Grammar 99). Berlin: Mouton de Gruyter.
- De Jong, R. (2000). *A Grammar of the Bedouin Dialects of the Northern Sinai Littoral: Bridging the Gap between the Eastern and Western Arab World*. Leiden, Boston: Brill.
- De Jong, R. (2011). *A Grammar of the Bedouin Dialects of Central and Southern Sinai*. Leiden, Boston: Brill.
- De Jong, R. E. (2007). Gahawa-Syndrome. In K. Versteegh, *et al* (Eds.), *Encyclopedia of Arabic Language and Linguistics* (Vol. II, pp. 151-153).
- Doron, E. (2003). Bare Singular Reference to Kinds. *Proceedings from Semantics and Linguistic Theory, 13*, 73–90.
- Eades, D. (2009). Retention of the passive verb in a Bedouin dialect of northern Oman. *Zeitschrift für Arabische Linguistik, 51*, 5-21. Harrassowitz Verlag. Accessed [January, 2016], at: <http://www.jstor.org/stable/43525855>.
- Eades, D. (2009b). The Arabic dialect of a Šawāwī community of northern Oman. In E. Al-Wer and R. De Jong (Eds.), *Arabic dialectology: in honour of Clive Holes on the occasion of his sixtieth birthday* (pp. 77-98). Leiden: Brill.
- Eades, D. (2009b). The Arabic Dialect Of Šawāwī Community of Northern Oman. In Al-Wer, E. & R. De Jong, R. (Eds), *Arabic Dialectology: in Honor of Clive Holes on the Occasion of his Sixtieth Birthday* (Vol. 53, pp.77-98). Brill.
- Eades, D. (2011). A Transitional Arabic Dialect of the Northern Omani Interior. *Zeitschrift für Arabische Linguistik, 54*, 27-41.

- Eckert, P. (1997). Age as a sociolinguistic variable. In F. Coulmas (Ed.), *The Handbook of Sociolinguistics* (pp. 151-167). Oxford: Blackwell publishing.
- Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics*, 12(4) 453–476. Doi: <https://doi.org/10.1111/j.1467-9841.2008.00374.x>
- Eckert, P. (2013). Ethics in linguistic research. In R. J. Podesva and D. Sharma (Eds.): *Research Methods in Linguistics*. (pp. 11-26). New York: Cambridge University Press.
- Eckert, P. & Labov, W. (2017). Phonetics, phonology and social meaning. *Journal of Sociolinguistics*, 21(4), 467-496. Doi: <https://doi.org/10.1111/josl.12244>
- Elabbas, B. & Choueiri, L. (2013). The Syntax of Arabic from a Generative Perspective. In J. Owens (Ed.), *The Oxford Handbook of Arabic Linguistics* (115-164). Oxford University Press. Doi: 10.1093/oxfordhb/9780199764136.001.0001
- E-LAN (Version 5.4) [Computer software]. (2018). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Accessed at: <https://archive.mpi.nl/tla/elan>
- Embarki, M. (2013). Phonetics. In Owens, J. (Ed.), *The Oxford Handbook of Arabic Linguistics* (115-164). Oxford University Press. Doi: [10.1093/oxfordhb/9780199764136.013.0002](https://doi.org/10.1093/oxfordhb/9780199764136.013.0002)
- Ethnologue. (2016). Spoken Omani Arabic. Accessed [December, 2016], at: <https://www.ethnologue.com/language/acx>
- Fischer, W. (1997). Classical Arabic. In R. Hetzron (Ed.), *The Semitic Languages* (pp. 187-219). London-New York: Routledge.
- Gardani, Francesco (2018). On morphological borrowing. *Language and Linguistics Compass* 12(10), 1–17. Doi:10.1111/lnc3.12302

- Gardiner, S. and Nagy, N. (2017). Stable Variation vs. Language Change and the Factors that Constrain Them. *University of Pennsylvania Working Papers in Linguistics*, 23(2).
Accessed at: <https://repository.upenn.edu/pwpl/vol23/iss2/10>
- Glover, B. (1988). *The Morphophonology of Muscat Arabic*. PhD Dissertation. University of California: Los Angeles.
- Hary, B. (1996). The *ḡīm/gīm* in colloquial urban Egyptian Arabic. *Israel Oriental Studies*, 16, 153–68.
- Hāmid, A. M. (2008). Lahjat Nizwā: dirāsa ṣawtiyya (The dialect of Nizwa: a phonetic study). *Majallat d-dirāsāt l-luḡawiyya*, 10(2), 27-116.
- Herin, B. (2010). *Le parler arabe de Salt, Jordanie: phonologie, morphologie et éléments de syntaxe*, pp. 39-41. PhD thesis. Université libre de Bruxelles, Brussels.
- Holes, C. (1980). Phonological variation in Bahraini Arabic: The [j] and [y] allophones of /j/. *Zeitschrift für Arabische Linguistik*, 4, 72-89.
- Holes, C. (1986). The Social Motivation of Phonological Convergence in three Arabic Dialects. *International Journal of the Sociology of Language*, 61(1), 33-51.
- Holes, C. (1987). *Language variation and change in a modernising Arab State: The case of Bahrain*. London and New York: Routledge
- Holes, C. D. (1989). Towards a Dialect Geography of Oman. *Bulletin of the School of Oriental and African Studies*, 52(3), 446-462.
- Holes, C. (1991). Kashkasha and the fronting and voicing of the velar stops revisited: A contribution to the historical phonology of the peninsular Arabic dialects. In Kaye, A. (Ed.), *Semitic Studies: in Honor of Wolf Leslau on the Occasion of his Eighty-fifth Birthday* (Vol. I, pp. 652-678). Wiesbaden: Harrassowitz.

- Holes, C. D. (1998). Retention and loss of the passive verb in the Arabic dialects of northern Oman and eastern Arabia. *Journal of Semitic Studies*, 43, 347-362.
- Holes, C. D. (2001). *Dialect, culture, and society in eastern Arabia*. Boston: Brill.
- Holes, C. D. (2004). *Modern Arabic Structures, Functions, and Varieties*. Washington, DC: Georgetown University Press.
- Holes, C. (2006 a.). Gulf States. In K. Versteegh, *et al* (Eds.) *Encyclopedia of Arabic Language and Linguistics* (Vol. II, pp. 210-216). Brill.
- Holes, C. (2006 b.). The Arabic Dialects of Arabia. *Proceedings of the Seminar for Arabian Studies*, 36, 25-34.
- Holes, C. D. (2007). Omani Arabic. In K. Versteegh, *et al* (Eds.), *Encyclopedia of Arabic Language and Linguistics* (Vol. III, pp. 478-491). Leiden, Boston: Brill.
- Holes, C. (2011). A Participle Infix in the Eastern Arabian Dialects- An Ancient Pre-conquest feature?. *Jerusalem Studies in Arabic and Islam*, 38, 75-97.
- Holes, C. (2012). The Omani Arabic Dialects in Their Regional Context: Yesterday, Today, and Tomorrow. In Zidjaly, N. (Ed.), *Building Bridges: Integrating Language, Linguistics, Literature, and Translation in English Studies* (pp. 233-250). Cambridge: Cambridge Scholars Press.
- Holes, C. (2016). *Dialect, Culture and Society in Eastern Arabia: Phonology, Morphology, Syntax, Style* (Vol. III). Brill: Leiden.
- Holes, C. (2018). *Arabic Historical Dialectology*. Oxford University Press.
- Honeybone, P. (2008). Lenition, weakening and consonantal strength: tracing concepts through the history of phonology. In De Carvalho, J. B., Scheer, T. & Ségéral, P. (Eds.),

- Lenition and fortition* (Studies in Generative Grammar 99) (pp. 9-92). Berlin: Mouton de Gruyter.
- Huehnergard, J. (2005). Features of Central Semitic. In *Gianto Biblical and Oriental Essays in Memory of William L. Moran* (pp. 184-186). Rome: Pontificio Istituto Biblico.
- Hussain, A. (2017). *The Sociolinguistic Correlates of Dialect Contact and Koineisation in Medini Arabic: Lenition and Resyllabification*. PhD thesis. University of Essex, UK.
- Ibrahim, H. M. (1986). Standard and prestige language: A problem in Arabic Sociolinguistics. *Anthropological linguistics*, 28(1), 115-126. Accessed at: <https://www.jstor.org/stable/pdf/30027950.pdf>
- Il-Hazmi, A. M. (1975). *A critical and comparative Study of the spoken dialect of the Ḥarb tribe in Saudi Arabia*. PhD thesis. University of Leeds, UK.
- Ingham, B. (1971). Some characteristics of Meccan speech Ingham. *Bulletin of the school of Oriental and African studies*, 34 (2), 273-297.
- Ingham, B. (1982). *North East Arabian dialects*. London: Kegan Paul International.
- Ingham, B. (1994). *Najdi Arabic: Central Arabian*. Amsterdam-Philadelphia: John Benjamins Publishing Company.
- Jastrow, O. (1997). The Neo-Aramaic Languages. In Hetzron, R. (Ed.), *The Semitic Languages* (pp. 334-377). London-New York: Routledge.
- Jayakar, A. (1889). The O'manee Dialect of Arabic. *Journal of the Royal Asiatic Society*, 649-87.
- Johnstone, T. M. (1965). The Sound Change $j > y$ in the Arabic Dialects of Peninsular Arabia. *Bulletin of the School of Oriental and African Studies*, 28(2), 233-241.
- Johnstone, T.M. (1967). *Eastern Arabian Dialect Studies*. London: Oxford University Press.

- Johnson, D.E. (2009). Getting off the GoldVarb Standard: Introducing Rbrul for Mixed-Effects Variable Rule. Analysis. *Language and Linguistics Compass*, 3, 359-383. Doi: <https://doi.org/10.1111/j.1749-818X.2008.00108.x>
- Johnson, D. E. (2013). Descriptive statistics. In R. J. Podesva and D. Sharma (Eds.), *Research Methods in Linguistics*. (pp. 288-315). New York: Cambridge University Press.
- Kaye. A. S. (1970). Arabic /žīm/: A Synchronic and Diachronic Study. *Linguistics* 79, 31-72.
- Kaye A. S. & Rosenhouse, J. (1997). Arabic Dialects and Maltese. In R. Hetzron (Ed.), *The Semitic Languages* (pp. 263-311). London-New York: Routledge.
- Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia. University of Pennsylvania Press.
- Labov, W. (1984). Field methods of the Project on Linguistic Change and Variation. In J. Bragh and J. Sherzer (Eds.), *Language in use* (28-53). Eaglewood Cliffs: Prentice Hall.
- Labov, W. (2001). *Principles of linguistic change: Social Factors* (Vol. II). Malden-Massachusetts: Blackwell.
- Lass, R. (1984). *Phonology: An introduction to basic concepts* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.
- Levin. A. Imāla. In K. Versteegh, *et al* (Eds.), *Encyclopedia of Arabic Language and Linguistics* (Vol. II, pp. 311-315). Leiden: Brill.
- Lorimer, J. G. (2003). *Gazeteer of the Arabian Gulf, Oman, and Central Arabia, Geographical* (Part 2, Vol. 6). Translated under Joint supervision of Sultan Qaboos University and Oxford University.
- Lorimer, J. G. (2013). *Gazeteer of the Arabian Gulf, Oman, and Central Arabia, Geographical* (Vol. 14). Beirut: Ad-Dār Al-‘Arabiyya lil-Mawsū‘āt.

- Manfredi, S. (2017). Demonstratives and the emergence of a definite article in JubaArabic and Ki-Nubi. *Journal of Pidgin and Creole Languages*, 32(2), 205-232.
- Marogy, A. E. (2012). *The Foundations of Arabic Linguistics*. Leiden: Brill. Doi: <https://doi.org/10.1163/9789004229655>
- Miles, S. B. (1966). *The Countries and Tribes of the Persian Gulf*. London: Harrison.
- Milroy, L. & Gordon, M. (2003). *Sociolinguistics: Method and Interpretation*. Blackwell Publishing.
- Ministry of Agriculture and Fisheries. (2015). 'Adā' l-Qiṭā' z-Zirā' i w s-Samaki. Accessed [January, 2016], at: <http://www.maf.gov.om/Pages/FactDetails.aspx?lang=AR&DIId=0&CIId=0&CMSId=269&id=21>
- Ministry of Information. (1995 a). *Years of Endeavours and Achievement: Al Batinah*.
- Ministry of Information. (1995 b). *Masīrat l-Xayr*. Muscat: Mazūn Publishing.
- Ministry of Information. (1998). *As-Suwēq l-Bahiyya Durrat l-Bātina l-'Abbiya*.
- Ministry of Information. (2006). *Oman 2006-2007*. Muscat: Oman Establishment for Press, Publication, & Advertising.
- Ministry of Information. (2010). *Oman 2010-2011*. Muscat: Oman Establishment for Press, Publication, & Advertising.
- Ministry of Social Development. *Omani Women Association*. Accessed [December, 2016], at: <https://www.mosd.gov.om/index.php/en/associations-and-clubs-2/women-associations/218-associations-and-clubs/omani-women-associations/782-omani-women-associations-in-the-governorate-of-al-batinah>

- Morano, R. (2019). *The Arabic dialect spoken in the al- 'Awābī district, northern Oman*. PhD thesis. University of Leeds, UK.
- National Center for Information and Statistics. (2012). *Figures and Stats: 2011*. Accessed at: https://www.ncsi.gov.om/Elibrary/LibraryContentDoc/ben_Facts%20and%20Figures%202011_6b8b6d1a-aea5-47b1-b158-1fe8b16f89b6.pdf, last accessed [25-04-2021].
- National Center for Information and Statistics. (2016). *Aš-Šabāb l- 'Umāni fī 'arqam* (Statistics of Oman youth). Accessed at: https://www.ncsi.gov.om/Elibrary/LibraryContentDoc/bar_Youth%20in%20Figures_5c270ff7-be44-4a41-9921-0cf415db2c10.pdf , last accessed [25-04-2021].
- National Center for Information and Statistics. (2016). *General Education Statistics: 2015-2016*, (6). Accessed [December, 2016], at: https://www.ncsi.gov.om/Elibrary/LibraryContentDoc/bar_School%20Education%202015%20-%202016%20_aea9e37d-d67e-4d15-9f4f-c0d029b1a4d1.pdf
- National Center for Information and Statistics. (2016). *Deaths and Births Statistics Bulletin*, (5). Accessed [December, 2016], at: https://www.ncsi.gov.om/Elibrary/LibraryContentDoc/bar_Deaths%20and%20Births%202015%20%205%20Issue_6dbbac6b-1ec4-4e25-a8ab-e10bd8ecbaaf.pdf.
- Ochs, P. (2000). *Maverick Guide to Oman*. Gretna: Pelican Publishing.
- Omar, M. (2008). *Mu 'ġam l-Luġa l- 'Arabiyya l-Mu 'āšira* (The Contemporary Arabic Dictionary, Vol. I-IV). Cairo: 'Ālam Al-kutub. Accessed at: <https://www.alarabimag.com/download/15187-pdf>
- Owens, J. (2001). Creole Arabic: The Orphan of All Orphans. *Anthropological Linguistics*, 43(3), 348-378. Accessed [December, 2020], at: <http://www.jstor.org/stable/30028796>

- Owens, J. (2006). *A Linguistic History of Arabic*. Oxford: Oxford University Press.
- Parrado, J. G. (2019). Reflexes of Old Arabic */ǧ/ in the Maghrebi Dialects. *Arabica*, 66, 137-169. Doi: <https://doi.org/10.1163/15700585-12341521>
- Pat-El, N. (2009). The development of the Semitic definite article: a syntactic approach. *Journal of Semitic Studies*, 54(1), 19-50.
- Peterson, J. E. (2004). Oman's Diverse Society Southern Oman. *Middle East Journal*, 58(2), 254-269.
- Phillips, W. (1986). *Rihla ilā 'Umān*. Translated by Dr Mohammed Amin. Oman: Ministry of Heritage and Culture.
- Potter, R. S. (2018). *A variationist multilocality study of unstressed vowels and verbal -s marking in the peripheral dialect of east Suffolk*. PhD thesis. University of Essex, UK.
- Procházka, T. (1981). The Shi'i Dialects of Bahrain and Their Relationship to the Eastern Arabian Dialects and the Omani Dialect of al-Ristāq. *Zeitschrift für arabische Linguistik*, 6, 16-55.
- Procházka, S. (2018). The Northern Fertile Crescent. In Holes, C. (Ed.), *Arabic Historical Dialectology* (pp. 257-292). Oxford University Press.
- Krug, M. & Lucas, C. (2018). Definite article (omission) in British, Maltese, and other Englishes. *STUF - Language Typology and Universals*, 71, 261 - 303.
- Rabin, C. (1951). *Ancient West Arabian*. London: Taylor's Foreign Press.
- Ratcliffe, R. R. (2005). Bukhara Arabic: A Metatypized Dialect of Arabic in Central Asia. In É. Á. Csató, B. Isaksson and C. Jahani (Eds.), *Linguistic Convergence and Areal Diffusion: Case Studies from Iranian, Semitic and Turkic* (pp. 141–159). London-New York: RoutledgeCurzon.

- Retsö, J. (1989). The apophonic passive in Arabic. In *Diathesis in the Semitic Languages*. Leiden, the Netherlands: Brill. Doi: https://doi.org/10.1163/9789004348264_005
- Retsö, J. (2013). What is Arabic. In J. Owens (Ed.), *The Oxford Handbook of Arabic Linguistics* (433-494). Oxford: Oxford University Press.
Doi:10.1093/oxfordhb/9780199764136.001.0001
- Rubin, A. (2005). *Studies in Semitic Grammaticalization*. Winona Lake, Indiana: Eisenbrauns.
- Rubin, A. (2008). The subgrouping of Semitic languages. *Language and Linguistics Compass*, 2(1), 79-102. Doi.: <https://doi.org/10.1111/j.1749-818X.2007.00044.x>
- Ryding, K. (2005). *A Reference Grammar of Modern Standard Arabic (Reference Grammars)*. Cambridge: Cambridge University Press.
- Ségéral, P. & Scheer, T. (2008a). positional factors in Lenition and Fortition. In De Carvalho, J. B., Scheer, T. & Ségéral, P. (Eds.), *Lenition and fortition (Studies in Generative Grammar 99)* (pp 131-172). Berlin: Mouton de Gruyter.
- Ségéral, P. & Scheer, T. (2008b). The Coda Mirror, stress, and positional parameters. In De Carvalho, J. B., Scheer, T. & Ségéral, P. (Eds.), *Lenition and fortition (Studies in Generative Grammar 99)* (pp. 484-518.). Berlin: Mouton de Gruyter.
- Shaaban, K.A. (1977). *The Phonology and Morphology of Omani Arabic*. PhD Dissertation. University of Texas: Austin.
- Schilling, N. (2013). Surveys and interviews. In R. J. Podesva & D. Sharma (Eds.), *Research Methods in Linguistics* (pp. 96-115). New York: Cambridge University Press.
- Sībawayh, A. (1982). *Kitāb Sībawayh (The Book of Sībawayh, Vol. IV)*. Hārūn, A. M. (Ed.). Cairo: Maktabat al-Xānji and Riyadh: Dār r-Rifā'ī.

- Spencer, A. (1996). *Phonology: Theory and Description*. Blackwell Publishing.
- Szigetvári, P. (2008). What and where?. In De Carvalho, J. B., Scheer, T. & Ségéral, P. (Eds.), *Lenition and fortition* (Studies in Generative Grammar 99) (pp. 131-172). Berlin: Mouton de Gruyter.
- Tagliamonte, S. A. (2006). *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press.
- Tagliamonte, S. & Roeder, R. V. (2009). Variation in the English definite article: Socio-historical linguistics in the speech community. *Journal of Sociolinguistics*, 13(4), 435–471.
- Tagliamonte, S. A. (2012). *Variationist Sociolinguistics: Change, observation, interpretation*. Sussex: Willey-Blackwell.
- Taqi, H. (2010). *Two ethnicities, three generations: phonological variation and change in Kuwait*. PhD thesis. University of Newcastle, UK.
- Tosco, M. & Manfredi, S. (2013). Pidgins and Creoles. In J. Owens (Ed.), *The Oxford Handbook of Arabic Linguistics* (495-519). Oxford: Oxford University Press. Doi: 10.1093/oxfordhb/9780199764136.001.0001
- Turner, M. L. (2013). *Definiteness Marking in Moroccan Arabic: Contact, Divergence, and Semantic Change*. MA thesis, University of Texas, Austin.
- Turner, M. L. (2018). *Definiteness in the Arabic Dialects*. PhD thesis, University of Texas, Austin.
- Valeri, M. (2009). *Politics and Society in the Qaboos State*. London: Hurst Publishers.
- Vanhove, M. (2009). Yemen. In K. Versteegh, et al (Eds.), *Encyclopedia of Arabic Language and Linguistics* (Vol. IV, pp. 750-758). Leiden: Brill.

- Watson, J. C. (2002). *The Phonology and Morphology of Arabic*. Oxford: Oxford University Press.
- Watson, J. (1993). *The Syntax of Ṣan'āni Arabic*. Wiesbaden: Harrassowitz.
- Webster, R. (1991). Notes on the Dialect and Way of Life of the Āl Wahība Bedouin of Oman. *Bulletin of the School of Oriental and African Studies*, 54 (3), 473-485. Accessed at: <http://www.jstor.org/stable/619056>
- Wilkinson, J. C. (1987). *Imamate Tradition of Oman*. Cambridge: Cambridge University Press.
- Woidich, M. & Zack, L. (2009). The *g/ğ* -question in Egyptian Arabic revisited. In E. Al-Wer & R.E. De Jong (Eds.), *Arabic Dialectology: In Honour of Clive Holes on the Occasion of His Sixtieth Birthday* (pp. 41-62). Leiden: Brill.
- Zaborski, A. (2006). Jīm. In K. Versteegh, et al (Eds), *Encyclopedia of Arabic Language and Linguistics*, II (pp. 494-496). Leiden: Brill.
- Zeroual, C. (2006). Affrication. In K. Versteegh, et al (Eds.), *Encyclopedia of Arabic Language and Linguistics* (Vol. I, pp. 26-28). Leiden: Brill.
- Z'ētar, W. (2014). *As-Suwayq: Al-Mawsū'a al-Maydāniyya Sāhīlan wa Sahlan wa Jabalan*. Amman: Juhainah Publishing.

Appendices

Appendix A: Distribution of (dʒ) per the preceding and the following segments

Preceding Segment	[j]	[gʲ]	Totals
pause	27	111	138
ʃ	10	2	12
tʃ	2	0	2
h	11	5	16
sʃ	2	0	2
ð	0	1	1
gʲ	0	3	3
kʲ	4	0	4
ʈ	2	0	2
ʃ	6	4	10
θ	1	1	2
a	204	107	311
a:	216	114	330
b	13	0	13
d	26	8	34
e	14	2	16
e:	4	2	6
f	2	5	7
g	2	0	2
h	6	0	6
i	168	95	263
i:	5	3	8
k	6	0	6
l	132	149	281
m	34	10	44
n	142	38	180
o	1	0	1
o:	2	0	2
q	1	0	1
r	37	3	40
s	11	29	40
t	173	55	228
u	3	3	6
u:	6	0	6
w	84	24	108
j	377	20	397
Totals	1801	709	2510

Following segment	[j]	[gʲ]	Totals
pause	3	8	11
ʧ	1	2	3
ħ	10	1	11
ð	2	1	3
gʲ	1	11	12
a	202	237	439
a:	406	86	492
b	11	2	13
d	34	6	40
e	11	28	39
e:	10	6	16
f	2	0	2
h	4	2	6
i	499	145	644
i:	329	50	379
k	2	0	2
l	56	30	86
m	3	21	24
n	6	7	13
o:	11	2	13
r	7	25	32
s	1	0	1
t	0	4	4
u	9	10	19
u:	108	23	131
w	9	1	10
x	1	0	1
j	58	2	60
z	4	0	4
Totals	1801	709	2510

Appendix B: Results for the use of [gʲ] per stress, manner of articulation, syllabic position, gemination and linguistic context

Stress

Syllable with (dʒ)	% [gʲ]	N for both variants
Stressed	26.4 (N=397)	1,506
Unstressed	30.6 (N=304)	993
		Total N=2,499

Linguistic context

Linguistic context	% [gʲ]	N for both variants
V-C	37.2 (N=108)	290
V-V	32.4 (N=210)	648
C-C	26.1 (N=6)	23
C-V	25 (N=350)	1,400
Pause_other	19.6 (N=27)	138
		Total N=2,499

Manner of articulation

Preceding manner of articulation	% [gʲ]	N for both variants
Sonorant	29.4 (N=594)	2,019
Obstruent	23.4 (N=80)	342
pause	19.6 (N=27)	138
		Total N=2,499

Position in syllable

Position in syllable	% [gʲ]	N for both variants
Coda	37.7 (N=112)	297
Onset	28.4 (N=571)	2,013
Branching onset	9.5 (N=18)	189

Gemination

Gemination	% [gʲ]	N for both variants
Singleton	28.6 (N=687)	2,405
Geminate	14.9 (N=14)	94
		Total N=2,499

Appendix C: Nominals with NULL in the whole dataset all used substantively; (cont.)**overleaf)**

argʌnti:n	Argentine	θa:min	eighth
atʃfa:l	children	θo:b	dress
alʃa:b	toys	θrajji	Pleiades
umu:r	things	gʃu:d	young camel
awwal	first; before	gʃa:ʃiʃ	dried sardines
bʃi:r	camel	gʃamb	side
batʃha	Al-Batha (village)	garʃha	Al-Qarha (village)
baħar	sea	ge:ðʃ	summer (date season)
ba:tʃna	Al-Batna (region)	gʃe:r	gear
ba:sʃa:t	buses	gʃe:s	an object used in a family game called 'kēram'
ba:b	door		coffee
ba:kʃir	tomorrow	ghawa	moon
ba:rda	Al-Barda (village)	gimar	penny (coin)
bada:wa	beduinness	giriʃ	Al-Quraihat (village)
baðra	offspring	gre:ħa:t	pancakes
badu	Bedouin	gru:sʃ	Friday
bakrah	young female camel	gʃimʃa	Al-Jurf (town; UAE)
bala:di:n	towns	gʃurf	eleventh
bana:gri	bangles	ħa:diʃaʃar	period (school)
bard	cold	ħisʃsʃa	Praise to Allah
(i)bduwija	Bedouin woman	ħimdilla:h	elderly women
badwija:t	Bedouin women	ħiri:m li-kʃba:r	wild (dessert) life
be:sa	Baisa (coin)	ħija: l-barrije	donkey
be:t	house	ħma:rah	goats
bgar	cows	ħo:ʃ	young camel
bgurah	cow	ħwa:r	line
bha:ra:t	spices	xatʃtʃ	fifth
bida:jeh	wells	xa:mis	Thursday
bidi	well	ximi:s	coins (50 baisa)
biga:ja	leftovers	ximasi:n	pocket in the ground
bi:ban	doors	xammu:d	Fifth.PL [Grade]
bina:t	girls	xawa:mis	taylor
birij	tower	xajja:tʃ	tailoring
bla:d	town	xija:tʃa	mat made of palm leaves
blu:ʃ	Balushis	dʃanah	yard
bo:ʃ	camels	daris	driver
bre:mi	Al-Buraimi (town)	dire:wil	chicken
bju:t	houses	dja:j	dates
ibtida:ʔi	primary.RELATION	rtʃab	fourth
ibtida:ʔijje	primary.ABSTRACTION	ra:biʃ	shade
tayru:d	<i>tagrūd</i> (camel art)	ra:jih	neck
ta:siʃ	ninth	rguba	camels
tamhi:di	kindergarden; reception	rkʃa:b	Rial
θanawi	secondary.RELATION	rja:l	time
θanawijje	secodary.ABSTRACTION	zima:n	investments
θawa:min	Eighth.PL [Grade]	istiθma:ra:t	hour (o'clock)
θa:liθ	third	sa:ʃa	seventh
		sa:biʃ	

Nominals with NULL in the whole dataset; all used substantively (Continued):

sa:biq	past	kru:k	krūk (trad. game)
sa:dis	sixth	kuskus	kuskus
sabit	Saturday	kwe:t	Kuwait
sawa:biʃ	Seventh.PL [Grade]	luʃa	language
sawa:dis	Sixth.PL [Grade]	mʃaʔʔab	mʻaddab (trad. oil)
se:h	sandy inland	myammu:ðʃa	hide and seek
siba:qa:t	races	myarib	dusk
subu:ʃ	week	msʻe:nʃa	Al-Miseen ʻa (village)
su:g	market	matʻhini	mill
swe:gi	Al-Suwaiq (town)	maqʻa:lis	guestrooms
ʃa:riʃ	road	maqʻlas	guestroom
ʃams	sun	maha:t	female relatives
ʃarqijje	Al-Sharqiya (region)	malʃab	pitch (football)
ʃja:di	Al-Sheyadi (tribe)	mana:tʻʃiq	regions
ʃti	winter	manfaʃ	Al-Manfash (village)
sʻba:h	morning	miʃbe:la	Al-Mi ʻbeela (town)
sʻaff	class/grade (school)	mitʻar	rain
sʻe:f	summer	misʻinʃa	Al-Misin ʻa (town)
sʻibiħ	morning	midi	range
ðʻhi:ra	noon	mikʻa:n	place
ðʻiħa	morning	mirka:ðʻ	track (camel racing)
ðihir	noon	mirka:z	related to traditional game of ʻwallamʻ
ðʻiħim	injustice	misḥa:b	tribal territory/confederation
ʃasʻir	afternoon	mizrad	mizrad (trad. jewellery)
ʃa:mri	Al- ʻAmri (tribe)	mka:ʃi	stomach/ intestines
ʃa:ʃir	Tenth.PL [Grade]	mna:tʻaħ	bull fighting
ʃambar	ʻambar (trad. Game)	msa:ʃda	Al-Msa ʻda (village)
ʃuru:sa:t	weddings	mse:blu	type of local plant used for animal food
ʃaʃir	ten		
ʃe:n	eye		
ʃisʻi:di	ʻiṣīda (trad. food)	musʻtʻlaħa:t	terms
ʃi:d	Eid	musalsala:t	TV series
ʃirðʻa	ʻirḍa (camel show)	muwa:sʻala:t	transportation means
ʃirs	wedding	imtiħa:na:t	tests
ʃma:nijja:t	Omani	na:s	people
iʃda:di	preparatory. RELATIVE	nxala	palme tree
ya:ʃri	Al-Ghafri (tribe)	ingʻili:zi	English
faʃir	dawn	hambal	hambal (camel art)
fu:dih	hearts	handasa	engineering
qaratʻ	Al-Qarat (village)	wallam	wallam (trad. game)
kʻabata:t	wardrobes	wihi:ba	Wahiba (tribe)
kʻalme	word	wla:d	children
ki:mja:ʃijje	chemistry.RELATIVE	jadd	hand
kʻina:z	date processing (storing)	jarmu:k	Al-Yarmouk (school)
ko:fi	chocolate	jizi	Al-Jizi (valley/town)
		jo:m	day

Appendix D: Cross-tabulation between type of noun and phrase type (NULL tokens)

Type of noun	Type of phrase					
	Construct phrase	Noun+adjective	Noun phrase	Demonstrative phrase	Nominalised adjective	Total
Adverbial	1	4	36	0	1	42
Camel	7	0	41	5	0	53
Domestic	4	3	41	4	0	52
Food item	1	2	6	3	0	12
Human	1	7	11	3	1	23
Nationality/ethnicity	0	0	1	0	1	2
Other abstract	2	5	9	2	0	18
Other animal	1	1	9	0	0	11
Other inanimate	2	18	15	2	0	37
Personal name	0	0	1	0	3	4
Scholastic term	4	59	4	0	20	87
Season	3	0	4	1	0	8
Special category	0	0	6	0	0	6
Topographical feature	2	0	13	0	0	15
Toponym	16	2	57	2	0	77
Traditional practice	15	39	12	0	26	92
Weekday	0	0	4	0	0	4
Total	59	140	270	22	52	543