

THE  
PHONOLOGY  
AND  
MORPHOLOGY  
OF  
ARABIC

---

*Janet C. E. Watson*

OXFORD  
UNIVERSITY PRESS

OXFORD  
UNIVERSITY PRESS

Great Clarendon Street, Oxford ox2 6DP

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 in

Oxford New York

Athens Auckland Bangkok Bogotá Buenos Aires Cape Town  
Chennai Dar es Salaam Delhi Florence Hong Kong Istanbul Karachi  
Kolkata Kuala Lumpur Madrid Melbourne Mexico City Mumbai Nairobi  
Paris São Paulo Shanghai Singapore Taipei Tokyo Toronto Warsaw

with associated companies in Berlin Ibadan

Oxford is a registered trade mark of Oxford University Press  
in the UK and in certain other countries

Published in the United States  
by Oxford University Press Inc., New York

© Janet C. E. Watson 2002

The moral rights of the author have been asserted  
Database right Oxford University Press (maker)

First published 2002

All rights reserved. No part of this publication may be reproduced,  
stored in a retrieval system, or transmitted, in any form or by any means,  
without the prior permission in writing of Oxford University Press,  
or as expressly permitted by law, or under terms agreed with the appropriate  
reprographics rights organizations. Enquiries concerning reproduction  
outside the scope of the above should be sent to the Rights Department,  
Oxford University Press, at the address above

You must not circulate this book in any other binding or cover  
and you must impose the same condition on any acquirer

British Library Cataloguing in Publication Data

Data available

Library of Congress Cataloging in Publication Data

Data applied for

ISBN 0-19-824137-2

1 3 5 7 9 10 8 6 4 2

Typeset in Times

by Peter Kahrel, Lancaster

Printed in Great Britain

on acid-free paper by

Biddles Ltd., Guildford & King's Lynn

# CONTENTS

<i>Preface</i>	xi
<i>Acknowledgements</i>	xii
<i>Abbreviations</i>	xiii
1. INTRODUCTION	1
1.1. The Semitic language family	1
1.1.1. Phonology	1
1.1.2. Morphology	3
1.1.3. Syntax	4
1.2. Arabic within Central Semitic	5
1.2.1. The spread of Arabic	6
1.2.2. The development of Arabic	7
1.2.3. The emergence of a standard language and diglossia	8
1.3. The present study	9
2. THE PHONEME SYSTEM OF ARABIC	13
2.1. Consonants	13
2.1.1. Bilabials	14
2.1.2. Labio-dental	14
2.1.3. Dentals	14
2.1.4. Interdentals	14
2.1.5. Sibilants	15
2.1.6. Palatals	15
2.1.7. Liquids	16
2.1.8. Velars	16
2.1.9. Uvulars	17
2.1.10. Pharyngeals	18
2.1.11. Glottals	18
2.1.12. Glides	19
2.2. The consonantal system of San'ani	19
2.3. The consonantal system of Cairene	20
2.4. Vowels	21
2.4.1. Short vowels	21
2.4.2. Long vowels	22
2.4.3. Diphthongs	22
2.4.4. Cairene long vowels	23

3. PHONOLOGICAL FEATURES	24
3.1. Root features	26
3.2. Stricture features	27
3.3. Laryngeal features	27
3.4. Place/articulator features	28
3.4.1. Primary versus non-primary place	29
3.4.2. [Labial]	31
3.4.3. The [coronal]–front vowel debate	32
3.4.4. [Dorsal]	35
3.4.5. [Guttural]	37
3.4.6. [Coronal]	39
3.4.7. The representation of the pharyngealized coronals	42
3.4.8. The representation of the uvulars	43
3.4.9. The representation of the pharyngeals	44
3.4.10. The phonetic interpretation of non-primary [guttural]	45
3.4.11. The representation of the vowels	47
3.5. Conclusion	48
4. SYLLABLE STRUCTURE AND SYLLABIFICATION	50
4.1. Association and the syllabic skeleton	50
4.2. Syllable structure	56
4.3. Superheavy and ‘super-superheavy’ syllable types	58
4.4. Syllabification	61
4.4.1. Syllable repair processes	64
4.4.2. Syncope	70
4.5. Conclusion	77
5. WORD STRESS	79
5.1. Word stress patterns	79
5.1.1. Word stress patterns in Cairene	79
5.1.2. Word stress patterns in San’ani	81
5.2. The theoretical model	84
5.2.1. The moraic model	86
5.2.2. The iambic/trochaic division	86
5.2.3. Degenerate feet and the minimal word	88
5.2.4. Extrametricality	90
5.2.5. Extrasyllabicity	92
5.3. Word stress in Cairene	93
5.3.1. Exceptions to the stress algorithm	96
5.4. Word stress in San’ani	98
5.4.1. Domain-final CVVC/CVCC	101

5.4.2. Domain-final CVV	110
5.4.3. Suffixed words with pre-antepenultimate CVV or CVG syllables	112
5.4.4. Stress fluctuation	114
5.4.5. Secondary stress	119
5.5. Conclusion	121
6. MORPHOLOGY	122
6.1. The morpheme	124
6.2. Root-and-pattern morphology	125
6.3. Non-concatenative morphology	126
6.4. Prosodic morphology	128
6.4.1. The minimal word	129
6.4.2. Basic stems	130
6.5. Level-one verbal morphology	133
6.5.1. Finite verb stems	137
6.5.2. Forms II, III, and IV	138
6.5.3. Forms VII, VIII, IX, and X	139
6.5.4. Forms V and VI	141
6.5.5. Form I imperfect template	142
6.5.6. Non-triliteral verbs	142
6.5.7. Verbal derivatives	152
6.6. Level-one nominal morphology	164
6.7. Conclusion	173
7. MORPHOLOGY 2	175
7.1. Level-two verbal morphology	176
7.1.1. Affixation of level-two verbal morphemes	176
7.1.2. Affixation of further morphemes	182
7.1.3. Allomorphy	184
7.2. Level-two nominal and adjectival morphology	186
7.2.1. Affixation of level-two nominal morphemes	186
7.2.2. Allomorphy	188
7.2.3. The <i>-i</i> morpheme in Cairene	190
7.2.4. A homophonous morpheme in San'ani	192
7.3. Additional suffixal morphemes in Cairene	193
7.3.1. Turkish suffixes	193
7.3.2. Additional native suffixes	195
7.3.3. Additional suffixal morphemes in San'ani	198
7.3.4. Additional native suffixes	198
7.4. Conclusion	199

8. LEXICAL PHONOLOGY	200
8.1. Prosodic processes	201
8.1.1. Pre-suffix vowel lengthening (in CA)	201
8.1.2. <i>y</i> -strengthening (in CA)	203
8.1.3. <i>n</i> -strengthening	205
8.1.4. *V-V resolution in the inflection of final-weak stems	206
8.1.5. Diphthong reduction and <i>n</i> -strengthening (in SA)	208
8.1.6. Pre- <i>{negative}</i> degemination (in SA)	210
8.1.7. <i>h</i> -disassociation (in SA)	211
8.2. Melodic processes	214
8.2.1. The role of the Obligatory Contour Principle	216
8.2.2. Assimilation of <i>-l</i> of the definite article	216
8.2.3. Assimilation of <i>t-</i> of the detransitivizing prefix	222
8.3. Conclusion	225
9. POST-LEXICAL PHONOLOGY	226
9.1. Prosodic processes	226
9.1.1. Unstressed long vowel shortening (in CA)	226
9.1.2. Resolution of V-V sequences	228
9.1.3. Gemination of clitic-final sonorant (in SA)	234
9.2. Melodic processes	235
9.2.1. Nasal place assimilation	235
9.2.2. Coronal sonorant assimilation	237
9.2.3. Assimilation of adjacent sibilants	240
9.2.4. Coronal place assimilation	242
9.2.5. Voicing assimilation (in CA)	245
9.2.6. Voicing, devoicing, and geminate devoicing (in SA)	248
9.2.7. Intervocalic voicing (in SA)	256
9.2.8. Palatalization	257
9.2.9. Labialization of [labial] and [dorsal] consonants (in SA)	263
9.2.10. Labialization of [dorsal] vowels	265
9.3. Conclusion	266
10. EMPHASIS	268
10.1. The articulatory correlates of emphasis	269
10.2. The acoustic correlates of emphasis	270
10.3. [Guttural] spread	270
10.4. The domain of emphasis spread	273
10.4.1. Emphasis spread from the primary coronal emphatics in Cairene	273
10.4.2. Emphasis spread from the secondary emphatics	275
10.4.3. Analysis	276
10.4.4. Spread from the pharyngeals and the guttural off-glide	277

10.5. Enhancing features and emphasis spread in San'ani	279
10.5.1. [Labial] spread and transparent segments	282
10.5.2. The directionality of [labial] spread	284
10.5.3. The pharyngeals	285
10.6 Conclusion	286
<i>References</i>	287
<i>Index of Authors</i>	299
<i>Index of Subjects</i>	302



## INTRODUCTION

In this chapter I sketch the development of Arabic from its Proto-Semitic ancestor to the present-day dialects. I begin by looking at common features of Semitic phonology, morphology, and syntax. I then consider the position of Arabic within the Semitic phylum as a Central Semitic language which also exhibits several shared traits with South Semitic. In Sections 1.2.1–3, I consider the spread of Arabic from the Arabian Peninsula, the development of the standard language and the phenomenon of diglossia. In Section 1.3 I introduce the main focus of this work, the modern dialects of Cairene and San’ani Arabic.

### 1.1 THE SEMITIC LANGUAGE FAMILY

Arabic is a member of the Semitic language family, which itself is part of the wider Afroasiatic phylum including Ancient Egyptian, Coptic, Cushitic,<sup>1</sup> Berber, and Chadic. Other principal members of the Semitic family are the East Semitic languages of Akkadian and Eblaite (both now long dead), and the West Semitic languages Aramaic, Ugaritic, the Canaanite languages (including Hebrew), ancient and modern South Arabian, and the Semitic languages of Ethiopia (for example, Ge’ez, Tigre, Tigrinya, and Amharic) (Hetzron 1992: 412–13;<sup>2</sup> Faber 1997: 6; cf. Beeston 1970: 11). Common features of Semitic in terms of the phonology, morphology, and syntax are set out in the following sections.

#### 1.1.1 *Phonology*

Semitic languages are marked by a limited vocalic system and a rich consonantal system. There are typically three basic vowels *a*, *i*, *u*, which are attested in both their short and long forms. Semitic languages are also marked by a rich inventory of guttural consonants, which includes both the laryngeals *ʔ*, *h*, the pharyngeals *ʕ*, *ħ* and the uvular fricatives *χ* and *ʁ*. The consonantal phonemes of Semitic languages usually constitute triads of voiceless, voiced and ‘emphatic’ in certain

<sup>1</sup> Hetzron (1992: 413) includes Omotic and Beja (‘if the latter two are separate branches’). According to Zaborski (p.c.), Beja is a Cushitic language and not an independent branch of Afroasiatic.

<sup>2</sup> Hetzron divides Arabic into the following fourteen dialect groups: Balkh, Classical, Cypriot, Eastern Colloquial, Egyptian, Hassaniya, Judeo-Moroccan, Judeo-Tunisian, Modern Standard, Moroccan, Northeastern Colloquial, Shua, Sudanese and Western Colloquial (Hetzron 1992: 416).

sub-sets of the coronal set, and in a few languages, including dialects of Arabic spoken in parts of south-west Yemen, in the dorsal set. ‘Emphatic’ sounds today are pharyngealized in the Central Semitic languages of Arabic and Neo-Aramaic, and glottalized in the South Semitic languages of Modern South Arabian and Ethiopian Semitic (Faber 1997: 8). Descriptions of eighth-century CE Classical Arabic suggest a velarized articulation for the emphatics in this dialect. A glottalized articulation of the emphatics is generally reconstructed for Proto-Semitic (Martinet 1959: 93; Dolgopolsky 1977; Hetzron 1992: 413; Faber 1997: 8). Common or Proto-Semitic appears to have had voiceless, voiced, and emphatic triads in four sub-sets of the coronal set (including a lateral sub-set) and in the dorsal set (Lipinski 1997: 107). The Proto-Semitic voiceless, voiced and emphatic triads are represented in Figure 1.1.

Within the lateral set, the voiceless lateral,  $*\text{ʔ}$ , and the emphatic lateral,  $*\text{ʔ}^h$ , were both realized as lateral fricatives, while the voiced lateral appears to have had a similar articulation to the plain lateral  $l$  attested in Semitic languages today. The Proto-Semitic emphatic lateral fricative,  $*\text{ʔ}^h$ , is the ancestor of the Classical Arabic phoneme known as  $\text{ḍād}$  (Rabin 1951; Moscati *et al.* 1964; Fischer 1997: 189). Descriptions by the Arab grammarians show unambiguously that  $\text{ḍād}$  continued to be articulated as an emphatic lateral fricative well into the eighth century CE (Rabin 1951: 33; Steiner 1977: 57 ff.). Rabin also claims that  $\text{ḍād}$  was articulated laterally by some twentieth-century Qur’anic readers (Rabin 1951: 33).  $\text{ḍād}$  continues to be articulated laterally in dialects of Arabic spoken in parts of the Hadramawt of southern Yemen (Landberg 1901; Habtoor 1989).

The voiceless lateral,  $*\text{ʔ}$ , is classified as one of the three sibilants of Proto-Semitic. It is referred to as  $*s^2$  in descriptions of ancient South Arabian languages (Moscati *et al.* 1964: 33) and there is considerable morphological evidence to show that it is the ancestor of Classical Arabic  $\text{šīn}$  (Moscati *et al.* 1964: 34; Lipinski 1997: 124, 131; Rabin 1951: 209, note 7; cf. Fischer 1997: 189), with the original palatoalveolar sibilant of Proto-Semitic ( $*\text{š}$  or  $*s^1$ ) apparently having coalesced historically with  $*s^3$  to become, over the course of time, the dental sibilant,  $s$  (Lipinski 1997: 124). The reflex of  $*s^2$  in modern South Arabian languages is a lateral fricative (Kogan and Korotayev 1997: 222; Simeone-Senelle 1997: 381–2). By the eighth century CE, the phoneme known as  $\text{šīn}$  had lost its laterality in Arabic,

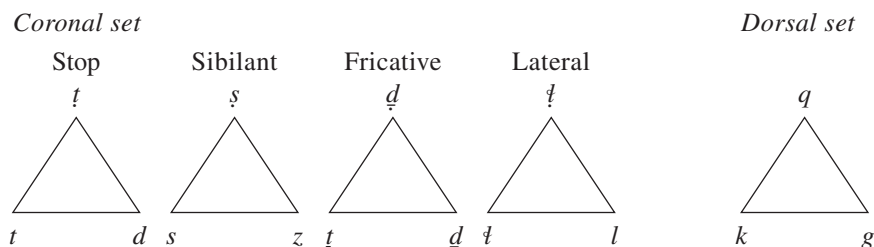


FIGURE 1.1. Proto-Semitic triads

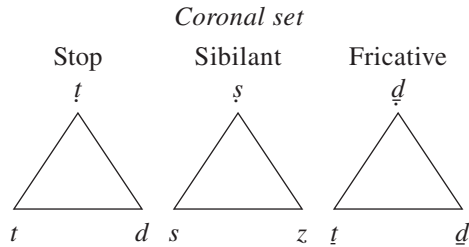


FIGURE 1.2. Eighth-century CE Classical Arabic triads

or at least in Classical Arabic<sup>3</sup> and, from what we can infer from the writings of the eighth-century CE Arab grammarian, Sibawayh, was articulated as a voiceless palatal fricative, with an articulation similar to German /ç/ (Watson 1992: 74; Lipinski 1997: 124, 130). The phoneme *jīm*, which probably had an original velar articulation in Proto-Semitic, moved forward and, according to Sibawayh, was produced between the middle of the tongue and the middle of the hard palate in eighth-century CE Classical Arabic (Sibawayh 1982: 433). This description is interpreted either as a voiced palatal occlusive (Gairdner 1925: 23; Fischer and Jastrow 1980: 105; Watson 1992: 73) or as a voiced palatalized velar occlusive (Schaade 1911: 73; Rabin 1951: 31, 126; Cantineau 1960: 58); I conjecture that Arabic *šīn* and *jīm* at this time constituted a voiceless–voiced palatal near pair, \*ç–\*j.

As a result of changes in the articulation of the non-emphatic lateral fricative, \*l, and the voiced velar stop, \*g, the eighth-century CE Arabic described by Sibawayh exhibited the three voiceless, voiced, emphatic triads in the coronal set shown in Figure 1.2 (cf. Rabin 1951: 209, n. 7).

### 1.1.2 Morphology

One of the main distinguishing features of Semitic languages is their root-and-pattern morphology. The root is a semantic abstraction consisting of two, three, or (less commonly) four consonants from which words are derived through the superimposition of templatic patterns (Holes 1995: 81). In Arabic, the root √KTb has the broad lexical sense of ‘writing’ from which the words for ‘book’ (KiTaaB), ‘written’ (MaKTuuB), ‘writer’ (KaaTiB), ‘office’ (maKTaB) and ‘document’ (KaTi-iBa) are derived. Nouns have feminine and masculine gender, and singular and plural number, and also dual in some Semitic languages. Adjectives are morphologically like nouns. Predicative adjectives agree with the noun subject in

<sup>3</sup> There is evidence that pre-Classical Arabic and other dialects of early Arabic had a lateral fricative. These include some second-century BCE to third-century CE inscriptions from Qaryat al-Faw, near modern Sulayyil, which ‘are written in fine monumental South Arabian script, capable of expressing the phonetic features of Arabic unambiguously’ and which ‘attest the preservation of š (s<sup>1</sup>) and s (s<sup>2</sup>)’ (Lipinski 1997: 73–4). Steiner considers the pair of doublets, *qišda* and *qilda*, in *Lisān al-‘Arab*, evidence ‘that at an earlier period (or in a different dialect) Arabic *šīn* was a fricative-lateral’ (Steiner 1977: 95).

TABLE 1.1 Semitic subject markers

	Prefixed/Suffixed		Suffixed	
	Singular	Plural	Singular	Plural
1.	<i>ʔa-</i>	<i>ni-</i>	<i>-ku</i>	<i>-na</i>
2m.	<i>ta-</i>	<i>ta- . . . -u:</i>	<i>-ta</i>	<i>-tuma:</i>
2f.	<i>ta- . . . -i:</i>	<i>ta- . . . -a:/na</i>	<i>-ti</i>	<i>-tin(n)a</i>
3m.	<i>ya-</i>	<i>ya- . . . -u:</i>	<i>-a</i>	<i>-u:</i>
3f.	<i>ta-</i>	<i>ta- . . . -u:/na</i>	<i>-at</i>	<i>-a:</i>

gender and number.<sup>4</sup> Attributive adjectives agree with the attributed noun in gender, number, case, and definiteness. Semitic languages typically have three sets of pronominal forms: independent subject pronouns, and bound possessive and object pronouns which are suffixed to nouns and verbs, respectively. Verbs have two conjugations for the subject: prefixes and suffixes for the non-past tense (also described as the imperfect aspect), and suffixes only for the past tense (also described as the perfect aspect). The Semitic subject markers are laid out in Table 1.1 (from Hetzron 1992: 414).

In Central Semitic (including most dialects of Arabic), *-k* of the first person singular suffix was replaced by *-t*, while *-t* of the second person suffixes was replaced by *-k* in South Semitic (Faber 1997: 11).

Three other typical morphological Semitic features which are found in Standard Arabic today are the following endings on nouns and verbs (from Holes 1995: 41):

- (1) (a) A set of final short vowel endings suffixed to the noun to indicate case;
- (b) A set of final short vowel endings suffixed to the verb to indicate mood;
- (c) A final nasal ending *-n*, (*tanwīn*), suffixed to the noun to indicate indefiniteness.

These endings have all but disappeared in modern Arabic dialects, though some dialects spoken in the Arabian peninsula, including dialects of the Yemeni Tihama and dialects spoken around Abha in Saudi Arabia, preserve a vestige of *tanwīn* (Greenman 1979; Behnstedt 1985: 60; Al-Azraqi 1998: 71–6).

### 1.1.3 Syntax

Although in Ethiopian Semitic the unmarked word order is S(ubject) O(bject) V(erb) (for example, Tigre; Raz 1997: 455), the original typical<sup>5</sup> word order in Semitic languages was V(erb) S(ubject) O(bject). For modern Arabic, as in

<sup>4</sup> In Standard Arabic, adjectives inflect for singular, dual, and plural number. In recorded modern dialects of Arabic, adjectives inflect for singular and plural number only.

<sup>5</sup> Zaborski (p.c.) points out that word order was not fixed: Proto-Semitic had full nominal inflections and the word order was more or less free with different variants.

Hebrew, it has been argued that the VSO word order is changing towards SVO (Loprieno 1995: 3, for instance); however, in San’ani and certain other, particularly bedouin, dialects of the Peninsula, word order is often dependent on factors such as the dynamism of the verb (dynamic verbs are more likely to occur before a noun subject than stative verbs), the text type (narratives with distinct events are more likely to have verb-initial clauses), and stylistics (Holes 1995: 210–11; Dahlgren 1998; Watson 2000: 11–15). Within phrases, a word which functions as the qualifier typically follows the qualified term. Thus, an adjective follows the noun it qualifies, as in the Standard Arabic noun phrase:

- (2) *al-baytu l-kabīru*  
 the-house the-large  
 ‘the large house’

and an object or complement follows the verb it complements, as in the Standard Arabic verb phrase, as in (3) and (4).

- (3) *kataba kitāban*  
 wrote-he book-indef.  
 ‘he wrote a book’
- (4) *ʔašbaḥa kātiban*  
 became-he writer-indef.  
 ‘he became a writer’

## 1.2 ARABIC WITHIN CENTRAL SEMITIC

Within the Semitic language family, it has traditionally been claimed that Arabic belongs to the South-Semitic or South-West Semitic branch as a sibling of Modern South Arabian and Ethiopian Semitic on the basis of three common factors:

- (5) (a) Almost total preservation of the proto-Semitic sound system with the exception that \**p* lenited to *f*, and \**š* merged with \**s*<sup>3</sup> (Arabic *šm* = proto-Semitic \**s*<sup>2</sup>);  
 (b) the derived *fā<sup>c</sup>ala* and *istaf<sup>c</sup>ala* patterns in the verb;  
 (c) the formation of the plural of nouns by internal vowel changes, as in the following examples from Arabic:<sup>6</sup>

- |     |                    |                  |                  |
|-----|--------------------|------------------|------------------|
| (6) | <i>Singular</i>    | <i>Plural</i>    | <i>Gloss</i>     |
|     | <i>madrasat-un</i> | <i>madāris-u</i> | ‘school/schools’ |
|     | <i>maktab-un</i>   | <i>makātib-u</i> | ‘office/offices’ |
|     | <i>miftāḥ-un</i>   | <i>mafātīḥ-u</i> | ‘key/keys’       |

<sup>6</sup> More recent research, however, has argued that the derived *fā<sup>c</sup>ala* form and the internal plurals go back to Proto-Afroasiatic, and therefore cannot be a feature of South-West Semitic only (Zaborski 1994, 1997).

On the basis of shared morphological innovations, however, Hetzron (1972, 1992) and others (Faber 1997, for instance) have demonstrated that it is more plausible to group Arabic as a sibling of North-West Semitic (including Hebrew, Ugaritic, Deir Alla and Aramaic) within a Central Semitic branch. Faber lists the following features which are peculiar to Central Semitic (Faber 1997: 8–9):

- (7) (a) The realization of the emphatics as pharyngealized rather than glottalized;
- (b) generalization of *-t* in the suffix conjugation verb to give *katabtu* ‘I wrote’ and *katabta* ‘you m.s. wrote’ (cf. 1.1.2);
- (c) a non-geminate prefix conjugation *yaqtul* for the non-past which replaced the inherited *\*yaqattal* non-past;
- (d) development of a compound negative marker *\*bal*;
- (e) within-paradigm generalization of vowels in the prefix conjugation: in Akkadian, the four prefixes which occur in active, non-derived prefix conjugation verbs are *ʔa-*, *ta-*, *ni-* and *yi-*, and Hetzron (1973) suggests that this *\*a-i* alternation reflects the Proto-Semitic state of affairs. In Central Semitic, however, all prefixes for a verb stem have the same vowel—either *a* or *i*—depending on the voice of the verb and, in Hebrew, the phonological shape of the verb stem.

Other features traditionally agreed to be shared by Arabic with North-West Semitic include the formation of the masculine plural suffix *-īn*, the internal passive, a definite article which developed out of the same demonstrative element before language separation, and the *puʿayl* diminutive (Versteegh 1997: 17).

### 1.2.1 *The spread of Arabic*

The original homeland of speakers of Arabic is the central and northern regions of the Arabian Peninsula. The lower half of the Arabian Peninsula was inhabited by speakers of languages known as Epigraphic South Arabian (Hetzron 1992: 412). The end of the sixth century CE, however, saw the rise of the new religion of Islam promoted by the Prophet Muhammad within the Arabian Peninsula in what is now Saudi Arabia. The new Islamic state spread rapidly throughout the Peninsula, and within 100 years had extended north into the Levant, east into Iraq and Khuzistan, and west into North Africa. Over the centuries, the religious frontiers of Islam stretched into Spain, Africa, India, and Indonesia, and across central Asia into Turkestan and China (Gibb 1978: 10). The rise and expansion of Islam was not only a religious and hence cultural conquest, but also a linguistic conquest, and within a few hundred years Arabic became both the official and the vernacular language of all Islamicized countries in the Middle East. Indeed, due to the prevailing tolerance on the part of the Muslims to Christians and Jews, arabicization was more complete a process and progressed at a greater rate than islamicization (Versteegh 1997: 93).

In the course of the spread of Islam, Arabic found itself in contact with a series of foreign languages which it has tended to supplant. In Egypt during the early

centuries of Islamic domination, the Coptic patriarchs communicated with the Arab conquerers through interpreters. By the tenth century CE, the Coptic bishop Severus of Eshmunein complained that most Copts no longer understood either Greek or Coptic, only Arabic. In Upper Egypt, Coptic was limited to a few small pockets in the countryside and to the clergy in monasteries by the fourteenth century CE (Versteegh 1997: 95). It is generally believed that by the sixteenth century CE the use of Coptic was restricted to liturgy in the Coptic church (cf. Loprieno 1995: 7). In North Africa, Arabic became the dominant language of the cities, but Berber managed to resist the spread of Arabic in the rural interior. In Morocco and Algeria, in particular, Berber has retained its vitality alongside Arabic to this day. Likewise in limited areas in the Fertile Crescent, dialects of Syriac have persisted and have influenced neighbouring Arabic dialects.

### 1.2.2 *The development of Arabic*

The Arabic of today is derived principally from the old dialects of Central and North Arabia which were divided by the classical Arab grammarians into three groups: Hijaz, Najd, and the language of the tribes in adjoining areas. Of these, the language of the Hijaz was considered to be the purest, while that of the neighbouring tribes was felt to have been considerably contaminated by other Semitic and non-Semitic languages. It has been estimated recently that Arabic is the native language of about 200 million people (Holes 1995: 1). Arabic is the sole or joint official language in twenty countries in a region stretching from Western Asia to North Africa. These are Morocco, Algeria, Mauritania, Tunisia, Libya, Egypt,



MAP 1. Countries of the Arab world

Sudan, Djibouti, Somalia, Saudi Arabia, Kuwait, Bahrain, Qatar, the United Arab Emirates, Oman, Yemen, Jordan, Syria, Iraq, and Lebanon. It is spoken by Israel's Palestinian population and by Palestinians living in the Occupied West Bank and Gaza. It also has speakers in the south-western corner of Iran, in southern Turkey, in Chad, in some areas in the south of the Sahara, in some enclaves of the Central Asian republics of the old Soviet Union, in francophone West Africa, and among Arab communities in Europe and America.

### 1.2.3 *The emergence of a standard language and diglossia*

The literary Arabic language began to attain a standard form through the development of grammatical norms in the eighth century CE (Fischer 1997: 188). This standard language can be termed Standard Arabic, the terms Classical Arabic and Modern Standard Arabic being used to describe its medieval and modern variants, respectively.<sup>7</sup> Classical Arabic was based primarily on the language of the western Hijazi tribe of Quraysh, with some interference from pre-Islamic poetic koiné and eastern dialects. The language was codified in the Qur'an, the holy book of Islam. Although the lexis and stylistics of Modern Standard Arabic are rather different from those of Classical Arabic, the morphology and syntax have remained basically unchanged over the centuries (Fischer 1997: 188). The vernacular Arabic dialects, by contrast, have developed markedly during this period. Like a number of other languages, therefore, Arabic came to have one standard variety and a large number of regional and social dialects. Unlike many such languages, however, no one in the Arab world is brought up speaking Standard Arabic as their mother tongue:<sup>8</sup> an Arab child's mother tongue will be the regional or social variety of Arabic of its home region, while Standard Arabic, if it is mastered at all, is learnt formally at school or at home as part of the child's education. Standard Arabic is confined to formal written and spoken occasions, and the regional/social variety of Arabic is used at all other times. Standard Arabic now differs considerably from regional and social colloquial varieties of Arabic in terms of its phonology, morphology, syntax, and lexicon. According to Lipinski (1997: 75), such diglossia in Arabic began to emerge at the latest in the sixth century CE when oral poets recited their poetry in a proto-Classical Arabic based on archaic dialects which differed greatly from their own (cf. also Vollers 1906; Wehr 1952; Diem 1973, cited in Fischer 1997: 188).

Dialects of Arabic form a roughly continuous spectrum of variation, with the dialects spoken in the eastern and western extremes of the Arab-speaking world being mutually unintelligible. On the basis of certain linguistic features, Arabic

<sup>7</sup> In this book, the term Standard Arabic is used when referring to the literary language in general; the terms Classical and Modern Standard Arabic are used for specific reference to the ancient or modern varieties of the language, respectively.

<sup>8</sup> The Hijazi dialect has developed markedly since the development of Classical Arabic, and Modern Standard Arabic is quite distinct from the modern dialect of Hijaz (Beeston 1970: 14).

dialects can be divided into two major geographical groups: the first comprises dialects spoken east of a line running from Salum in the north to roughly the Sudan–Chad border in the south; the second comprises the Maghribi dialects spoken to the west of this line. The main phonological features which distinguish the western dialect group from the eastern include the typical reduction of the triangular system of short vowels, *a, i, u*, which is found in eastern dialects, to a two-vowel system (Fischer and Jastrow 1980: 33); and a contrast between an iambic word-stress system in the western group and a trochaic word-stress system in the eastern group. Thus, a word such as *katab* ‘he wrote’ will be typically stressed as *ka'tab* in western dialects, but as *'katab* in eastern dialects.<sup>9</sup> In western dialects, the combination of an iambic stress system together with a tendency to delete unstressed vowels leads to word-initial consonant clusters which are not typically attested in eastern dialects: in the Moroccan Arabic dialect of Lmnabha, *smin* ‘fat’ (Elmedlaoui 1995: 139) is the cognate of Cairene *simīn*; and the word for ‘outside’ is realized as *br̥ra* in Lmnabha (Elmedlaoui 1995: 157), but as *barra* in Cairene.

Dialects of a language which has speakers as ethnically and socially diverse as Arabic, however, cannot be divided in purely geographic terms. Dialects are also commonly distinguished along a bedouin–urban axis: bedouin dialects tend to be more conservative and homogenous, while urban dialects show more evolutive tendencies and usually exhibit fairly clear intra-dialectal variation based on age, gender, social class, and religion. Typical Bedouin features include the voiced reflex of Classical Arabic *qāf*, preservation of the Classical Arabic interdentials, and a gender distinction in the second and third persons plural of the verb, pronouns, and pronoun suffixes (Versteegh 1997: 144). Distinctions between bedouin and urban dialects appear to be less marked in the East, however, particularly in the Peninsula, than they are in North Africa (Fischer and Jastrow 1980: 24).

### 1.3 THE PRESENT STUDY

Most accounts of the phonology and morphology of Arabic are fragmentary, with the information given in unpublished theses, journal articles, and works which address particular aspects of phonology or morphology taking examples from Arabic. In this book, I seek to provide a more comprehensive and integrated account. I focus on two dialects from the eastern group: Cairene, and the dialect spoken within the old city of San'a (the capital of the Republic of Yemen). Where relevant I draw comparisons with Standard Arabic, and other modern varieties of near-eastern Arabic, including Central Sudanese, Palestinian, the Saudi Arabian dialect of Abha and other dialects of Yemeni Arabic.

<sup>9</sup> There are, however, a number of eastern dialects (including that of the Negev Bedouin) and some dialects spoken in Upper Egypt and Oman, in which iambic stress is attested today (Fischer and Jastrow 1980: 59–60).

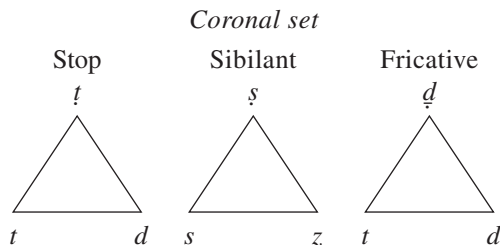


FIGURE 1.3. San'ani triads

San'ani is a dialect of the Arabian Peninsula, an area which has received little attention in generative work on the phonology and morphology of Arabic. It is closer to the descriptions we have of Classical Arabic than is Cairene. It also has considerably fewer speakers (circa 100,000 as opposed to a probable figure of over 12 million speakers of Cairene). Partly as a result of this and partly as a result of its history and tenacious hold on its own traditions, San'ani has experienced a far slower rate of linguistic change than Cairene. In many respects, the dialect is very conservative, exhibiting a number of features of phonology, morphology, and syntax typically considered rural or bedouin (cf. Fischer and Jastrow 1980: 24). The reflex of the Classical Arabic phoneme *qāf*, for example, is a voiced velar /g/, and the reflex of Classical Arabic *jīm* a palatoalveolar affricate /dʒ/; the original form II *fīccāl* and form V *tīfīccāl* verbal noun patterns are more commonly used than the *tafīl* and *tafīccul* patterns found in Modern Standard Arabic; verb-initial as opposed to subject-initial clauses are typically used in narrative texts (cf. Holes 1995: 210); and in possessive constructions direct annexion is often favoured over the use of an 'of' word (*ħagg* in San'ani)<sup>10</sup> (cf. Versteegh 1997: 143). In its consonantal phoneme system, San'ani maintains the triadic opposition attested in eighth-century CE Classical Arabic between voiceless, voiced, and emphatic consonants in three sub-sets of the coronal set—see Figure 1.3.

In contrast to San'ani, Cairene is an innovative, urban dialect. It has maintained the voiceless uvular stop /q/ in religious terminology and other loan words from Standard Arabic, and through the influence of foreign languages has gained seven additional marginal or quasi-phonemes. These are the emphatic /l/ used almost exclusively in the word *allāh* 'God' (cf. Testen 1997: 219–20) and derivatives, as in the majority of Arabic dialects, the emphatics /t/, /b/ and /m/, the voiceless bilabial stop, /p/, and the voiced palatoalveolar fricative, /ʒ/, and labio-dental fricative, /v/.

Through merger, Cairene has lost the Classical Arabic interdental phonemes \*t, \*d, and \*ḏ. This historical loss of the interdental fricatives has led Cairene to

<sup>10</sup> For example, 'my house' translates more commonly in San'ani as *baytī* 'house-my' than as *al-bayt ħaggī* 'the-house-of-me'.

	Stops		Sibilants	
Emphatic	<i>t̤</i>	<i>d̤</i>	<i>ʃ</i>	<i>ʒ</i>
Non-emphatic	<i>t</i>	<i>d</i>	<i>s</i>	<i>z</i>
	voiceless	voiced	voiceless	voiced

FIGURE 1.4

develop a four-way distinction for two sub-sets of the coronal set:<sup>11</sup> voiceless-plain, voiced-plain, voiceless-emphatic, voiced-emphatic, as in Figure 1.4.

In terms of word-stress, Cairene contrasts with San'ani and the majority of other eastern dialects of Arabic which exhibit trochaic word-stress systems (for example, Classical Arabic, Central Sudanese, Palestinian, Saudi Arabian dialects) in its treatment of peripheral feet. In a word comprising more than a single binary metrical foot, such as *madrasa* 'school', stress is assigned to the head of the final, peripheral foot in Cairene to give *mad'rasa* 'school'. In San'ani, the final foot is deemed extrametrical and is therefore not taken into account in word-stress assignment. Thus, stress in San'ani is assigned to the rightmost non-extrametrical foot to give *'madrasah* 'school'.

In terms of its morphology, Cairene contrasts with Proto-Semitic (3), Classical and Modern Standard Arabic, San'ani, and a number of other bedouin-type Peninsula dialects in making no gender distinction in the second and third person plural independent and bound pronouns. Thus, while San'ani has the pronouns *antayn* 'you f.pl.' and *antū* 'you m.pl.', *hum* 'they m.' and *hin* 'they f.', Cairene simply has *intu(m)* 'you pl.' and *humma* 'they' to refer to both genders. Cairene has an enriched concatenative nominal morphology due in large measure to extended contact with other languages and cultures. In addition, much of its non-concatenative morphology has been simplified, particularly in the derivation of verbal participles of derived verbs: where the active participle was once distinguished from the passive participle by the final stem vowel (*i* for active, *a* for passive), today the active participle of most derived verbs is distinguished from the passive participle by syntactic or pragmatic context alone. However, the loss of non-concatenative morphology in one part of the morphology is occasionally balanced by the development of non-concatenative morphology in another. In common with many eastern urban dialects of Arabic, Cairene has long since lost the apophonic (or internal) passive (Retsö 1983): thus, 'a letter was written' is *kutibat risālatun* in Standard Arabic with the /u-i/ vocalism of the verb indicating perfect passive, while in Cairene it is *itkatab gawāb* with the passive indicated by affixation of a passive prefix to the

<sup>11</sup> In Cairene, *qāf* lost its dorsal articulation historically to become a glottal stop, /ʔ/; thus, although the Cairene reflex of *jīm* is a voiced velar plosive, /g/, as in Proto-Semitic (see Table 1.1), and the voiceless velar stop, \*k, has been preserved as such, the Proto-Semitic dorsal triad \*/g, k, q/ was also lost in Cairene.

basic verb. By contrast, a number of simple intransitive verbs are distinguished from their transitive counterparts, not by a different verbal form, as in many other eastern dialects of Arabic, but by a palatal (/i/) versus a guttural (/a/) vocalism (cf. Willmore 1905: 120, 121).<sup>12</sup> Consider the paired examples in (8).

- (8) *bi<sup>c</sup>id* ‘to be/become distant’      *ba<sup>c</sup>ad* ‘to take away, remove’  
*i<sup>c</sup>ib* ‘to be/become tired’      *ta<sup>c</sup>ab* ‘to tire, wear out’  
*ḥimi* ‘to be/become hot’      *ḥama* ‘to heat’

The present work, therefore, is a study of the phonology and morphology of one progressive and one conservative near-eastern dialect of Arabic. The two dialects have a number of similarities which enable us to classify them on one level as a group. As we have seen, these include a trochaic stress system, and the presence of a short vowel triad, /a, i, u/. Other similarities include the same basic syllable inventory, and a simplex vocalic melody in basic form I verb stems. Differences between the two dialects include, as discussed, the consonantal phoneme inventories, the treatment of peripheral feet in the stress system, and the presence or absence of gender distinction in second and third person plural independent and bound pronouns. Other differences include the number of long vowels (five in Cairene, three in San’ani), the identity of the default vowel (/a/ in San’ani, /i/ in Cairene), and the tolerance or intolerance of the dialect to initial consonant clusters.

<sup>12</sup> In many cases at least, the /a/ type verb represents a restructured form IV.