OXFORD LINGUISTICS

ARABIC HISTORICAL DIALECTOLOGY

Linguistic and Sociolinguistic Approaches

Edited by

CLIVE HOLES

OXFORD STUDIES IN DIACHRONIC & HISTORICAL LINGUISTICS

Arabic Historical Dialectology

OXFORD STUDIES IN DIACHRONIC AND HISTORICAL LINGUISTICS

RECENTLY PUBLISHED IN THE SERIES

23
Micro-change and Macro-change in Diachronic Syntax
Edited by Eric Mathieu and Robert Truswell

24
The Development of Latin Clause Structure
A Study of the Extended Verb Phrase
Lieven Danckaert

25
Transitive Nouns and Adjectives
Evidence from Early Indo-Aryan
John. J. Lowe

26
Quantitative Historical Linguistics
A Corpus Framework
Gard B. Jenset and Barbara McGillivray

27
Gender from Latin to Romance
History, Geography, Typology
Michele Loporcaro

28
Clause Structure and Word Order in the History of German
Edited by Agnes Jäger, Gisella Ferraresi, and Helmut Weiß

29
Word Order Change
Edited by Ana Maria Martins and Adriana Cardoso

30
Arabic Historical Dialectology
Linguistic and Sociolinguistic Approaches
Edited by Clive Holes

For a complete list of titles published and in preparation for the series, see pp. 423–6

Arabic Historical Dialectology

Linguistic and Sociolinguistic Approaches

Edited by
CLIVE HOLES

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2018

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2017958460

ISBN 978–0–19–870137–8

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

Contents

Series preface viii
List of maps and tables ix
List of abbreviations xi
Transliteration and transcription conventions xiii
The contributors xvii

1. Introduction 1
Clive Holes

1.1 The early history of Arabic: Myth and reality 2
1.2 Early written evidence: 'Middle Arabic'/'Mixed Arabic' 9
1.3 Early dialectal Arabic 14
1.4 Substrates and borrowing 18
1.5 Dialect geography and typology 20
1.6 Dialect change, large-scale and small-scale 22
1.7 Types of internal linguistic change 23
1.8 Indexicalization 24
1.9 Pidginization and creolization 26
1.10 Concluding remarks 27

2. The Maghrebi dialects of Arabic 29
Jordi Aguadé

2.1 Introduction 29
2.2 Sources 32
2.3 General features and classification 33
2.4 Substrata 34
2.5 The Arabization of the Maghreb 36
2.6 Phonology 43
2.7 Nominal morphology 49
2.8 Verb morphology 55
2.9 Word order: VSO or SVO 59
2.10 Loanwords 59

3. The formation of the Egyptian Arabic dialect area 64
Peter Behnstedt and Manfred Woidich

3.1 Introduction 64
3.2 The eastern Delta 72
3.3 The western Delta 75
3.4 The Fayyūm and the northern Nile Valley (NME1) 81
3.5 Northern Middle Egypt (NME2) 82
3.6 Southern Middle Egypt (SME) 82

3.7 Upper Egypt 83
3.8 The oases 89
3.9 The role of substrata 93
3.10 Conclusion 94

4. The adnominal linker *-an* in Andalusi Arabic, with special reference to the poetry of Ibn Quzmān (twelfth century) 96
Ignacio Ferrando

4.1 Introduction 96
4.2 The corpus 98
4.3 The adnominal linker *-an* in IQ 100
4.4 Conclusions 111

5. The Arabic dialects of the Gulf: Aspects of their historical and sociolinguistic development 112
Clive Holes

5.1 Ancient inherited elements 112
5.2 Dialect typology 133
5.3 Summary and conclusions 146

6. Judaeo-Arabic 148
Geoffrey Khan

6.1 Introduction 148
6.2 Periods of written Judaeo-Arabic 149
6.3 Spoken Judaeo-Arabic dialects 161
6.4 Further reading 168

7. The Levant 170
Jérôme Lentin

7.1 The geographical area covered 170
7.2 On the eve of Arabicization 170
7.3 The sources and their interpretation 172
7.4 Levantine and Egyptian dialects 175
7.5 Phonetics and phonology 176
7.6 Pronouns 183
7.7 Morphophonology 185
7.8 Verb morphology 186
7.9 Auxiliary verbs 187
7.10 Preverbs 187
7.11 Existential particle *fī* 192
7.12 Conjunctions 193
7.13 Genitive particles 194
7.14 Adverbs 196
7.15 Lexicon 197
7.16 The Aramaic substrate 199
7.17 Finale 204

8. Dialects (speech communities), the apparent past, and grammaticalization: Towards an understanding of the history of Arabic 206
Jonathan Owens

8.1 Preamble 206
8.2 Terminology and basic background 209
8.3 Five case studies of *b- 212
8.4 Reconstruction 237
8.5 Speech communities and dialectal apparent past 247
8.6 Grammaticalization 251
8.7 Conclusion 253

9. The Northern Fertile Crescent 257
Stephan Procházka

9.1 Introduction 257
9.2 History of settlement 260
9.3 Linguistic description of features relevant to historical dialectology 262
9.4 Conclusions 291

10. Historical and typological approaches to Mauritanian and West Saharan Arabic 293
Catherine Taine-Cheikh

10.1 Origin(s) and the Arab legacy 294
10.2 A very gradual Arabization 304

11. South Arabian and Arabic dialects 316
Janet Watson

11.1 Phonological links between South Arabian and Yemeni dialects 317
11.2 Morphological links between South Arabian and Yemeni dialects 320
11.3 Lexical links between South Arabian and Yemeni dialects 324
11.4 Syntactic similarities between South Arabian and Yemeni dialects 327
11.5 Word stress and prosodic links 330
11.6 Conclusions 332

Glossary 335
Bibliography 355
Index of names 410
Index of subjects 416

Series preface

Modern diachronic linguistics has important contacts with other subdisciplines, notably first-language acquisition, learnability theory, computational linguistics, sociolinguistics, and the traditional philological study of texts. It is now recognized in the wider field that diachronic linguistics can make a novel contribution to linguistic theory, to historical linguistics, and arguably to cognitive science more widely.

This series provides a forum for work in both diachronic and historical linguistics, including work on change in grammar, sound, and meaning within and across languages; synchronic studies of languages in the past; and descriptive histories of one or more languages. It is intended to reflect and encourage the links between these subjects and fields such as those mentioned in the previous paragraph.

The goal of the series is to publish high-quality monographs and collections of papers in diachronic linguistics generally, i.e. studies focussing on change in linguistic structure, and/or change in grammars, which are also intended to make a contribution to linguistic theory, by developing and adopting a current theoretical model, by raising wider questions concerning the nature of language change, or by developing theoretical connections with other areas of linguistics and cognitive science as listed at the beginning of this preface. There is no bias towards a particular language or language family, or towards a particular theoretical framework; work in all theoretical frameworks, and work based on the descriptive tradition of language typology, as well as quantitatively based work using theoretical ideas, also feature in the series.

Adam Ledgeway and Ian Roberts

University of Cambridge

List of maps and tables

Maps

2.1 Morocco 30

2.2 Algeria 31

2.3 Tunisia 31

3.1 The Egyptian Arabic dialect area today 66

3.2 Spring pastures of the tribes in immediate post-conquest Egypt 68

5.1 The 2fsng object/possessive pronoun enclitic: forms and their incidence in eastern and southern Arabia today 125

5.2 The dialectal *-in(n)-* active participle infix: incidence in eastern and southern Arabia today 125

8.1 Arabic-speaking areas of north-eastern Nigeria and Cameroon 231

9.1 Languages spoken in the Northern Fertile Crescent today: Arabic, Turkish, Kurdish, Aramaic 258

9.2 'Bedouin' and 'sedentary' Arabic dialect areas in the Northern Fertile Crescent 259

Tables

2.1 Independent pronouns in Djidjelli, eastern Algeria 51

2.2 Geminate verbs, s-stem, Muslim dialect of Marrakech 56

2.3 Geminate verbs, s-stem, Jewish dialect of Sefrou 56

2.4 French and Spanish loan words in the dialects of Tangiers and Casablanca 62

4.1 Ibn Quzmān's poetry: frequency of head nouns by gender and number 100

4.2 Ibn Quzmān's poetry: frequency of types of nominal adjunct 103

5.1a Mesopotamian lexical influences on Gulf Arabic: agriculture-related 115

5.1b Mesopotamian lexical influences on Gulf Arabic: seafaring- and fishing-related 116

5.1c Mesopotamian lexical influences on Gulf Arabic: general items 117

5.2 Some salient A/B dialectal contrasts in Bahrain and Oman 136

5.3 Some typical communal dialect contrasts in Bahrain 140

5.4 The development of OA /*q*/, /*k*/, /*g*/ in Gulf 'sedentary' dialects 145

8.1 Number of predicates, tokens of *b-* in two Emirati Arabic corpora 215

8.2 *b*-C in Nigerian Arabic 229

8.3	Grammaticalization of *b-, initial account	238
8.4	Reconstructed development of *b-	242
9.1	2/3pl pronoun forms, Shāwi bedouin (Behnstedt 2000: 454)	265
9.2	2fsng pronominal suffix	265
9.3	'Bedouin' (*gilit*) and 'sedentary' (*qəltu*) dialect verb contrasts	266
9.4	The phonological development of /ġ, x, q, g, k/ in the 'bedouin' Arabic dialects of the Northern Fertile Crescent	271
9.5	Personal pronouns in Cilicia (Procházka 2002a: 64)	273
9.6	Personal pronouns in Tillo/Anatolia (Lahdo 2010: 68)	273
9.7	Independent and bound 3rd-person pronouns	274
9.8	2nd-person pronouns	274
9.9	Pronominal suffixes without initial *-h-*	275
9.10	Presentative particles I	278
9.11	Presentative particles II	278
11.1	Sabaic relative pronouns	328

List of abbreviations

a	article
adj	adjective
ap	active participle
aux	auxiliary (verb)
b	broken (plural)
C	consonant
com	common
constr	construct (state)
def	definite
dem	demonstrative
du	dual
f	feminine
fut	future
indef	indefinite
indir	indirect
interrog	interrogative
m	masculine
n	noun
neg	negative
obj	object
p-stem	prefix-stem (of the verb)
part	particle
perf	perfective (particle)
poss	possessive
pref	prefix
prep	preposition / prepositional
pron	pronoun
pl	plural
ps	passive
q	question (clitic)
reflex	reflexive
rel	relative
s	sound (plural)
s-stem	suffix-stem (of the verb)

sng	singular
subj	subject
trans	transitive
V	vowel
v	verb
Ar.	Arabic
Aram.	Aramaic
Akk.	Akkadian
Ber.	Berber
Eng.	English
Fr.	French
Gk.	Greek
H.	Hindi
It.	Italian
L.	Latin
Pers.	Persian
Sp.	Spanish
Syr.	Syriac
Sum.	Sumerian
T.	Turkish

Transliteration and transcription conventions

1. ARABIC TEXT

All Arabic speech data in this book have been transcribed using the italic symbols in the left-most column of the chart in the next section, which is the system used in the *Zeitschrift für arabische Linguistik*. The corresponding IPA symbols are in the next column, and a phonetic description of each sound, using traditional nomenclature, is in the third column. The chart summarizes all the Arabic consonant sounds[1] which occur in this book—but of course, not all of them occur in every Arabic dialect. Though some of them are (a) shared by virtually all dialects, and (b) correspond to the same original sounds in Old Arabic[2] (OA), others have different OA correspondences. For example, the grapheme ج (*ǧīm*), whatever its OA realization(s) may have been, corresponds regularly to several different modern dialectal sounds, depending on the dialect: *ǧ, ž, g,* or *y*—so that written Arabic جار 'neighbour' can be pronounced in normal speech as *ǧār* (e.g. in Baghdad), *žār* (Damascus), *gār* (Cairo), or *yār* (Gulf). The grapheme ق (*qāf*) may correspond to any of the dialectal sounds *q, g, ʾ, k, ǧ, ǵ,* or *ġ*, and here the determining factor is (in some cases) not only geography but the social profile of the speaker. So while written Arabic قريب (normally pronounced *qarīb* in formal spoken Arabic) 'near' is typically pronounced *ʾarīb* in Beirut and Damascus speech, *qarīb* in Muscat, and *girīb* in Baghdad, things are more complicated in the small Gulf State of Bahrain: there it can be pronounced *ǧirīb, karīb, garīb*, or *ġarīb*, depending on who is speaking. In Bahrain, the *ǧ* pronunciation is the 'heritage' pronunciation of the older members of the so-called ʿArab (tribally descended Sunni) communities, and occurs, as in this example, in front-vowel environments; the *k* pronunciation is typical of many Baḥārna villages (all Shīʿa) in all environments; the *g* pronunciation occurs in urban Baḥārna speech and 'educated' Baḥārna speech more generally; and *ġarīb* is a purely ʿArab pronunciation in certain categories of word.[3]

[1] A number of non-Arabic sounds/symbols occur in certain chapters in which data from related languages (e.g. Modern South Arabian Languages in chapter 11) are exemplified and compared. These are explained *in situ*.

[2] In this book, 'OA' is a cover term used for varieties of immediately pre- and early Islamic spoken Arabic for which we have almost no direct evidence and few reliable reports. It is, however, clear that the phonology of the OA tribal dialects was not identical with the Classical Arabic system codified by the Arab grammarians some two to three hundred years later (from the late eighth century AD onwards) as we know from the grammarians' own remarks (see §1.3.1). In this book, reconstructed OA forms (see especially chapter 8) are in roman, preceded by an asterisk.

[3] This pronunciation makes the word sound like its exact opposite in meaning, *ġarīb* 'stranger'. But there is no confusion, as *ġarīb* 'stranger' in the ʿArab dialect of Bahrain is pronounced by many speakers as [qarīb]: initial and medial [ɣ] is realized as [q] or [G], so *qanī* 'rich', *qēr* 'other than', *qasal* 'he washed'—there has been a merger of *ġ* and *q* in non-final position in the ʿArab dialect (Holes 2016: 53–4). The late Sheikh Isa bin Salmān, Ruler of Bahrain (1961–99), would often, even in formal public speeches, confuse *q* and *ġ* in this way, e.g. he pronounced *taqaddum* 'progress' as *taġaddum*.

Looking at things the other way round, a single dialectal phoneme can be 'multivalent' across different dialects: the sound *g*, for example, corresponds to historical ج in Cairo, most of Oman, and a large part of Yemen, but in other dialects it corresponds to historical ق, e.g. those of southern Egypt, those of most of Arabia, and all 'bedouin' Maghrebi dialects; and it also occurs widely in borrowings which are not originally Arabic at all. Details of the phonological systems of the individual dialects dealt with in this book are explained in each chapter where they are relevant to the subject at hand.

CONSONANTS

Symbol	IPA value	Phonetic description
ʾ	[ʔ]	glottal plosive
b	[b]	voiced bilabial plosive
p	[p]	voiceless bilabial plosive[4]
t	[t]	voiceless dental-alveolar plosive
t̲	[θ]	voiceless interdental fricative
ǧ	[ʤ]	voiced palato-alveolar affricate[5]
ž	[ʒ]	voiced palato-alveolar fricative
ḥ	[ħ]	voiceless pharyngeal fricative
x	[x]	voiceless velar fricative
d	[d]	voiced dental-alveolar plosive
d̲	[ð]	voiced interdental fricative
r	[r]	voiced dental-alveolar tap/rolled
z	[z]	voiced dental-alveolar fricative
s	[s]	voiceless dental-alveolar fricative
š	[ʃ]	voiceless palato-alveolar fricative
ṣ	[sˤ]	voiceless velarized[6] dental-alveolar fricative
ḍ	[ɗˤ]	voiced velarized dental-alveolar plosive
ṭ	[ʈˤ]	voiceless velarized dental-alveolar plosive
ḏ̲	[ðˤ]	voiced velarized interdental fricative
ẓ	[ʐˤ]	voiced velarized dental-alveolar fricative[7]

[4] *p* occurs only in foreign borrowings in some Arabic dialects and is sometimes replaced by *b*, e.g. *pāča/bāča* 'offal stew' (Iraq) < Pers. *pāčā* 'feet of sheep, calves or other animals, especially when boiled'; *banka* 'fan' (Gulf) < H. *pankhā; bančar* 'puncture' (Iraq, Gulf) < Eng. *puncture*; *paxxāxa* or *pakka* 'chameleon' (Morocco) (unknown etymology).

[5] As already noted, in some eastern Arabian dialects *ǧ* can be a reflex of *g* < OA *q*, e.g. *ǧilīl* 'little, few', and forms a pair of alveolar affricates with *č* < OA *k*.

[6] Also termed 'emphatic', 'emphatized', or 'pharyngealized' and indicated by a subscript dot. In relevant contexts, secondary velarization of other consonants is also so indicated, e.g. North African *bḅᵂa* 'daddy', *mṃᵂi* 'mummy'.

[7] *ẓ* occurs in Egypt and the Levant as a reflex of *ḏ̲*.

Symbol	IPA value	Phonetic description
ʿ	[ʕ]	voiced pharyngeal fricative
ġ	[ɣ]	voiced velar fricative
f	[f]	voiceless labiodental fricative
q	[q]	voiceless uvular plosive
g	[g]	voiced velar plosive
ǵ	[dz]	voiced dental affricate[8]
k	[k]	voiceless velar plosive
č	[ʧ]	voiceless palato-alveolar affricate[9]
ć	[ʦ]	voiceless dental affricate[10]
l	[l]	voiced dental-alveolar lateral
m	[m]	voiced bilabial nasal
n	[n]	voiced dental-alveolar nasal
h	[h]	voiceless glottal fricative
w	[w]	voiced bilabial glide
v	[v]	voiced labiodental fricative[11]
y	[j]	voiced palatal glide

VOWELS AND SYLLABLE STRUCTURE

The vowel systems of some Arabic dialects bear a resemblance to that of Classical Arabic (CLA), but others are quite different. Some, like CLA, have a system of three short vowel phonemes (*a, i, u*) and three long (*ā, ī, ū*). But others have three short and five long, with the addition of *ē* and *ō*, derived from historical *ay and *aw respectively. A major difference with CLA is that in many dialects, which unlike CLA are natively spoken living forms of speech, non-final unstressed short vowels are liable to neutralization (to ə) and/or to deletion, and long vowels in certain positions to shortening, but the conditions in which these processes occur differ from one

[8] *ǵ* occurs in the dialects of Najd, central Saudi Arabia, as a reflex of *g* < OA *q* in front vowel environments, e.g. *ǵīma* 'value'.

[9] In many dialects of northern and eastern Arabia including the Gulf, southern Iraq, and parts of the Fertile Crescent, *č* is a common reflex of OA *k*, e.g. *čibīr* 'big'. In the 'bedouin'-descended dialects it normally occurs in front-vowel environments only, but in some 'sedentary' dialects of eastern Arabia and central Palestine, it can occur in any vocalic environment.

[10] *ć* is the voiceless member of the *ǵ-ć* pair of dental affricates in Najd and is a reflex of OA *k* in front-vowel environments, e.g. *ćibīr* 'big'.

[11] *v* is a relatively rare sound in Arabic dialects. It occurs regularly in some Turkish Arabic dialects as a reflex of *f* where *f* occurs in contiguity with a voiced consonant, e.g. *vzaʿt* < *fzaʿt* 'I was afraid'; but in Ḥassāniyya (the Arabic dialect of Mauritania) it is the *normal* reflex of OA *f* in most positions, e.g. *vrīg* pl *vərgān* 'encampment', the *f* allophone occurring in only a limited range of environments.

dialect to another. In particular, some Maghrebi dialects allow vowel deletion and consequential consonant clusters to a much greater extent than is true of Mashreqi ones, and this is a very salient feature of how they sound. In the Mashreq, many 'bedouin' dialects have a general rule which disallows sequences of more than two short syllables, e.g. CV-CV-CV(C), which becomes CCV-CV(C) or CVC-CV(C), and restricts the occurrence of *a* in open syllable to certain consonantal environments. There are many other such local phonological rules,[12] and, where relevant, these are commented on in the chapters which follow. In general, however, the examples are transcribed in a broad system, except in a few cases.[13]

GLOSS LINES

Where they are used, gloss lines employ the abbreviations listed earlier, and include only as much detail as is required for the morphological analysis in a particular case, which may differ from one chapter to another, e.g.

(from chapter 8)
al-insān lo ma yi-ba *yi-mūt* *b-yi-mūt* *guwwa*
the-man if not 3msng-want 3msng-die *b*-3msng-die force
'If man doesn't want to die, he'll die against his will'

(from chapter 9)
ḥak-at-li *ḥakkōy-ət* *lə* *staxbar-tū-wa* *ʿala-ya*
tell.s-stem-3fsng-to-1sng story-constr rel ask.s-stem-1sng-3fsng on-3fsng
'She told me the story about which I had asked her.'

2. PROPER NAMES IN THE TEXT AND BIBLIOGRAPHIC REFERENCES

Well-known place and other proper names are spelled in accordance with normal practice, so Cairo, Baghdad, Damascus. Less well-known ones are in a modified version of the orientalist convention, in which long vowels are marked with a macron, velarized consonants with a subscript dot, e.g. al-Fusṭāṭ, but there are some differences from the (italicized) transliteration of actual Arabic text: th is used rather than *ṯ*, dh rather than *ḏ*, kh rather than *x*, j rather than *ǧ*, sh rather than *š*, gh rather than *ġ*.[14] The same applies to the names of Arab historical figures and authors, e.g. al-Jāḥiḍ, Ibn Khaldūn, al-Qalqashandī, and to the titles of books transliterated from the Arabic, e.g. Al-Taghribirdī's *Al-Nujūm al-Zāhira fī Mulūk Miṣr wa l-Qāhira*.

[12] For example, the so-called *buḳara-* and *gahawa*-syndromes (see Glossary).

[13] In chapter 10, for example, *ä* indicates a raised variety of short *a*, and *ĕ* a front variety of intermediate height.

[14] Where, in a few cases, alternative symbols are used in the quotation of the work of others, e.g. ʤ (for j), these are left as is, as the phonetic value intended is obvious.

The contributors

Jordi Aguadé is holder of the Chair of Moroccan Arabic at the University of Cádiz, Spain, and has published widely on the Moroccan dialects, particularly of Skoura and Casablanca. Professor Aguadé is co-author of a Moroccan-Spanish dictionary and was co-editor of the *Estudios de Dialectología Norteafricana y Andalusí* and of the collection of essays entitled *Peuplement et Arabisation au Maghreb Occidental*. He is also editor of the *Taʾrīkh* ('History') of Ibn Ḥabīb—the first Arab chronicle written in al-Andalus (Islamic Spain)—as well as author of several articles on early Islamic eschatology.

Peter Behnstedt, PhD in Romance Philology (University of Tübingen), Dr. habil. in Arabic studies (University of Hamburg), has been a research fellow in various German universities (Tübingen, Heidelberg, Erlangen, Hamburg). He lived and worked in the Arab world (Egypt, Yemen, Syria, Tunisia, Morocco) from 1974 to 1988. He has co-authored with Manfred Woidich the five volumes on Egyptian Arabic within the Tübinger Atlas des Vorderen Orients (TAVO) project of Tübingen University, an introduction to Arabic dialect geography, and the three volumes of the *Wortatlas der arabischen Dialekte* (*WAD*). He has also published dialect atlases of Syria and Yemen, dialect monographs, and glossaries, as well as articles on Tunisian, Moroccan, Palestinian, and Sudanese Arabic.

Ignacio Ferrando is Professor of Arabic Language at the University of Cádiz, Spain, where he has been teaching Arabic grammar and related subjects since 1997. He has published extensively on the Arabic language, including an *Introduction to the History of the Arabic Language* (in Spanish) and a *Comprehensive Arabic-Spanish Dictionary* (in collaboration with Federico Corriente), as well as a number of papers in Spanish, French, English, and Arabic on various lexical, syntactic, and morphological aspects of Arabic, notably the so-called broken plural and the history of Arabic with a focus on its patterns of evolution. His current areas of research interest include Andalusi Arabic from both synchronic and diachronic perspectives.

Clive Holes was Khālid bin ʿAbdullah Āl Saʿūd Professor for the Study of the Contemporary Arab World, University of Oxford (1997–2014), and before that Lecturer, then Reader in Arabic at the University of Cambridge (1987–96). He is now an Emeritus Professorial Fellow of Magdalen College, Oxford. He lived and worked in the Arab world (Bahrain, Kuwait, Oman, Iraq, Algeria) for ten years in the 1970s and 1980s and has published extensively on the Arabic language, sociolinguistics, and dialectology, notably *Dialect, Culture and Society in Eastern Arabia* (Brill, 2001–16, 3 volumes) and *Modern Arabic: Structures, Functions and Varieties* (2nd edition Georgetown, 2004). He has also co-authored two volumes of annotated English translations of dialectal poetry by poets from Jordan, Sinai, and the UAE. He was a member of the advisory board of the *Encyclopedia of Arabic Language and Linguistics* (2006–9). Professor Holes was elected a Fellow of the British Academy in 2002.

Geoffrey Khan is Regius Professor of Hebrew at the University of Cambridge. His research includes philological and linguistic studies of Hebrew, Aramaic, and Arabic. He has a broad

interest in all periods of the Hebrew language and is editor-in-chief of the *Encyclopedia of Hebrew Language and Linguistics*, which was first published in 2013 by Brill. His interests in Aramaic focus for the most part on the documentation of spoken Neo-Aramaic dialects. In the field of Arabic his main published research has been on medieval Arabic documents. His work on Arabic language includes studies on the Judaeo-Arabic of various periods. Among his honours are election as Fellow of the British Academy (1998) and election as Honorary Fellow of the Academy of the Hebrew Language (2011).

Jérôme Lentin is Professor Emeritus of Near Eastern Arabic at the Institut National des Langues et Civilisations Orientales (INALCO), Paris. His research interests include Arabic dialectology and sociolinguistics, Middle Arabic, and Comparative Semitics. He has been working and publishing extensively on the Arabic dialect of Damascus, a city where he lived for a dozen years in the 1970s and 1980s and is preparing, with Claude Salamé, a *Dictionnaire d'arabe dialectal syrien, parler de Damas*. In 2004, he founded, with Jacques Grand'Henry, the International Association for the Study of Middle and Mixed Arabic (AIMA), over which he now presides. He was a member of the advisory board of the *Encyclopedia of Arabic Language and Linguistics* and, since 1997, has been president of the Groupe Linguistique d'Études Chamito-sémitiques (GLECS).

Jonathan Owens is retired professor of Arabic linguistics at Bayreuth University, Germany. Besides Bayreuth, he has worked in the Arabic-speaking world (Libya, Jordan, North East Nigeria), and in the USA. His interests cover various aspects of Arabic linguistics, including the Arabic linguistic tradition, the history of Arabic, Arabic sociolinguistics, and Arabic dialects, especially the western Sudanic dialects. His books include *The Foundations of Grammar: An Introduction to Medieval Arabic Grammatical Theory* (Benjamins, 1988) and *A Linguistic History of Arabic* (Oxford University Press, 2006–9), and he is editor of the *Oxford Handbook of Arabic Linguistics* (Oxford University Press, 2013).

Stephan Procházka studied Arabic and Turkish in Vienna, Tunis, and Istanbul. Since 2006 he has held the Chair of Arabic at the University of Vienna. The main foci of his research are Arabic dialectology and the popular culture of the Arab world. He has published several books and numerous articles on Arabic dialectology, including comparative and syntactic studies as well as descriptions of local dialects, based on data gathered in the course of long-term fieldwork. Many of his studies are dedicated to the different Arabic varieties spoken in Southern Turkey. He has served as the leader of a number of externally funded projects devoted to the study of the Tunisian and Moroccan Arabic dialects. Professor Procházka was elected a Fellow of the Austrian Academy of Sciences in 2017.

Catherine Taine-Cheikh entered Le Centre National de la Recherche Scientifique (CNRS) in 1996, after having been a researcher based in Nouakchott, Mauritania (1978–94). She is now Emeritus Director of Research at LACITO (*Laboratoire Langues et Civilisations à Traditions Orales* of the CNRS). She initially studied the spoken Arabic of Mauritania, Ḥassāniyya, before embarking on the study of Zenaga, a variety of Berber spoken in Mauritania which at that time was rapidly disappearing. She has published dictionaries of both Ḥassāniyya and Zenaga, as well as numerous articles in the fields of descriptive and theoretical linguistics, comparative dialectology, linguistic typology, and historical linguistics. The first two volumes of her *Études de linguistique ouest-saharienne* (volume 1 : *Sociolinguistique de l'aire hassanophone* (2016),

volume 2 : *Onomastique, poésie et traditions orales* (2017)) have recently been published in Morocco. These contain revised editions of twenty-seven of her previously published articles.

Janet Watson studied Arabic and Islamic Studies at the University of Exeter, and then moved to SOAS, London, to study linguistics and complete a PhD on the phonology and morphology of Yemeni Arabic dialects. She has published widely on Arabic linguistics and the dialects of Yemen, and since 2008 has been working on the Modern South Arabian Languages (MSALs). Her book, *The Structure of Mehri*, was published in 2012. She has held academic posts at the Universities of Edinburgh, Durham, and Salford, and held visiting posts at the universities of Heidelberg (2003–4) and Oslo (2004–5). She took up the Leadership Chair for Language@-Leeds at the University of Leeds in 2013, and was elected a Fellow of the British Academy in the same year.

Manfred Woidich was Professor of Arabic Language at the University of Amsterdam (1983–2008). He lived and worked in Egypt for eleven years from 1969 until the mid-1980s and has been doing fieldwork there ever since both in Cairo and in rural Egypt. He has published very extensively on Egyptian Arabic, notably on Cairo, Upper Egypt, and the Oases of the Western Desert. His publications include *Das Kairenisch-Arabische: Eine Grammatik* (2006). With Peter Behnstedt he co-authored the five volumes of *Die ägyptischen-arabischen Dialekte* (1984–99), *Arabische Dialektgeographie* (2005), and the three volumes of the *Wortatlas der arabischen Dialekte* (2011, 2012, 2014). He was an associate editor of the *Encyclopedia of Arabic Language and Linguistics* (2006–9). Professor Woidich was elected a Corresponding Fellow of the British Academy in 2017.

1

Introduction

CLIVE HOLES

This book is a first step down the road of describing the history of Arabic as a *spoken* language. It would obviously be impossible in a book of this modest length to attempt to cover everything that has a claim to importance, whether from the point of view of dialect geography, history, or language processes. The ten chapters which follow all treat substantial topics and are by leading authorities in their fields. Each has a distinct focus, but there are many shared elements and themes, and cross-referencing enables the interested reader to follow them across the whole work. The bibliography is comprehensive and a glossary of Arabic linguistic, cultural, and historical terms provides the non-specialist with a 'one-stop shop' for quick reference.

The authors were asked to adopt a historical perspective in their descriptions and analyses, though the degree to which this is feasible depends in part on the availability of reliable historical sources, which is highly variable. The time frame covered is from the advent of Islam in the early seventh century AD[1] up to the present day. Seven of the chapters are geographical in focus (Taine-Cheikh on Mauritania; Aguadé on the Maghreb; Behnstedt and Woidich on Egypt; Lentin on the Levant; Procházka on the Northern Fertile Crescent; Holes on the Gulf; Watson on South Arabia); one is 'communal' (Khan on Judaeo-Arabic); and two each deal with an individual linguistic feature (Owens on the origin and evolution of the *b-* verbal prefix in the Gulf, the Levant, Egypt, Yemen, and Nigeria; Ferrando on the 'adnominal linker' *-an* in medieval Andalusian Arabic, similar forms of which, with identical functions, occur in other old and geographically peripheral dialects of Arabic, as far away as Central Asia). Unfortunately, it was not possible to obtain separate chapters on the linguistic history of Malta and the Sudan, which exemplify important leitmotifs not covered elsewhere in the book: Malta, because it is a (literal) example of an Arabic *Sprachinsel* which, around the end of the eleventh century, was summarily cut off from 'heartland' Arab linguistic influence and has thus preserved evidence of what the spoken Arabic of North Africa was like a millennium ago; Sudan, because in

[1] Al-Jallad (2018) gives a summarizing overview of what is known about Arabic in the centuries before the advent of Islam.

Arabic Historical Dialectology. First edition. Clive Holes (ed.).

medieval times it acted as a conduit for migrations from Egypt and the Arabian Peninsula to the Sāḥil region of West Africa. Certain Sudanese dialects (Reichmuth 1983) have absorbed linguistic influences from both, and are, it would appear, the link in the diffusion of certain unusual dialect features from Arabia to the Sāḥil and northern Nigeria. Sudan is important in the history of Arabic for another reason: it is one of the few places in the Arabic-speaking world where a process of full-blown pidginization and creolization occurred, the results of which have survived up to the present day. The process started in nineteenth-century military camps, and produced so-called Juba Arabic (for a description, see Manfredi 2017).

The main purpose of this introductory chapter is to outline, principally for the benefit of the non-specialist reader, the cultural hinterland of Arabic language history. Many issues discussed in detail in the individual chapters are mentioned here in passing and briefly exemplified, but this introduction is not a summary of their arguments, far from it; indeed, some authors disagree with each other on how certain linguistic developments in Arabic are best explained. The objective here is rather to set the chapters as a whole in their cultural, historical, and scholarly context, and, as the subtitle of the book suggests, adumbrate the linguistic and sociolinguistic approaches to Arabic dialectology which underpin them. Also included here, where relevant, are historical data from Arab lands not specifically covered in any individual chapter (e.g. Iraq), and certain general topics not dealt with in detail either (e.g. dialect typology, indexicalization, pidginization).

1.1 THE EARLY HISTORY OF ARABIC: MYTH AND REALITY

The history of Arabic is long, complicated, and in some periods and locations, obscure. But its history in the imagination of the ordinary man in the Arab street is anything but: it bursts on to the scene, seemingly from nowhere, with the seventh-century Revelation of God's word to Muḥammad, the Arabian Prophet, vouchsafed by the Angel Gabriel, and recited by him in Arabic, at first to a sceptical audience, and later to a growing band of believers. Later collected together and written down, these oral recitations, 'sent down' over a twenty-two-year period, became the scriptures of the new religion of Islam. Though the Arabian poets of the pre-Islamic 'Time of Ignorance' (*al-ǧāhiliyya*) were and still are celebrated for the magnificence of their mono-rhyming odes (*qaṣāʾid*), it was the revelation of the Qurʾān, and *in Arabic*, which, in the popular imagination, moved the language and the people who spoke it to centre stage in world history.

The first attestation of the word *qurʾān* is in the Qurʾān itself, where it means 'reading' or 'recitation'. It may have been calqued on the cognate Syriac word *ḳeryānā* 'scripture reading, lesson', and modelled on the similar Arabic verbal noun pattern CuCCān. The word *ʿarabiyy* ('Arab', 'Arabic') occurs eleven times in the Qurʾān, and in six of these instances it specifies that the Prophet's recitation is an 'Arabic Qurʾān' (*qurʾān-un ʿarabiyy-un*) and, furthermore, 'an Arabic Qurʾān that contains no

crookedness'[2] (*qurʾān-an ʿarabiyy-an ġayra ḏī ʿiwağ-in*). In five others it collocates with other nouns: its orality is suggested by the phrase 'a clear Arabic tongue' (*lisān-un ʿarabiyy-un mubīn-un*); and its purpose was to provide 'a decisive Arab/Arabic judgement' (*ḥukm-an ʿarabiyy-an*).

That the Qurʾān was revealed through the medium of Arabic has thus, from the very beginning, been an intrinsic part of its message. But whether the word *ʿarabiyy* 'Arab, Arabic' meant at the time of the Revelation what it means today is, among the scholarly community, a moot point. Retsö, in his monumental study of the Arabs in antiquity,[3] contends that *ʿarabiyy* in the Qurʾān probably 'refers to a language connected with the *ʿarab* which was known as a vehicle for messages from the non-human world' and was not simply 'the normal everyday speech in Mecca and its surroundings' (2003: 592). In his opinion, another rather problematic word, *ʾaʿğamī*, which in the Islamic Middle Ages came to be used to designate 'speaker of a language other than Arabic' (e.g. of Greek, Persian, etc) is in the Qurʾān used to refer to forms of Arabian speech which deviated from *ʿarabiyy*, and may have referred to the Arabian speech in everyday use at the time (Retsö 2013: 434–5). In the same vein, Webb (2014: 156ff.), on the strength of another trawl through the early sources, and a study of the words derived from the root *ʿ-r-b* in the Qurʾān, claims that *ʿarabiyy* may have originally denoted not a community of speakers or their language but rather God's message, and one which the Qurʾān repeatedly states that the *aʿrāb* ('bedouin') largely failed to embrace, making it therefore illogical, in his view, that the relational adjective *ʿarabiyy* should be understood as relating to their language: '… the Qurʾān's "*ʿarabiyy*" is an adjective of revelation, not a people. "*ʿArabiyy*" is an adjective for the Qurʾān's sacred idiom…. The Qurʾān's conception of Arabic is that of a language possessing miraculous clarity that conveys the Sacred Message, and prompts its listeners to comprehend and respond by embracing Islam'. Self-evidently, the text of the Qurʾān is in a variety of the language we now call 'Arabic', but, even if one remains sceptical of the Retsö/Webb hypothesis that it had once a special sacred or esoteric sense, the word *ʿarabiyy* at the time of the Revelation may indeed have had a different meaning from what it has now or had in the early Islamic centuries.

It is indeed probable that the meaning of *ʿarab* and *ʿarabiyy* as 'Arab/Arabic' in the general ethnic and linguistic senses we know today developed late—well after the death of the Prophet—and were not current in pre-Islamic Arabia. Macdonald (2009) has documented the wide range of peoples and communities to which these terms were applied before late antiquity (i.e. before about AD 106), which varied depending on who was using them, for what purpose, and in what context. There seems at this point to have been no unified concept of what 'Arabs' were: the term was applied by outsiders of many different backgrounds to peoples some of whom didn't speak Arabic at all, and many of whom, to judge from the descriptions, were not nomadic pastoralists either. As a *self-descriptor*, the ethnonym 'Arab' at this

[2] The translations are those of Pickthall.

[3] For a critical but generally very positive review of this landmark work see Donner 2007.

period was rare. If one can generalize, it seems to have often been applied as an unflattering term to 'the other'—an exotic stranger not part of one's own community. Hoyland (2015: 26) notes that in the famous funerary inscription at Namāra, 120 km south-east of Damascus, dating to AD 328—much later than Macdonald's material but still three centuries before the Revelation—a certain Marʾu l-Qays bar ʿAmrū is praised as being 'king of all the Arabs' (*mlk ʾl-ʿrb kl-h*). This text is written in Nabatean (i.e. Aramaic) script but the language seems to be Arabic (except for *bar,* Aramaic 'son', for Arabic *ibn*, and the word *ʿkdy* ('thereafter' (?)), which has not been attested in any variety of Islamic-period Arabic). But who were these fourth-century 'Arabs' that Marʾu l-Qays was king of, and where did they live?

Schiettecatte and Arbach (2016) have outlined the tribal geography of Arabia in *c.* the mid-third century, based on a recently discovered Sabaic inscription. This locates the 'Arab' tribes of Asad$^{\bar{a}n}$, Nizār, Maḏḥiǧ, Maʿadd, and Ṭayy in the north and west of Arabia. Three of these, Asad, Maḏḥiǧ, and Maʿadd are also mentioned by name in the Namāra tomb inscription as having been 'subdued' or 'mastered' by Marʾu l-Qays. This suggests a degree of stability in the demography of Arabia through at least the third and fourth centuries. In the Qurʾān, however, revealed three centuries later in the early to mid-seventh century AD, the noun *ʿarab* (in contrast with its adjective *ʿarabiyy*) is absent. This is surprising, but probably not significant, as by this time the generic concept of 'Arabs' based on a common language and shared elements of culture (such as the tribal poetry which was then circulating) must have existed at least in embryonic form. More likely than any esoteric meaning, it seems to this writer that the Qurʾān's repeated insistence on the 'Arab/Arabic' nature of its message was intended to promote a distinctive linguistic facet of this emerging shared culture, to which it was now adding a new religious dimension. As noted earlier, in several of the verses where it occurs in the Qurʾān, the word *ʿarabiyy* is further specified as something 'clear', something that everybody who listened to it would easily understand.[4] It is significant in this context that the Qurʾān delivers a damning judgement on poets, famous as clever wordsmiths and 'seers of the unseen': 'As for poets, the erring follow them. Hast thou not seen how they stray in every valley, and how they say that which they do not?' (Q 26: 223–5). The Prophet's own expressed opinions were similar: 'It is better for a man to fill the inside of his body with pus than to fill it with poetry' (the *Ṣaḥīḥ* of al-Bukhārī Volume 8, Book 73,

[4] Recently, the standard translation of *mubīn* as 'clear', and the assumption behind this translation that this 'clarity' inheres in the fact that the Revelation was in Arabic has been challenged (Dichy 2009): 'Si l'on compare cette fois le participe *mubīn* et l'adjectif *bayyin*, qui exprime l'état caractéristique de « clarté », une différence cruciale se fait jour. Du fait qu'il est un participe actif, *mubīn* peut prendre deux valeurs:

- la valeur d'état résultatif, qui est produite par un événement accompli (et donc passé): le Livre est dit alors « ayant fait la clarté », à partir de l'événement de sa propre énonciation;
- la valeur progressive, qui dénote un événement en cours: le Livre est dit, mot-à-mot, « faisant la clarté » en s'énonçant, ou « en train de faire la clarté »'.

It is not, on this reading, the linguistic fact of 'being in Arabic' which renders the message 'clear'; the act of 'making clear' the message is a process, and 'clarity' a result, which are both predicated of the Divine, and achieved through the agency of His Prophet's revelations, and not through the choice of language forms in which this message is vouchsafed.

No 175). This poetry was composed in the same variety of Arabic as the Qurʾān, but the Prophet wished at all costs to distinguish his message from it; for him, Arabic poetry was a dangerous waste of time, lies even, and in that respect quite unlike the Qurʾān, whose message was described by itself as clear and straightforward.[5]

The Revelation of the Qurʾān was one of two events of the greatest linguistic significance in the history of Arabic; the other was the Arab conquests which began shortly after the death of the Prophet in 632 and were largely complete by about 715. The Arab armies of conquest and their camp followers exported Arabic to a vast region stretching from Spain in the west to Central Asia in the east. But what kind of Arabic was it that they spoke? We have almost no direct contemporaneous evidence. Modern popular culture, as evinced in the 1976 film *al-Risāla* ('the Message' in the English language version) and countless TV historical dramas, depicts the Arabs of that time as speaking perfect Classical Arabic (CLA) at all times with full case- and mood-endings. Many native speakers still believe implicitly that this cinematic scenario reflects the reality, and it has even been espoused by some academic historians of the language: cf. Versteegh 1984: 17 'before the coming of Islam there was a single Arabic language, which was used both as a colloquial and literary language'. On this view, CLA was there *ab initio* as a spoken as well as a written language and was gradually corrupted, owing to imperfect learning by the non-Arab subjects of the early Islamic Empire,[6] who vastly outnumbered the native Arabic speakers who formed the first wave of migrants. The end point of this process, according to the popular belief, was the Arabic dialects spoken today. This account—of a distant golden age of linguistic perfection corrupted by foreigners—is one of several popular myths about the language:[7] for many Arabs, 'the Arabic language' (*al-luġa l-ʿarabiyya*) refers only to the modern version of this 'pure and eloquent' (*faṣīḥ*) form of the language, and not to the 'corrupt' forms they themselves use in everyday speech (which they refer to more informally as *ʿarabiyy*).

If that is the myth, what was the reality, insofar as we can glean it from the historical record? We know that initially, Greek, Coptic, and Persian were retained as *written* languages for administrative purposes in the territories of the Byzantine and

[5] By the end of the ninth century, another linguistic facet of the Qurʾān, the 'inimitability' (*iʿǧāz*) of its style, had become established doctrine, and in consequence it came to represent for (Muslim) Arabs the apogee of *faṣāḥa*, a word which combines notions of purity and eloquence. The traditions of melodic Qurʾānic cantillation and the artistic calligraphy of the text which developed subsequently have become ubiquitous aspects of Arabo-Muslim religious culture, and learning to recite the Qurʾān from memory, beginning at the age of six or seven at religious schools (the *kuttāb*) was the foundation of pre-modern education all over the Arab world. The Qurʾān is the only Arabic text invariably written with full vocalization and the full set of orthographic symbols, so as to ensure its accurate rendition.

[6] Versteegh (1984) went further and hypothesized that CLA underwent a process of mass pidginization and creolization after the conquests which resulted in the Arabic dialects. There is, however, no evidence that pidginization and creolization in the normally accepted definition of these terms occurred on anything like the scale envisaged. For counter-arguments, see the reviews of Goodman 1986, Holes 1986b; 2004a: 23–9, and Hopkins 1988.

[7] For an amusing account of several others, see Ferguson 1968 [1959]. Our main purpose in writing this book is to show how a more historically and socially grounded linguistic approach, despite the gaps in the historical record, can help trace the long-term dynamics and some of the detail of what happened.

Sasanian empires which the Arabs had conquered. Then, at the turn of the eighth century, they were phased out, and Arabic was imposed as the language of state records (Hoyland 2015: 213–14; *EI* art. Ḳibṭ). As far as *spoken* languages are concerned, the big picture was of widespread bilingualism/multilingualism, which persisted in some parts of the conquered territories for very many centuries. This has continued so up to the present day in parts of North Africa, where various varieties of Berber never died out in the mountainous regions, in contrast to Vulgar Latin, which quickly disappeared from the coastal plains. Coptic seems to have survived for longer in Middle and Upper Egypt, but by no later than *c.*1300 the Christian Copts had shifted to Arabic (Richter 2006: 495).[8] In parts of Lebanon, Syria, Iraq, and Turkey, however, pockets of neo-Aramaic speakers still survive.[9] Other large land masses—Persian- and Turkish-speaking Central Asia—were more thinly colonized and simply too vast ever to be completely (or even largely) Arabicized at the level of the ordinary population. In these areas, the indigenous languages eventually re-emerged as the literary standard as well as remaining the main spoken languages, but some pockets of Arabic-speakers still live on there (Seeger 2002; Ingham 2006; Zimmermann 2009).

From around the middle of the eighth century, the Arab grammarians of southern Iraq began codifying the rules of CLA—what eventually (but long after their time) became known as *al-luġatu l-fuṣḥā*, 'the pure and eloquent language'. The timing of this enterprise was not fortuitous. Although the grammarians embarked upon it for reasons which they never make explicit, prime among them must have been the need to standardize the language for governmental, administrative, legal, religious, and literary use in what had by then become a geographically far-flung and ethnically diverse empire, without which written Arabic might have developed regionally in undesirably uncontrolled ways, or might not have been used at all. The grammarians were highly prescriptive: their aim was to specify, and enjoin the use of, the 'best' language, as it had come down to them in various sources: the ancient pre-Islamic poetry, the Qurʾān itself, the *ḥadīṯ* and *sīra* literature,[10] early accounts of the *ayyām al-ʿarab,*[11] and by direct elicitation from bedouin tribesmen of their time, whom they judged still to speak the 'purest' (*ʾafṣaḥ*) and 'most trustworthy' (*mawṯūq bihā*) Arabic from the range of tribal dialects known to them. We know from their accounts and those of other early writers, such as the Iraqi polymath al-Jāḥiḏ (776–868), that dialectal variation in the spoken Arabic of their time was rife,[12]

[8] The date when Coptic finally ceased to be a living language is, however, disputed. It has been claimed that there were still villages in Upper Egypt which were Coptic-speaking in the eighteenth century (see *EI* art. Ḳibṭ).

[9] Or did until recently. It is unclear at the time of writing (summer 2017) whether the activities of the so-called Islamic State have effectively ended their existence in Syria and Iraq.

[10] *ḥadīṯ* pl *aḥādīṯ* are the authenticated 'sayings' of the Prophet, of which there are several large collections. The *sīra* literature consists of early accounts of the Prophet's life.

[11] lit. 'the days of the Arabs': early quasi-legendary accounts of pre-Islamic battles fought between the Arab tribes.

[12] For example, in his treatise *Kitāb al-Bayān wa l-Tabyīn* ('The Book of Eloquence and Exposition'), al-Jāḥiḏ notes in passing the pronunciation of /r/ as /ġ/ by certain elite communities in the Baghdad of his time, a feature which still survives in some dialects of Iraq and the Fertile Crescent today (see Procházka, this volume).

even if they generally give little or no detail of it and what they do say is confused and sometimes even contradictory.[13] As Rabin (1951: 13) has commented:

> It would be difficult, if not impossible, to discover why the philologists recorded just those dialect features they did. There is certainly no system in it. They never considered dialects as a form of speech in their own right, but as a collection of deviations from the literary language. All their data are measured on CLA, and we can often see quite clearly that they failed to see anything that did not fall within the categories of that idiom . . . The net result is that we have a great deal of information on minor points of dialect usage, but get only occasional glimpses of the major forms. We cannot reconstruct the complete paradigm of any tense in any dialect; we can hardly say with certainty what a complete word may have sounded like. The few glimpses we obtain prove that there were profound differences, the full nature of which will probably never be revealed to us.

This attested yet tantalizingly undescribed dialectal variation[14] cannot have appeared out of the blue: if this was the situation in the late eighth century, the spoken dialects must have been evolving since well before that date, and, indeed, well before the revelation of the Qurʾān—an inference supported by an ever-growing body of evidence from the decipherment of early inscriptions.[15] If so, the CLA of the eighth-century grammarians was a culturally and politically necessary distillation of an ancient elevated register of the language no longer identifiable (if it ever had been) with any particular dialect, but, as suggested earlier, long used in poetry, soothsaying, and other types of formal ritualized performance, in the way that, *mutatis mutandis*, it still is today. One of the main sources of the grammarians' data, the Qurʾān itself, shows orthographical peculiarities and grammatical deviations from the norms the grammarians subsequently established, which probably reflects the fact that the Meccan vernacular spoken by the Prophet did not fully conform to these later imposed norms, which reflected the usage of central/eastern Arabia, where most of the ancient poets came from, rather than that of the Hejazi coastal towns. If this is correct, dialectal variation in Arabic is again shown to be an ancient, not a recent phenomenon, and not simply the end result of the 'corruption' of CLA as a consequence of the imperfect learning of it by the conquered peoples, as has been proposed in traditional Arab and Western accounts alike. So, in answer to the question posed earlier, the language which the conquerors spoke and took with them to the new

[13] Kofler 1940–2 collected together the scattered information the early grammarians give about the tribal dialects of Arabia.

[14] For an attempt at a historical reconstruction of variation in the pre-Islamic Arabic dialects, see Magidow 2013.

[15] The earliest known reference to an Arab (Krebernik 2008: 257) mentions a chieftain called 'Gindibu' who contributed a thousand camels to an anti-Assyrian military coalition at the battle of Qarqar. It is also the first known record of an Arabic word: 'Gindibu' = CLA *ǧundub* 'grasshopper', and occurs in a cuneiform inscription on the Kurkh monolith of the neo-Assyrian monarch, Salmanassar III dated to 853 BC. The word *ǧundub* is attested among the Semitic languages only in Arabic, so it appears that Musil's remark (1928: 243) on the naming practices of the Rwala bedouin of the Syrian desert in the 1920s, that 'there is no beast nor plant after which a child cannot be named' may already have applied to the Arabs of two and a half millennia earlier. Inscriptions in Arabic for the first three centuries of the Common Era, though limited in number, also exist. See Al-Jallad 2018 for a summary of what is now known about the structure of Arabic at the very earliest period of its history.

territories outside Arabia must have been various regional/tribal dialects of Arabic.[16] As we shall see in the chapters which follow, there is evidence from a study of the geographical distribution of features in the modern dialects, when compared with what the medieval Arab historians say about population movements, to show that this was indeed probably the case.

Against this, however, a theory put forward by Ferguson (1978 [1959]) contained the proposition that the tribal dialects were koineized in military camps very early on, and that it was this shared koine, which came about 'through a complex process of mutual borrowing and levelling' over the first few Islamic centuries (Ferguson 1978: 51), and not CLA or individual tribal dialects, which formed the basis from which all the modern 'sedentary' Arabic dialects have developed since. How else, Ferguson argued, was it possible to explain how these dialects, now spread far and wide, share so many 'complicated, systematically isolated' features (he selects fourteen), none of which are shared with CLA, if they did not all originate from a common but non-Classical source? The unconnected polygenesis of so many novel and widespread features, he argues, seems highly unlikely. Ferguson excluded from his argument Arabian 'bedouin' dialects, which, he claimed, did not undergo koineization and followed a different trajectory. This theory was rebutted by Cohen (1962). Cohen argued that (a) some of Ferguson's fourteen features could easily be explained as the result of natural linguistic change ('drift') over time; (b) others were far less widely shared than Ferguson claimed; and (c) yet others were a priori too all-encompassing for it to be proposed that they had been formed in a relatively short period of time as part of a koine formed in military camps. In sum, Cohen argued, it looks more likely that Ferguson's features are the recent *end point* of a process of partial convergence than an ancient single *start point*. What has become clear from recent research on the modern 'bedouin' dialects is that many of Ferguson's features occur in them also, even though, according to his theory, they did not undergo the phase of koineization. Nonetheless, they show similar developments to those of the 'sedentary' dialects, even if they have lagged behind in the speed at which these developments have occurred. Furthermore, it has also been shown by recent research that some early regional Arabic dialects were the product of the local mixing of elements from very specific and still easily identifiable locations. An example is the surviving Central Asian Arabic dialects, some of which have maintained, over thirteen centuries, a mixture of features which clearly mark them as of Iraqi and eastern/south-eastern Arabian origin, which is in line with what the early Arab historians had to say about the origins of the soldiery which conquered Transoxania (Holes 2011a; see also §1.3.1). From several points of view, Ferguson's theory of a unitary ancient koine now seems untenable as a general explanation of how the modern Arabic dialects evolved.

Whilst the serious study of CLA in Europe began in the mid-seventeenth century, attention started to be paid to the spoken Arabic dialects only in the late nineteenth

[16] These dialects were not confined to those of Arabia. There is evidence for Arabic having been spoken in the Levant since Roman times, and even much earlier (see Lentin, this volume). The Levant was one of the sources of the early Muslim soldiery.

and early twentieth centuries.[17] Teaching manuals of the major urban dialects (especially Cairene) for the use of European expatriates also began to appear at this time.[18] Meanwhile, the publication in 1906 of Karl Vollers' *Volkssprache und Schriftsprache im alten Arabien* caused an earthquake by claiming that the Qurʾān was recited by the Prophet in a variety of Arabic which lacked the CLA vowel inflections of case and mood (*iʿrāb*), its most distinctive morphological feature, and that the linguistic character of the canonical text we have was actually a post hoc fabrication of the Arab philologists. Vollers' claim was never accepted by most scholars, but the debate it engendered led ultimately to a critical re-examination of the popular view of Arabic language history outlined earlier, and was one of the first steps towards the elaboration of a theory based on actual evidence. This came in two forms: early non-literary written documents which were discovered throughout the course of the twentieth century, and, just as importantly, and complementary to these, an ever-increasing number of detailed descriptions of the modern Arabic dialects. We will briefly examine each of these sources of evidence in turn.

1.2 EARLY WRITTEN EVIDENCE: 'MIDDLE ARABIC'/'MIXED ARABIC'

Non-literary papyri and paper documents in Arabic for the first three Islamic centuries (up to *c.*AD 912) now exist in abundance,[19] and one of the first detailed comparative analyses of their language can be found in Hopkins 1984. Though the range of their provenance is for climatic reasons limited to a few areas (mainly the Nile Valley and Palestine), these texts come in a wide variety of types: private and business letters, property deeds, marriage contracts, administrative surveys, lists and registers, passports, petitions, tax receipts, demands for payment, etc. The very fact that their content is so mundane gives them two advantages over literary material as linguistic evidence: (i) they are likely to reflect a register of the language closer to the

[17] In Germany by such scholars as Wetzstein (1868) (Syrian bedouin), Reinhardt (1894) (Oman), Socin (1900–1) (central Arabia), and Meissner (1903) (southern Iraq), and in France by Marçais (1902, 1908, 1911, among others) (N. Africa), Feghali (1928) (Lebanon), Barthélemy (1935–69) (the Levant), and Cantineau (1934, 1936, 1937, 1946) (Syria). And we must not forget the work of that eccentric but brilliant linguist and ethnographer, the Swedish aristocrat Le Comte de Landberg (1901, 1905, 1909, 1913, 1919, 1920–42) (South and Central Arabia). All of these scholars and more began the task of describing, on the basis of fieldwork done *in situ*, and long before the era of tape-recorders, how ordinary Arabs actually spoke to each other.

[18] Typical early examples are Wahrmund 1880, Spitta-Bey 1880, Vollers 1890, translated into English by Burkitt 1895, Cameron 1892, Spiro Bey 1895, Willmore 1901, and Gairdner 1926. Scholarly efforts to record and teach the Arabic dialects were, and in some quarters still are, dismissed by Arab opinion as part of a Western plot to divide and rule the Arabs through encouraging the use of the regional colloquials. Typical is the comment of Muḥammad Riḍa al-Shabībī, in an article entitled 'The chaos of the dialects' in the *Review of the Egyptian Language Academy* 12 (1966: 135): 'This call [to adopt the colloquial] is basically a colonialist scheme (*dasīsa istiʿmāriyya*) which has only been accepted by a few people terrified by the call of Islam.' Saʿīd (1964) expresses similar sentiments in a book-length treatment of the subject.

[19] At the time of writing, Hopkins (1984: xl) estimated the number of extant early papyri at 16,000 and early paper documents at *c.*33,000, in addition to several hundred texts on leather, parchment, linen, and other materials.

everyday spoken usage of their time; (ii) they have no literary value, so have escaped 'correction' by subsequent generations of editors wishing to make them conform to the norms of codified CLA.[20] There are even among the papyri a few bilingual texts, notably an early Greek–Arabic psalm fragment (Violet 1901) in which the Arabic text has been transcribed into Greek uncials, which was perhaps intended as a crib for those who spoke Arabic and were literate in Greek, but could not read the Arabic script. It is very rare among early documents in that it throws light on phonetic and phonological features, such as vowel quality and the realization of certain consonants (e.g. *qāf*), and morphological ones, such as the presence or absence of case and mood inflections, features not recoverable from ordinary unvocalized texts in the Arabic script. Hopkins' verdict (1984: xlvi) on these early written materials is unequivocal:

> In almost every case in which the language of the Arabic papyri deviated from CLA, it deviated unmistakably in the direction of Middle Arabic, typologically akin to most of the modern colloquials. This language, therefore, lies fully within the mainstream of Middle Arabic, of which it is the earliest representative. A large proportion of the features attested later in medieval Jewish, Christian, and to a lesser extent, Muslim Middle Arabic, many of which are familiar today from modern dialects, occur here for the first time. This fact speaks for a very impressive continuity in colloquial Arabic usage, and the roots of the modern vernaculars are thus seen to lie very deep.

These comments require some scholarly context. The term 'Middle Arabic' first gained currency in the late 1950s/early 60s, in work by Joshua Blau, as a description of the language of medieval texts written in Judaeo-Arabic (= Arabic written by Jews in the Hebrew script, see Khan, this volume) and texts of a generally earlier date written by Christians typically in the Arabic script, but sometimes in the Syriac script, known as *karšūnī* (Blau 1981 [1965], 1966, 1988 (which is a collection of articles published over the previous twenty-five years)). These texts showed many systematic deviations from normative CLA but were seemingly not in 'pure dialect' either, leading Blau to label them the 'missing link' and 'the intermediate stage'[21] between CLA and the modern dialects. At this point in the scholarly debate, Middle Arabic was thus conceived of as a chronological stage in the development of Arabic, inviting analogies to 'Middle English' in the history of English. Subsequent work by many scholars (e.g. Lentin and Grand' Henry 2008), including Blau himself, has, however, highlighted three things:

(a) The divergences of Middle Arabic from CLA which are dialectal are not random, but, in the main, rather specific, typically affecting certain classes of morphological items and syntactic structures, and reflecting typological similarities shared by a wide range of spoken dialects.
(b) Middle Arabic texts contain many forms which are neither Classical nor dialectal but 'hypercorrect' or 'hybrid' (see Khan, this volume), and they occur in texts of different geographical provenances over a long period of

[20] Examples of 'deviant' features in early recensions of the ancient Arabic poetry, by contrast, have indeed often been ironed out in later copies as a result of this process.

[21] Blau 1988: 61. This article ('The importance of Middle Arabic dialects for the history of Arabic') was first published in 1961.

time. Mid-nineteenth-century documents from the Arabian Gulf sheikhdoms, for example (Holes 2008a), show a range of hypercorrect and hybrid features similar to those of texts from Egypt and the Levant of a much earlier period, and, for that matter, to those which occur in 'mixed' styles of Arabic spoken today (Mejdell 2008). They are the result of the long-term coexistence of and contact between two systems, the naturally acquired dialectal one and the institutionally acquired CLA one, and often do not belong to a chronological phase of any spoken dialect. Over a long period, it seems that many of these hypercorrect and hybrid features have become part of the inventory of 'mixed' forms typical of an informal style of writing as well as a semi-formal style of speaking.

(c) There is no linguistically principled way of distinguishing the language of Middle Arabic texts from that of non-CLA texts which predate and others that post-date them: all are composed in varieties of Arabic on a spectrum containing different mixtures of CLA, dialectal, hypercorrect, and hybrid features.

Blau's most recent pronouncement (Blau 1999: 223–4) is that 'MA [Middle Arabic] is not an exclusively medieval phenomenon. It refers rather to Arabic texts of mixed character in general, and MA can be found at the present time as well'. The 'communal' aspect of MA was the dominant focus of debate early on: it was hypothesized that the relative paucity of documents in Middle Arabic written by Muslims compared to Jews and Christians was due to their better knowledge of, and culturally instilled closer adherence to, the norms of CLA. Later discoveries and further work have called this idea into question also. In fact, Muslim Middle Arabic documents exist in quantity, and, with the obvious differences of religion-specific vocabulary, are typologically akin to those written by non-Muslims. It has also become clear that choice of style—whether a text is in 'mixed' Arabic or a style closer to CLA on the one hand, or to dialect on the other—is related to contextual factors: what the text was about, why it was written, and who it was intended for.

A few typical features of early (from *c.*tenth century) and 'Middle Arabic' texts selected from various sources[22] are:

(a) reflecting widespread non-CLA dialectal features:
  (i) in phonology, as reflected in the orthography:
    - loss of the glottal stop in all positions, e.g. بقا (*baqā*) '(continued) life', cf. CLA بقاء (*baqāʾ*), راس (*rās*) 'head', cf. CLA رأس (*raʾs*), الول (*alawwal*) 'the first', cf. CLA الاول (*al-ʾawwal*);
    - initial *ʾalif* to indicate a prosthetic vowel before a consonant-cluster resulting from the deletion of an unstressed short vowel in a CV syllable in initial position, e.g. خمسة احمور (*xamsat əḥmūr*) 'five asses', cf. CLA خمسة حمور (*xamsatu ḥumūr*); اتقول = *itʾūl/ itqūl/ itgūl* (depending on the dialect) 'you (m) say' cf. CLA تقول = *taqūlu/a*.
  (ii) morphosyntax:

[22] Mainly the articles in Lentin and Grand'Henry 2008 and Hopkins 1984. These sources contain examples from as early as the eighth and as late as the nineteenth century.

The NP

- sound pl suffixes not marked for case: always *-īn* not *-ūn/-īn*, and dual *-ayn* (or *-ēn*) not *-ān / -ayn*;
- the *-n* of the sound pl and dual suffixes retained in annexation structures (unlike the situation in CLA), e.g. *ḥāmilīn al-awsāq* 'carriers of burdens', *ṣayyādīn as-samak* 'fishermen', *miyatayn ǧulūd* 'two hundred skins';
- *-in/-an* adnominal linker suffixed to indefinite nouns with an adjunct: *yawm-an kāmil* 'a complete day', *ḏanb-an kabīr* 'a great sin', *taʿārīf-in minnā* 'notifications from us';
- use of genitive particles *mtāʿ, btāʿ*, etc. instead of the annexation structure, e.g. *as-sāniya mtāʿ aṭ-ṭūbanǧiya* 'the farm belonging to the gunners';
- N+ def article +N for the def NP with an attributive adjective: *šahr al-fulānī* 'the such-and-such month', *waṣāyā l-ḥasana* 'the good pieces of advice'.

The pronoun

- personal pronouns follow the dialectal systems, e.g. *iḥna, niḥna* 'we'; *intu* 'you (pl)';
- 'what?', 'why?', and other interrogative pronouns based on forms derived from *š* (< *šayy* < *šayʾ* 'thing') e.g. *ayš, wēš, lēš, lāš*, and similar;
- invariant relative pronoun *illi/ alli*.

The verb and verb-phrase

- *-ī* and *-ū* not *-īn* and *-ūn* and in the 2fsng and 2 and 3pl ending of the p-stem verb;
- loss of gender distinctions in the pl of the verb, and loss of the dual e.g. *-tū* (com pl) not *-tum* (mpl), *-tunna* (fpl), *-tumā* (dual) in the s-stem;
- use of preformatives to mark mood and tense in the p-stem verb, e.g. *b-*;
- *mā* the default negative particle for all verbs, p-stem or s-stem;
- loss of the apophonic passive, with a compensatory increase in the use of the passivizing prefix of pattern VII, *in-*;
- loss of pattern IV of the verb, and its replacement by patterns I or II.
  (iii) in lexis
- dialectal rather than CLA items for very common verbs, e.g. *ǧa* 'to come', *ǧāb* 'to bring', *šāf* 'to see', *rāḥ* 'to go'.

Regional particularities:
Documents may also reflect more localized dialectal forms:

- loss of the interdentals: ت (*t*) for ث (*ṯ*), د (*d*) for ذ (*ḏ*), ض (*ḍ*) for ظ (*ḏ̣*): in documents from Egypt and the Levant, reflecting the urban dialects of those regions;
- *n- . . . / n- . . . u* for 1sng and 1pl p-stem verbs: in Andalusian and Maghrebi texts;
- intrusive *-in(n)-* in active participle + suffixed pronoun forms, e.g. *wāṣlatinnak* 'has come to you': in the Gulf, reflecting the local spoken use of this form;
- *hāḏi* 'this (f)' used to refer to both m and f nouns: in the Gulf, reflecting spoken usage.

(b) substandard, hypercorrect and 'hybrid' forms:
- the CLA msng relative pronoun *allaḏī* used substandardly as an all-purpose relative pronoun, i.e., not the full CLA system of gender and number agreement but a saliently CLA form substituting for the dialectal invariant form *illi/alli*, e.g. with f antecedent noun: *hāḏihi l-marra allaḏī* . . . 'this occasion which . . . '; with grammatically mpl antecedent: *al-qubaysāt allaḏī hum* . . . 'the Qubaysat who are . . . '. These two examples are from the nineteenth century, but the same types of substandard use occur in much earlier Egyptian/Levantine material, e.g. with grammatically f antecedent noun: *fa inna š-šiqāq allaḏī baʿaṯṯ bihā* . . . 'the kerchiefs which you sent . . . ';
- hypercorrect use of CLA case-marker *tanwīn-an* (with orthographic *ʾalif*) on indefinite adjectives and nouns, seemingly as a marker of emphasis, e.g. *ana ḥayrān-an* 'I am confused'; *hāḏi ṯābit-an ʿind al-ǧamīʿ* 'this is well-known to all': early modern Levant and nineteenth-century Gulf; CLA demonstratives *ḏālika* (m) and *tilka* (f) 'that' used interchangeably and substandardly with nouns of either gender: the Levant;
- CLA negative particle *lam* used hypercorrectly with the s-stem verb forms, e.g. *lam atā maʿāhu* 'he did not come with him': Egypt, Levant, Gulf;
- CLA 3rd person *laysa* '(is) not' used hypercorrectly as an invariable negative particle with any person of the p-stem verb, and in nominal sentences, e.g. *laysa rāḍiyīn* 'they are not satisfied', *ǧanābukum laysa tarḍūna* 'Your Honour will not be content', *wa laysa aḥtāǧ* 'and I do not need . . . ': Gulf, Egypt, Levant;
- dialectal verb preformative *b-* combined with saliently CLA p-stem verb morphology to form 'hybrids' which are neither fully dialectal nor CLA, e.g. *b-yuʿād* 'will be sent back': Gulf.

Non-literary early and Middle/Mixed Arabic texts can thus yield valuable evidence for the early occurrence of modern dialectal features. The drawbacks are that the documents are representative of certain locations only, and some aspects of their morphophonology remain hidden by the nature of the Arabic script. And one must always bear in mind that even non-literary materials are not exact mirrors of speech. Judging what the stylistic value of some items is requires local context: the syntax of the relative pronoun *allaḏī*, for instance, which might look like a 'substandard' CLA feature in all the examples listed, may not have had quite the same stylistic value in the nineteenth-century Gulf as it did in medieval Cairo and the pre-modern Levant, since we know that in some of the modern Gulf dialects, variant forms such as *ilaḏī, iladī*, and *illidi*, cognate with CLA *allaḏī*, were still in everyday spoken use as recently as the 1970s alongside the more widespread *illi* (Holes 2016: 93, 387–91). In the nineteenth-century Gulf documents, the written form الذي (*allaḏī*) may simply have been a representation of the spelling of a common dialectal form, not a conscious Classicism, as it probably was in Cairo. A similar issue is the *-īn* and *-ūn* endings on p-stem verbs: these endings in Middle Arabic documents from Egypt and the Levant may be an attempt at a 'high' CLA-like style, since the contemporaneous spoken dialects probably lacked these endings; but the same inflections in Gulf documents would not have had the same stylistic value, as the *-īn* and *-ūn* endings are the normal dialectal forms in that region up to the present day.

1.3 EARLY DIALECTAL ARABIC

As we have already noted, the early and medieval Arab grammarians give us only meagre information about the Arabic dialects of their day, typically in the form of the 'peculiarities' of a certain tribe's dialect (when measured, of course, against their codified CLA). Some of these peculiarities are still current today. An alternative, richer, and more specific source of information about dialect history is from surviving Arabic *Sprachinseln*—Arabic dialects which, for a variety of reasons, were historically cut off from later Arab influence, whether through the immigration of speakers of 'heartland' Arabic dialects or the homogenizing tendencies brought by the spread of literacy in CLA. These *Sprachinseln* have preserved what must, in some cases, be early dialectal forms. We will look briefly at both of these data sources.

1.3.1 OA DIALECTAL FEATURES

The early grammarians conceptualized certain common features of the Arabian tribal dialects of their time as incorrect deviations from CLA. These features were mainly phonological and were given names such as *laxlaxāniyya, ṭumṭumāniyya, kaškaša, taltala, ʿağrada, rutta*, and many others. In some cases these names seem to have been mimetic of the sounds they denote, but in others they are vague, used by different writers to mean different things, and/or are ascribed to different tribes. Many of them seem to mean little more than English lay terms like 'speaking with a lilt', 'drawling', 'gabbling', etc. (Rabin 1951: 10). However, there are a few which can be related to still extant dialectal features:

- *kaškaša*: most probably[23] this was the affrication of /k/ to an alveolar /č/ when the /k/ which forms part of the 2fsng pronoun enclitic *-ki* was in pause and its final vowel was dropped by general rule, e.g. *bayt-V-ki (where 'V' is the case vowel) → *bayt-V-k# → *bayt-V-č# 'your (fsng) house'. The tribal dialects which the grammarians said had this feature were in the north/north-east of Arabia. Sībawaih (d. 793) explained *kaškaša* as the 'addition' of /š/ to /*ki*/, the 2fsng pronoun enclitic, when in utterance-final position, so, e.g. bayt-u-ki 'your (f) house' → bayt-uk# →baytu-kš#. To his way of thinking, the reason for this 'addition' was that speakers needed to preserve the gender distinction with the m form *baytuka* which would otherwise have been lost, since the m form also lost its final short vowel in pause by general rule. Sībawaih's descriptive difficulty was that there is no CLA sound /č/ (= IPA [tʃ])—which is what he was probably actually hearing—so he got around this problem by explaining it as the phonetically similar [k+ʃ], i.e. not as a sound change but as the *suffixation* of [ʃ] (which *is* a CLA sound) to the pausal form of the CLA 2fsng enclitic. Such reductive reasoning was not untypical of him.

[23] See Hopkins 2004 for a more detailed discussion of what the grammarians may have meant by this term.

This /č/ for the 2fsng enclitic, but now generalized to all positions, not just pause, is found today throughout north-eastern and eastern Arabia, southern Iraq, and parts of the UAE, e.g. *bēt-ič* 'your (fsng) house', *abū-č* 'your (fsng) father'. Dialects which have these forms also affricate /k/ in many common lexical items, generally in front-vowel environments, e.g. *čibīr* 'big, old', *hāḏīč* 'this (f)'. Significantly, the alveolar affricate /č/ < /k/ also occurs in some of the Central Asian Arabic dialects and must have been an early 'export' (see *Sprachinseln* §1.3.2, and Holes, Procházka, this volume). Many of the Arabian dialects which affricated etymological /k/ →/č/ in front vowel environments also affricated /g/ (< OA /q/) → /ǧ/, thus forming a pair of voiced–voiceless alveolar affricates.

- *kaskasa*: similar to *kaškaša*, but involving the shift of the /k/ of the 2fsng suffix to a dental affricate /ć/ (= IPA [ts]) in front-vowel environments, so, e.g. *bayt-ić 'your (fsng) house'. The grammarians ascribed this feature to various north Arabian tribes. Today, it is typical of the dialects of Najd, central Saudi Arabia, e.g. *bēt-ić* 'your (fsng) house', and like the alveolar affricate, it also occurs in many common lexical items, e.g. *ćibīr* 'big, old'. Dialects which underwent this change also have a dental affricate /ǵ/ (= IPA [dz]) < /g/ < OA /q/ to form a pair of dental affricates.
- *šinšinna*: this apparently referred to the general Yemeni change *k → š, which produced a dialectal reflex /š/ of the /k/ for the 2fsng enclitic pronoun, e.g. *bayt-iš 'your(fsng) house'. The term *fašfaša* was used to refer to the same phenomenon, and was sometimes called *fašfašat šihr* (Rabin 1951: 50) after the southern Yemeni port of Shiḥr, where it was part of the dialect. Today, this /š/ reflex of the 2fsng pronoun enclitic is a typically south Arabian feature found in Yemen and south-western Saudi Arabia, Oman, parts of the UAE, and in the Baḥārna (Shīʿī) dialects of Bahrain and neighbouring eastern Saudi Arabia, e.g. *bēt-iš* 'your (fsng) house', *abū-š* 'your (fsng) father'[24] (see Holes, Watson, this volume).
- *ʿanʿana*: defined as the substitution of the voiced pharyngeal fricative /ʿ/ for the glottal stop / ʾ/. This feature is well known today in eastern Arabian dialects in a few high-frequency items, e.g. *ʿaǧal*/ *ʿayal* for *ʾaǧal* 'so, then, well', *ʿan* for the complementizer *ʾan*, 'that', *ʿafar* (< *ʾaṯar*) 'it may be that', and lives on in modern borrowings such as *ʿangrēz* 'English (people)', *ʿaskrīm* 'ice-cream', *ʿananās* 'pineapple' (Holes 2016: 55, 63).
- *taltala*: an ancient feature exhibited by certain old tribal dialects of central, eastern, and northern Arabia (Rabin 1951: 61) whereby the prefix vowel of the p-stem (imperfect) verb was *i* rather than *a* as in CLA. This feature is still widespread in the dialects of eastern Arabia today but may well have been

[24] It seems unlikely, given its geographical distribution, that this /š/ was a development of /č/. It is more likely to be a substrate element of the ancestral forms of the Modern South Arabian languages, which also have /š/ or similar forms for the 2fsng enclitic, and which survived in early Yemeni Arabic dialects, from whence it spread east and north via later migrations (see Holes 1991 for a detailed exposition of the history of the various Arabian Peninsula reflexes of /k/ in the 2fsng enclitic pronoun).

'exported' early (and/or have arisen separately) in many areas outside Arabia (see Lentin, this volume).

- *al-qāf at-tamīmiyya*, lit. 'the *qāf* of the tribe of Tamīm': this was most probably a voiced velar stop *[g], which is the normal reflex of *qāf* in all the modern typologically 'bedouin' dialects of northern, central, eastern, and south-eastern Arabia, and many other dialects outside the Arabian Peninsula, all of which were originally 'bedouin' or underwent historical bedouinization. Examples: *gāl* 'he said', *ṣagar* 'falcon', *ḥagg* 'right, entitlement' (see Holes, Procházka, Taine-Cheikh, this volume).

A number of other archaic features of the modern Arabian Peninsula dialects turn up, rather surprisingly, in dialects remote from modern Arabia in both space and time. The features concerned still exist in the modern dialects spoken in the areas of Arabia from which the early migrants to these far-flung places originally came, insofar as this is known (Spain, for example, had an influx of 'Yemenis',[25] Central Asia had one of tribesmen of Iraqi and eastern Arabian origin). The geographically peripheral position of the dialects concerned—Central Asia, West Africa, and, in the past, Andalusia—meant that they remained relatively untouched by later homogenizing influences and retained these unusual features. Some examples:

- the combination of a /*q*/ reflex of OA *qāf* and a /*g*/ reflex of OA *ǧīm*, the norm in both south Yemeni and Omani 'sedentary' dialects, is typical of some varieties of Andalusian Arabic (cf. Corriente 1977; 50–1, 53–4, who specifically notes the Andalusian /*g*/ reflex of *ǧīm* as of 'Yemenite' origin[26]);
- reflexes of the CLA pre-verbal particle *qad*[27] (= *qad, gad, ǧid, čid*) are still found widely in central, eastern, and southern Arabia though with changed functions; the 'parent' form *qad* is frequent in Andalusian Arabic (cf. Corriente 1977: 129);
- *qaṭ/ gaṭ* < *qaṭṭ*, a particle of negative emphasis in CLA, is still widely used in eastern and southern Arabia, e.g. *lā gaṭ min ḥayātna marrēna bēt* 'We never once visited anyone's house' (Holes 2001: 429); in Spanish Arabic its cognate *qaṭ* could be used in the same way to reinforce any negative, e.g. in nominal sentences like *lis qaṭ maʿi šuġal* 'I never have any work' (Corriente 1977: 145);

[25] That is, 'Yemeni' according to medieval Arab historical sources. It seems there was indeed a geographically and linguistically Yemeni element among the incomers, but 'Yemeni' also had a political meaning, 'Yemeni' origin being fraudulently asserted by some who sought to gain an advantage in status by claiming a Yemeni and therefore 'pure Arab' origin when they were not from Yemen at all.

[26] A caveat must be entered here, however. This combination of reflexes of *qāf* and *ǧīm* is rather uncommon in the modern Arabic dialects, but the fact that it occurs today mainly in parts of Yemen and Oman does not prove that these places must have been its origin in Andalusian Arabic. All that this shows is that in all these dialects—Andalusian, Yemeni, Omani—this combination of reflexes is a shared retention, and it may have been much more geographically widespread thirteen centuries ago, when Spain was first colonized by Arabic speakers, than it is now. In and of itself, the presence of these reflexes in Andalusian Arabic proves nothing about their origin (see van Putten 2017 for a rebuttal of Corriente's arguments concerning the claimed 'Yemeni' origin of many features of Andalusian Arabic).

[27] See Holes 2016: 266–72 for data and references to the literature.

- the invariant negative particle *lēs* 'not' (cf. the CLA verb *laysa* 'to not be') survives in southern and eastern Arabia in this fossilized form, e.g. *lēs lēh bidd* 'he has no alternative', *lēs ṣaḥīḥ* 'that isn't true' (Holes 2001: 487 for Bahrain; see also Reinhardt 1894: 282 for Oman) and also turns up in Sinai in the form *lās* and the reduced form *s-*: *lās bidd, s-bidd* 'there is no alternative, no doubt' (Holes and Abu Athera 2009: 238; Bailey 1991: 430).[28] This particle is also found in the invariant forms *lis/ las* in Andalusia (cf. Corriente 1977: 144; 2006: 106). As we have already noted, its CLA cognate, *laysa*, is used in the same way as an invariant negative particle in Middle Arabic texts.
- the 'adnominal linker' *-in/-an* (see §1.2) is common in central, eastern, and southern Arabia, and occurs with a similar form and function in a wide range of typologically 'bedouin' or 'bedouin'-descended dialects in Jordan and the northern Fertile Crescent (Procházka, this volume). It also occurs with an identical form and function in the Central Asian dialects (Holes 2011a: 90, and for references), in eastern Sudan (Reichmuth 1983: 190), in Nigeria (Owens 1993b: 111, 140, 144) and in early Andalusian Arabic (Corriente 1977: 121–2; 2006: 109; Ferrando, this volume). Examples: *bint-in zēna* 'a nice girl' (Gulf), *zamān-an āxar* 'another time' (Andalusia); *lafd-in ʿarabiyye* 'an Arabic dialect' (Khorasan, northern Iran); *bitt-an kabīra* 'a grown-up girl' (eastern Sudan); *bagar-an kubār* 'big cattle' (Nigeria).
- the infix *-in(n)-* inserted between the active participle and a suffixed pronoun when it has verbal force e.g. *bāyginnah* 'he has stolen it', *māxḏatinnah* 'she has taken/married him', a rare feature, typical only of some eastern Arabian, south Yemeni, and virtually all Omani dialects (see Holes 2011a and this volume), is also found in all the dialects of Central Asia (Holes 2011a: 83–4 and for references), eastern Sudan (Reichmuth 1983: 284), and Nigeria (Owens 1993a: 102).
- definite NPs which lack the definite article on the defined noun but have it on the adjective, e.g. *bāb iš-šarği* 'the eastern gate' (Baghdad) are common in eastern and south-eastern Arabian dialects (Holes 2016: 213–15). This feature is also found widely in the 'sedentary' dialects of the Levant (Feghali 1928: 134–6; Procházka, this volume); Iraq (Erwin 1963: 365, 367; Blanc 1964: 126–7), and again in Andalusia (Corriente 1977: 123; 2006: 109).

These features have been selected because of their relative rarity in the modern Arabic dialects as a whole, and their unusual distributional profile. How is one to decide whether such features are instances of polygenesis or monogenesis followed by diffusion? As a general principle, for features widely dispersed but shown by comparative dialectology to be formally and functionally similar and at the same time rare, the single origin explanation is likely to be the correct one (Owens 2006: 161–2). On this view, the features exemplified were probably 'exported' from Arabia (or in the case of the last one, possibly more than one 'eastern' location) at an early date, and then, in some cases, to judge by what is known of the demographic history of where they are now found, re-exported at a later date by further waves of

[28] Holes and Abu Athera 2009: 238 sub the lemma *bdd*.

migration/re-migration. This almost certainly explains the distribution of the rare *-in(n)-* infix construction already noted (see Holes, this volume, for details). The presence of many of the features listed in early Andalusian Arabic is especially significant, as it testifies to their age.

1.3.2 SPRACHINSELN

Another important source of evidence for what the Arabic dialects were like in the past are *Sprachinseln* ('speech islands')—remote communities of Arabic speakers which were cut off and did not receive any fresh infusions of 'linguistic blood' from 'heartland' Arabic-speaking areas after a certain point in their history, and in which literacy in CLA was never widespread. A good example is Malta, where a variant of Siculo-Arabic was spoken until the end of the eleventh century and the Normans' conquest, after which all contact with Arabic-speaking communities ceased. Cyprus also has a 'relic' Arabic dialect, originally spoken in Kormakiti in the north-east of the island, which may originally have been brought there by Maronites fleeing persecution in the ninth century, though a later influx in the twelfth/thirteenth century has been proposed by some, see Borg (2006: 536–7), who comments: 'the sociocultural parallel between Cypriot Arabic and Maltese is particularly close . . . since, in both cases, we are dealing with an Arabic vernacular surviving in complete isolation from the Arabic-speaking world, exposed to interaction with a variety of Indo-European (Italian and English in the case of Maltese, Greek in that of Cypriot Arabic), and spoken by Catholic Christians in a Mediterranean and insular sociocultural habitat'. Other examples of long-term isolation are the scattered communities of Arabic speakers in the vast Oxus valley of Central Asia (in modern N. Iran, Uzbekistan, Afghanistan), conquered by the Arabs in the early eighth century. Here, contact with Arabic speakers external to the region virtually ceased after the tenth century, and the influence of Persian and Turkish became very strong. Nonetheless, these are still today instantly recognizable as Arabic dialects of mixed Iraqi and eastern/south-eastern Arabian origin, such is their similarity to the dialects still spoken in those places today. This fits with what the historian al-Ṭabarī (839–923) tells us about the tribal affiliations of the troops cantoned in Basra in the late seventh/ early eighth century that went on to conquer Transoxania, of which around 30% were from the Azd and ʿAbd al-Qays, two large tribes living at that time in what is now eastern Saudi Arabia, Bahrain, and the UAE/Oman (Holes 2011a: 85–8). A local Omani history written in the eleventh century (Hinds 1984; 1991: 14–15) states that 3,000 Omani soldiers crossed directly to the south coast of Iran from Julanda (modern Rās al-Khayma) in AD 694 without going via Basra.

1.4 SUBSTRATES AND BORROWING

The language(s) which predated the arrival of Arabic have often left vestiges in it, mainly lexical but in some cases phonological and morphological. In the dialects of North Africa, there is a modest Berber substrate, the result of 'imperfect group

learning during a process of language shift' (Thomason and Kaufman 1988: 20ff.; see Aguadé, Taine-Cheikh, this volume) that is mainly lexical, but has a few morphological elements too, as well as a smattering of Vulgar Latin vocabulary from earlier colonists (Aguadé, this volume). At the other end of the Arab World, a substantial example of a substrate is to be found in certain north Yemeni dialects, where fundamental elements of inflectional morphology have been affected, producing inter alia the so-called *k*-perfect verb paradigm, unique in the Arabic-speaking world, and the /š/ reflex of the 2fsng enclitic pronoun referred to in §1.3.1 as *šinšinna*, which may then have been 'exported' from Yemen to the east and south-east of Arabia via later migrations (see Watson, Holes, this volume). The problem here, though, is our limited knowledge of the language which, historical accounts of Yemen's demography suggest, lies behind these apparently substrate elements: Ḥimyaritic (Robin 2007), a Semitic relative of Arabic whose speakers dominated the south-west of the Arabian Peninsula from the first century BC to the sixth century AD. We know from the tenth-century Yemeni historian al-Hamdānī (d. 946) that the remote locations where 'pure Ḥimyaritic' was spoken in his day (see Rabin 1951: 42–53, esp. the map on p.46 extrapolated from al-Hamdānī's description) are precisely the regions of Yemen where the putative substrate Ḥimyaritic elements are found today (cf. Behnstedt 1985: 116, map 68).[29]

In the Gulf region, there is evidence of non-Arabic Semitic vocabulary in certain domains, notably agricultural practices and toponyms, but also, arguably, in morphosyntax (Holes, this volume), of Mesopotamian origin. This is unsurprising given the centuries of political and economic contact and control which the Babylonians and their predecessors exercised over the northern Gulf littoral. But, again, because we know so little about the early language history of this region, it is difficult to be sure whether these (now obsolescent) Gulf Arabic words with cognates in Aramaic/Akkadian are the remains of (a) a substrate; (b) ancient contact-induced borrowing; (c) more recent borrowing (i.e. an 'adstrate') from other Arabic dialects (or Persian) into which the Aramaic/Akkadian items had been borrowed at an earlier period; or (d) a combination of more than one of these processes occurring at different periods of history (see Holes, this volume).

Aramaic also had an influence on the lexis and possibly certain morphological and syntactic structures of the Arabic dialects of the Levant and northern Fertile Crescent (i.e. of northern Iraq, Syria, and southern Turkey (see the chapters of Procházka, Lentin, this volume)). On the other hand, the Egyptian dialects show virtually

[29] The Ḥimyarites left many inscriptions, but the vast majority of them are written in late forms of Sabaic, a much older 'imperial' language of south-west Arabia, to which they considered themselves heirs. Only three inscriptions (all poetry) seem to be in the different idiom, which has been dubbed, because of these differences, 'Ḥimyaritic', but the texts are very difficult to decipher with any certainty. It has recently been argued (Stein 2008) that the differences in these three texts are in fact illusory, and that in all probability the Ḥimyarites spoke a form of late Sabaic as well as using it in their inscriptions. Whatever the truth of this claim, it does not affect the argument advanced here, since late Sabiac/Ḥimyaritic contains exactly the same 'non-Arabic' morphological elements (e.g. the *k*-perfect, *daw* as a negative particle, see Watson, this volume) which occur in no other Arabic dialect and which must therefore be the substrate source of these same elements in the Yemeni Arabic dialects which have them.

nothing which could be described as a Coptic substrate (Behnstedt 2006a; Behnstedt and Woidich, this volume), but rather small-scale borrowing in limited areas, especially rural terminology—measures of size and weight, tools, irrigation terms, soil and field types. This situation is probably a consequence of the early, rapid, and large-scale migration of the Arabs into Egypt.

1.5 DIALECT GEOGRAPHY AND TYPOLOGY

As with any language, major differences between Arabic dialects are associated with geography. Thus it is possible, on the basis of a few basic and more or less coinciding isoglosses, to draw a distinction between the 'Maghrebi' dialects of Mauritania, Morocco, Algeria, Tunisia, and the western part of Libya, on the one hand and the Mashreqi dialects of eastern Libya, Egypt, the Levant, Iraq, Arabia, on the other. On either side of this broad geographical division, many subdivisions and subdivisions of subdivisions could be drawn almost ad infinitum, down to the level of differences between neighbouring villages. But the reality is that differences on individual variables are by nature sometimes neither binary nor sharp, and there is a vast array of many hundreds of individual variables, on each of which the isoglosses marking where the use of one variant begins and another ends do not exactly coincide. If a dialect can be defined as a bundle of variants shared by a community of speakers, this does not alter the fact that boundaries between dialects are almost always fuzzy (as are the boundaries of their associated 'communities of speakers'); indeed, the very concept of distinct dialects may be more a convenient descriptive idealization of linguists than it is an observable reality on-the-ground. That said, it is of course true that if one looks at the dialects of a pair of Arabic-speaking cities remote from each other, such as Casablanca and Cairo, the differences between the two are many and striking, and can reach the point of mutual incomprehensibility in the case of uneducated speakers of their plain colloquials. Nonetheless, there are no 'dialect borders' between the two that correspond to geographical, still less political ones—just a continuum of imperceptible, gradual change from one place to the next as the traveller moves across the thousands of kilometres between the two cities. That much is familiar from research on the dialect geography of virtually any language. In Arabic, however, the picture is complicated by the existence of a pervasive and ancient typological distinction which cross-cuts the geographical one: that between so-called 'bedouin' (*badawī*) and 'sedentary' (*ḥaḍarī*) dialects. There are even subdivisions here, too: in the Levant, the 'sedentary' dialects can be subdivided into distinct 'urban' (*madanī*) and 'rural' (*fallāḥī*) varieties.

Little is known about how, when, and where the 'bedouin' versus 'sedentary' distinctions first arose, but statements are often made to the effect that 'bedouin' dialects are 'more conservative' than 'sedentary' dialects, on the grounds that 'they retain many 'Classical' features lost elsewhere' (Rosenhouse 2006: 259).[30] Such statements are

[30] For a detailed account of so-called 'bedouin' features, see Rosenhouse 1984.

problematic. Part of the problem is that they imply that CLA existed as an invariant linguistic monolith once spoken by a community of native speakers at some unspecified point in the distant past, rather than the outcome of a process of selection and codification by eighth-century schoolmen in Iraq. In this scenario, CLA is seen as the 'parent' of both dialect types which descended from it, whereas the reverse is arguably the case: the old dialects of Arabic predated CLA (or more precisely predated its codification by the grammarians), because old features such as the adnominal linker *-in/an* and the *-in(n)-* infix occur in both dialect types and, circumstantial evidence suggests (Holes, this volume), must have been in spoken Arabic by no later than the end of the seventh century and probably well before this—even if they were not deemed sufficiently 'pure' to make it into the grammarians' codified CLA. Secondly, even if we were to accept the eighth-century grammarians' CLA as the yardstick of conservatism, it is not always true that the 'bedouin' dialects are 'more conservative'. An oft-quoted example is the bedouin 'retention' of the CLA interdental fricatives /*ṯ, ḏ, ḍ*/ in contrast to the 'innovating' 'sedentary' dialects which have stops, /*t, d, ḍ*/. This may be true for large Levantine cities, but it is not elsewhere in that region: many Palestinian village dialects, otherwise 'sedentary' in character, also have the CLA/'bedouin' fricatives. Further afield, all Omani dialects without exception, whether they are otherwise typologically 'bedouin' or 'sedentary', also have the fricative series. In morphology too, the 'bedouin' dialects are said to be more conservative because, for example, they retain, like CLA, gender distinctions in plural forms of the verb, whereas the 'sedentary' ones have lost them. But again, this is an oversimplification: there are typologically 'bedouin' dialects (e.g. of Sunnī Bahrain, of Kuwait, of Muslim Baghdad) which have 'lost' these gender distinctions and many 'sedentary' ones (e.g. of southern Yemen and Oman) which have 'retained' them. It all depends on the choice of which 'bedouin' or 'sedentary' dialects one looks at, and of which dialectal features. Region is also a factor within the 'bedouin' type: the phonotactic feature called the '*gahawa/ghawa* syndrome' (see Glossary) is unquestionably a 'bedouin only' feature,[31] but it is regional, being typical of the eastern Arab World only, its westernmost attestation being in Cyrenaica (de Jong 2007: 151). The 'bedouin' dialects of Tunisia, Algeria, Morocco, and Mauritania do not seem to have it at all. This total absence is perhaps an indication that these Maghrebi 'bedouin' dialects were brought from areas of Arabia (the west?) which never had this feature.

The problem is that the distinction between 'bedouin' and 'sedentary' dialects, though to some degree still valid and useful, is not expressed everywhere by the same set of contrasts (Holes 1996). This is because of the different demographic and social histories of different areas of the Arabic-speaking world, in which the ecological bedouin/sedentary lifestyle divide was not always sharp—indeed, it is quite normal in southern Arabia for one tribe to have sedentary and bedouin sections living in a symbiotic relationship and for both to use exactly the same dialect (the Omani Durūᶜ are an example), a set-up which is not found in northern Arabia. Individual 'bedouin'

[31] Some 'sedentary' dialects of Upper Egypt have it (de Jong 2007: 151), but here it appears to be the result of the long-term 'bedouinization'.

dialects, however, despite the huge distances which separate them, often show extraordinary similarities. To give one personal example: when, in the mid-1970s, my employers transferred me from Kuwait to Algeria, a distance of several thousand miles, I had no difficulty, if I spoke in Gulf Arabic, in making myself understood to (and in understanding) ordinary Algerians in southern oasis towns such as Ourgla and Touggourt, even though most of them had never left Algeria in their lives: we were all speaking 'bedouin' dialects. But the Arabic of the city of Algiers, only a few hundred miles to the north, and where I was based, is of North African 'sedentary' type, and was so incomprehensible to me (as was my Gulf Arabic to the Algérois) that throughout my two-year residence there I found it easier to speak French. This discontinuous 'patchwork quilt' patterning of 'bedouin' and 'sedentary' dialects is now being eroded by dialect contact brought about by the expansion of the cities and new patterns of work and residence. Nonetheless, these original and ancient typological differences remain a feature of the sociolinguistic landscape of the Arabic-speaking world, even if in some locations they have historically undergone social reinterpretation, and continue to do so (see §1.8).

1.6 DIALECT CHANGE, LARGE-SCALE AND SMALL-SCALE

The modern Arabic dialects stand at the end point of thirteen centuries of evolution in the dialects which migrating Arabs, beginning in the mid-seventh century, took with them to Spain, North Africa, Egypt, the Levant, Iraq, Iran, and Central Asia. Since that remote period and in that vast area, migrations, wars, invasions, colonial settlements, and other upheavals have often been the cause of 'macro' dialectal change. Two well-known medieval cases are the expulsion of the trouble-making Banī Hilāl and Banī Sulaym tribes from Egypt to the Maghreb by the Fāṭimid Caliphate in the eleventh century, and the devastation and depopulation of Baghdad wrought by the invasion of the Mongols in February 1258, followed by a second episode of destruction at the hands of Tamerlane in 1400–1. These events had profound effects on the dialect geography of North Africa and Iraq respectively, the results of which we still see today. The Bani Hilāl migration transplanted into the Maghreb the 'bedouin' dialect type, quite different from the 'pre-Hilālī' 'sedentary' dialects brought there and to Spain by the first Arab armies two to three centuries earlier. These bedouin, whose migration continued intermittently over about a century, gradually spread their dialects throughout the North African countryside and also eventually influenced the dialects of neighbouring towns, in a process known as 'bedouinization'. In Iraq, the mass slaughter of the Muslim population of Baghdad (the Christians and Jews were largely spared) left a depopulated city which was gradually refilled over the next five centuries by in-comers, the majority of whom were Muslims and spoke dialects 'bedouin' in character and different from those of the original Muslim population who, like the Baghdadi Christians and Jews, had spoken a 'sedentary' *qəltu*-type dialect, as they still do in the north of Iraq today (Blanc 1964: 168–71). The long-term result for Baghdad was the creation of a new communal indexicalization of dialect-type, Muslim v.

Christian/Jewish; a minor one, Christian v. Jewish, continued the pre-1258 situation. These are just two of the many examples one could cite of 'macro' language change triggered by *force majeure*.

Turning now to less dramatic causes of language change: gradual, accretive urbanization and the contact phenomena associated with it have been a main driver. Among studies of urbanization and its effects, we note the work of Enam Al-Wer on Amman, now a city of several million which was no more than a village as recently as the 1930s, but grew exponentially as a result of inward migration, much of it Palestinian in the aftermath of the wars with Israel of 1948 and 1967. In Amman, a process of levelling has occurred which is resulting in entirely novel forms, such as the 'compromise' focussing of the 2 com pl enclitic pronoun *-kum*, which is now replacing original *-ku* (typical of the Jordanian 'input' dialects) and original *-kon* (typical of Palestinian 'input' dialects) as the default Ammani form in the speech of the younger generations. This *-kum* is an ethnically neutral adopted form (possibly borrowed from MSA) and forms part of an embryonic new linguistic identity for the city (Al-Wer 2007: 70–3). Something similar occurred on a smaller scale in the Iraqi town of Hīt (Khan, this volume): the Arabic dialect of the Karaite Jews of that town developed novel forms, apparently as a result of local bedouin influence. All the Jewish dialects of Iraq have *q* as a reflex of OA*/q/ and *-tu* as the 1sng s-stem form, thus all typically have 'sedentary' *qəltu* 'I said'. But the Karaite Jews of Hīt innovated the form *qilit*, a compromise based on the Iraqi 'bedouin' *gilit* but with 'Karaite' *q* for *g*. Khan attributes the motivation for this innovation to the openness of this particular Karaite community to assimilation into the surrounding culture, as shown, for example, by their use of the Arabic script in many of their writings. Other Iraqi Jewish dialects, spoken by socially less open and inclusive Jewish communities, did not develop such hybrid forms.

1.7 TYPES OF INTERNAL LINGUISTIC CHANGE

The processes of historical change which affected the morphosyntax of Arabic dialects as a whole fall into several categories, some of them already exemplified in the Middle/Mixed Arabic materials that have been discussed:

- the grammaticalization of lexical items to create invariant particles or clitic morphemes to express grammatical categories such as tense, aspect, and modality. These were in many cases transparently derived from common core dialectal verbs such as *kān* 'to be', *ʿamal* 'to work', *qaʿad* 'to sit', *rāḥ* 'to go'. The origin of other particles, however, is more obscure and disputed, e.g. the so-called *b*-prefix on p-stem verbs, which has multiple semantic/pragmatic values and occurs over a vast area (see Owens, Lentin, this volume).
- increased use of clitics and particles in general, e.g. genitive particles which differ from one region to another, e.g. *btāʿ, tabaʿ, mtāʿ, dyāl, māl, ḥagg*, etc. in an 'analytic' noun-phrase structure to express possession and other semantic relationships originally expressed by an annexation structure; loss of the apophonic

passive and its replacement with a system of prefixes/infixes such as *in-* and *it-*, e.g. *inxalaq* 'it/he was created, born' instead of *xuliq/ xiliq, itqaddam* 'it/he was presented' instead of *quddim*.

- simplification and analogical reduction resulting in greater paradigm symmetry, e.g. the loss of the dual and fpl as separate categories of the verb, both replaced by the mpl; reduction in the number of verb conjugations via analogical merger, so, e.g. *garēt* 'I read', *habbēt* 'I loved', *xallēt* 'I allowed', and even (in some Iraqi and Gulf dialects) *gālēt* 'I said', *širbēt* 'I drank', and *tzawwiǧēt* 'I married', whereby several different s-stem verb paradigms were collapsed into one before consonant-initial personal inflections; the reshaping of the p-stem verb conjugation in the Maghrebi dialects so that there is greater symmetry in the expression of person and plurality, e.g. in Morocco: 1sng *n*..., 1pl *n*...*u*, 2msng *t*..., 2fsng *t*...*i*, 2 com pl *t*...*u*, 3msng *y*..., 3fsng *t*..., 3 com pl *y*...*u*.

Processes such as these are typically triggered by long-term contact and dialect mixing, especially in cosmopolitan urban milieus with a history of continuous inward migration (e.g. Cairo, Damascus, Beirut, Casablanca). 'Bedouin'-type dialects, on the other hand, which existed until recent times in relative isolation from external influences, have been less susceptible to them, and slower to develop a more 'analytic' morphosyntax, though this is now changing (Ingham 1994a: 109; Holes 2004a: 123–5).

1.8 INDEXICALIZATION

A widespread sociolinguistic phenomenon in the Arabic-speaking world, touched on already, is the 'indexicalization' of dialectal forms, whereby differences which were originally geographical can acquire a new communal marking, often religious. We look here at a couple of examples.

We have already noted the case of Baghdad, where what was probably historically a minor degree of communal dialect differentiation became a major one because of the knock-on effects of the cataclysm of 1258: a large influx of new inhabitants who spoke a different ('bedouin') type of dialect, and whose speakers happened to be Muslims. An analogous process occurred in Bahrain following the arrival there in the mid-eighteenth century of the Āl Khalīfa and their bedouin tribal allies, all of them of central Arabian origin and confessionally Sunnīs. The already existing major population element in the islands was the Baḥārna, that is, an indigenous Arabic-speaking 'Twelver' Shīʿa community, whose dialects were different from those of central Arabia. After the Āl Khalīfa arrived, the new reality in Bahrain was of two separate communities divided along sectarian lines. From the start, there was no social integration between them: they lived in different settlements, did not intermarry, and pursued different life-styles. Over time, dialect differences—geographical in origin, as was the case in Baghdad—came to be seen as one of the most salient markers of this social fragmentation, and became so labelled: it eventually became possible to speak of the *Sunnī* and *Shīʿī* dialects of Bahrain. By no later than the mid-1970s and almost certainly for many decades before, the 'Sunnī' dialect had become