Clive Holes

MODERN
Arabic

Structures,
Functions, and
Varieties

Georgetown Classics in Arabic Language and Linguistics
Karin C. Ryding and Margaret Nydell, series editors

With global awareness now refocused on the Arab world, and with renewed interest in Arab culture, society, and political life, it is essential to provide easy access to classic reference materials such as dictionaries and reference grammars, and language teaching materials. The key components of this series of classic reprints have been chosen for quality of research and scholarship, and have been updated with new bibliographies and introductions to provide readers with resources for further study. Where possible, the original authors have been involved in the reproduction and republication process.

Georgetown University Press hopes hereby to serve the growing national and international need for reference works on Arabic language and culture, as well as provide access to quality textbooks and audiovisual resources for teaching Arabic language in its written and spoken forms.

Books in the series:
Arabic Language Handbook
        Mary Catherine Bateson
Audio CDs for A Basic Course in Moroccan Arabic
        Richard S. Harrell
Audio CD for A Reference Grammar of Syrian Arabic
        Mark W. Cowell
A Basic Course in Iraqi Arabic with MP3 Audio Files
        Wallace M. Erwin
A Basic Course in Moroccan Arabic
        Richard S. Harrell with Mohammed Abu-Talib and William S. Carroll
A Dictionary of Iraqi Arabic: English–Arabic
        B. E. Clarity, Karl Stowasser, and Ronald G. Wolfe, editors
A Dictionary of Iraqi Arabic: Arabic–English
        D. R. Woodhead and Wayne Beene, editors
A Dictionary of Moroccan Arabic: Moroccan–English / English–Moroccan
        Richard S. Harrell and Harvey Sobelman, editors
A Dictionary of Syrian Arabic: English–Arabic
        Karl Stowasser and Moukhtar Ani
Formal Spoken Arabic FAST Course with MP3 Files
        Karin C. Ryding and Abdelnour Zaiback
Modern Arabic: Structures, Functions, and Varieties, Revised Edition
        Clive Holes
A Short Reference Grammar of Iraqi Arabic
        Wallace M. Erwin
A Short Reference Grammar of Moroccan Arabic with Audio CD
        Richard S. Harrell

Modern Arabic Structures, Functions, and Varieties

REVISED EDITION

Clive Holes

Georgetown University Press / Washington, D.C.

Contents

Figures and Tables

Foreword to the Georgetown Classics Edition

As I write these words in 2004, a series of strategic decisions initiated in Washington, D.C., have had the effect of turning Samuel Huntington's thesis of a "clash of civilizations" from debatable theory into an increasingly disturbing global reality. In such a context (and however we may wish to interpret its causes and ramifications), one thing is abundantly clear: there is an urgent need for us to acquire a great deal more knowledge than is available about the cultures of the non-Western world in order to have a better grasp of those cultures' motivations to behave and to react in the ways that they do and have done. The republication of Clive Holes's already well-known study *Modern Arabic* is, thus, to be welcomed. It contributes to that learning process by providing an excellent introduction to the language that is not only the canonical linguistic vehicle of the Islamic faith in both the oral and textual dimensions, but also the common means of communication within the Arab world and, indeed, beyond it. Arabic is the most prevalent language in that region of the world that *we* have dubbed the "Middle East" (why else, one might ask, is it "middle"?). Using a more indigenous yardstick, it is the predominant language in that area of the globe that the late president of Egypt, Jamal `Abd al-Nasir (Nasser), described as stretching "from the [Atlantic] Ocean to the [Arabian/Persian] Gulf." At the outset of the twenty-first century, the strategic significance of this region—the impact of its oil reserves on the global economy, its role as the physical center of one of the world's largest religious communities, among others—make the Arab world and Arabic langugae enormously important.

The great strength of Holes's study is that it insists on studying its topic within a thoroughly modern linguistic frame of reference. It is important at this point to note that the Arabic language itself has traditionally "stacked the deck" when it comes to describing itself. The term for today's standard form of the written language, historically founded on the inimitable model of the Qur'anic text, remains what it has been for centuries: "*fushā*" (or, in full, "*al-lughah al-fushā*"), the *most* eloquent, culturally acceptable form of the language (the emphasis is mine, drawing attention to the use of the superlative form), whereas the colloquial dialects are generally referred to as " `*ammiyyah*" (implying "ple-

beian") or "*darijah*" (implying "current"). This bifurcation of the language into two levels and the heavy overlay of cultural value that has traditionally been added to the former as a natural consequence is, in some ways, enshrined in the concept of "diglossia," a term originally coined by Charles Ferguson in a famous study of this particular category of languages.

Holes follows the general outline of this approach to the language but the importance of his work lies in the fact that he refuses to be bound by many of the norms and expectations that have characterized so many previous studies of it. Indeed, in chapter 9 he examines some of the flaws in the basic model that have emerged as the result of recent linguistic research. As he leads his readers through surveys of the history, phonology, morphology, syntax, and lexicography of the Arabic language, Holes insists on presenting the linguistic situation as it is: the standard written patterns are discussed alongside those of the dialects. In the case of the latter, it is also important to point out that the examples used to illustrate the various features of the language are culled from across the breadth of the Arab world. Thus, we find references to the language situation in both the Maghrib (northwest Africa) and the Gulf alongside those of the more often cited "central regions" (such as Egypt and Syria).

This eminently scholarly but highly readable account of the Arabic language as it is rather than as cultural and religious norms might have liked to believe it is presents its readers, be they within academe or part of a broader general reading public, with a clear and well-illustrated picture of one of the world's most important language systems. Its republication is indeed a welcome event.

Roger Allen
University of Pennsylvania

Preface to the Second Edition

The print run of the first edition of *Modern Arabic: Structures, Functions, and Varieties* was exhausted around 1999. The original publishers declined on economic grounds to reprint it. However, demand for the book has not ceased, and I am grateful to Georgetown University Press for approaching me with a view to publishing a reprint, and then for agreeing, at my suggestion, to a revised edition.

The aim of this second edition remains the same as the aim of the first: to present a balanced, dispassionate, and accurate picture of the structures, functions, and varieties of the contemporary Arabic language and in so doing, to grind no theoretical axe or espouse any particular position. To use the contemporary jargon, my approach is data-driven rather than theory-driven, but certainly not, I hope, theory-free.

This second edition has been updated to take account of research published in several areas of Arabic linguistics since the first edition came out in 1995, but I have not felt it necessary to make any substantial alterations. Some small errors of fact and a few typographical errors that crept into the first edition have also been corrected. I am indebted to a number of academic reviewers for their comments and criticisms, in particular to Simon Hopkins.

Oxford
2004

Preface to the First Edition

It is now more than twenty years since the publication of A. F. L. Beeston's *The Arabic Language Today*. That slim but masterly volume, a model of concise and accurate description, is long since out of print. Although Beeston's book has been a constant source of inspiration in the writing of this one, I have made no attempt to imitate it in style or in content. For one thing, this book's scope, unlike Beeston's, is not limited to Modern Standard Arabic (MSA), but also describes in outline the structure of the modern Arabic dialects in their multifarious varieties, geographical and social. Furthermore, I have been as much concerned with explaining as with describing the synchronic and, to a lesser extent, diachronic relationship between the written language and the spoken dialects, and cross-dialectal influences. To achieve these ends, I have made heavy use of the wealth of descriptive material and more theoretically oriented studies of both MSA and the dialects that have appeared since Beeston wrote his book. Despite this welcome upsurge in Arabic linguistic studies, much of the work done on Arabic, it seems to me, has suffered, and continues to suffer, from a reluctance on the part of native and western linguists alike (though for different reasons) to recognize the contemporary linguistic realities of the Arab world and take these as the starting point for their descriptive and analytical researches. This book is a modest attempt to redress the balance and view the language as an integrated whole—which is, after all, how its native speakers use, experience, and think about it.

Cambridge
1994

Transliteration Conventions, Gloss Lines, and Abbreviations

Transliteration conventions

(a) Phoneme, word, and sentence examples

All illustrative Arabic examples, standard or dialectal, from single phonemes to whole sentences, are printed in italics, as are the occasional examples of Arabic linguistic and other technical terminology.

In transliterating the consonantal inventory of Modern Standard Arabic (MSA), I have adopted a slightly modified form of the International Phonetic Alphabet (IPA), which is laid out in table 2.1. I have followed the same system in transliterating dialectal examples, including those cited from sources that use different systems, which I have in most cases retransliterated to conform to my own. This may occasionally have resulted in a broader transcription than in the source text, but not in any matter relevant to the point that the example is cited to illustrate.

Unless otherwise indicated, the quality of the vowels is similar to that of IPA convention. Where the dialects contain sounds that are never used in MSA (e.g., *č*), the symbol again reflects its IPA value (here a voiceless alveolar affricate).

In the chapters on morphology in which the root-pattern system is described (and elsewhere also when appropriate), capital letters are used to represent root consonants, and small letters to represent nonroot consonantal affixes and vowels. The paradigm triliteral root is represented as $C_1C_2C_3$ (or simply CCC) and v = unspecified short vowel. Doubled consonants are written double and long vowels marked by a colon. Thus *KTB* is a root (a purely theoretical concept) that has the broad lexical value 'write'. From it may be derived such forms as the verbal suffix-stem *KvTvB-*, which may be realized actively as *KaTaB-* 'wrote' or passively *KuTiB-* 'was written'; to these must be added personal inflections in order for an actual verb form (e.g., *katabtu* 'I wrote') to be formed.

(b) Names

When not cited as examples, well-known Arabic names are spelled as they usually appear in nontechnical English transliteration and are printed in Roman

(e.g., Baghdad, Cairo, the Hejaz, Naguib Mahfouz). The names of historical figures that have no standard transcription in English are transliterated according to orientalist convention (e.g., 'Umar ibn al-Khaṭṭāb [rather than *'umar ibn al-xaṭṭa:b*, as the name would be in the transliteration system used for illustrative examples]). The titles of Arabic books in the notes to each chapter and in the list of works consulted are similarly given in the orientalist transliteration system, because this is the convention in the literature.

Gloss lines

Where it is necessary to indicate the morphosyntactic structure of a word or sentence, a gloss line is provided in which the glosses are placed directly under the transliterated Arabic words to which they correspond. The third line, where three are supplied, is an idiomatic English translation. For example:

> *sa?usa:firu ?ila: lqa:hira bukra*
> FUT-1sg-travel to the-Cairo tomorrow
> 'I shall be leaving for Cairo tomorrow'

In this MSA example, the gloss of the first word *sa?usa:firu* is to be read as "*sa* (future particle) —*?u* (1st person singular) —*sa:firu* (verb "travel")." In written MSA, this word is a single grapheme; the gloss reflects this by hyphenating the constituents. The verb in this case is prefix-stem, requiring a prefixed inflection, therefore the placing of "1sg" before the verb below *?u*; if the verb is suffix-stem, the gloss of the personal inflection, here *tu*, is placed after it, as in:

> *sa:fartu ?ila: lqa:hira ssana lma:ḍiya*
> traveled-1sg to the-Cairo the-year the-past
> 'I traveled to Cairo last year'

I emphasize that, for the sake of both descriptive accuracy and presentational clarity, I have adopted a realistic, and at the same time minimalist, approach in transliterating and glossing. The principle followed is that the amount of morphological detail given in the Arabic examples and the gloss lines is sufficient just for the purposes for which the example is cited. So, in the first example above, the verb form -*sa:firu* is not further broken down into stem (*sa:fir*) and indicative mood marker (-*u*), and would be so only if the point were to illustrate that particular aspect of morphological structure.

So, for example, the Arabic for "Greater Cairo," which in all varieties and styles of Arabic is grammatically a definite noun phrase, feminine in gender, would appear in transliteration in several different forms, depending on the amount of phonological and morphological detail relevant to the exposition at any given point. As a phrase of Modern Standard Arabic, it would appear as:

(a) with minimal detail (usually in running text) as a lexical citation form, with only word boundaries distinguished:

> *al-qa:hira l-kubra:* ('Greater Cairo')

(b) in more detail (with or without the word boundaries marked), reflecting the phonological facts of performance in carefully pronounced MSA:

> *?al-qa:hiratu l-kubra:*
> the-Cairo the Greater
> 'Greater Cairo'

(c) in maximal detail, where the detail of its morphological structure is being highlighted:

> *?al-qa:hir-at-u l-kubra:*
> the-Cairo-f-NOM the-greater-f
> 'Greater Cairo'

The same phrase in dialectal Arabic would be transliterated:

> *il-qa:hira l-kubra:*
> the-Cairo the-Greater
> 'Greater Cairo'

I have followed a similar principle, in order to avoid extraneous detail, in illustrating larger scale aspects of syntactic structure such as clause coordination and subordination. In many cases (especially in chapter 7), I do not provide a transliteration, let alone a gloss line, of the Arabic example sentences: a close English translation is provided with only the relevant element (clause complementizer, conjunction, etc.) written in Arabic in order to illustrate (e.g.) its position and function relative to other elements.

Abbreviations

A few abbreviations (e.g., a.part., p.part.) appear in both upper- and lower-case: lowercase in running text, and uppercase in the gloss lines of example sentences. They are given here in lowercase.

a.part.	active participle
acc.	accusative case
adj.	adjective
adv.	adverb
Amer. Eng.	American English
Ar.	Arabic
asp.	aspect
aux.	auxiliary verb
C	consonant
CLA	Classical Arabic
coll.	collective
com.	common plural
COMP	complement
d.	dual
def.	definite
dem.	demonstrative
det.	determiner
DO	direct object
enc.	enclitic
Eng.	English
EXTRA	extraposed element
f.	feminine
Fr.	French
FUT	future marker
gen.	genitive case
imp.	imperative
indic.	indicative mood
interrog.	interrogative
intrans.	intransitive
IO	indirect object
It.	Italian
juss.	jussive mood
lit.	literally
loc.	locative

m.	masculine
MSA	Modern Standard Arabic
n.	noun
neg.	negative
nom.	nominative case
NP	noun phrase
O	object
OA	Old Arabic (the putative ancestor of the modern colloquials in matters of pronunciation)
PASS	passive marker
Pers.	Persian
pl.	plural
p.part	passive participle
Port.	Portuguese
poss.	possessive
prep.	preposition
pron.	pronoun
prov.	proverb
p-stem	prefix-stem
PTCLE	particle
Q	question particle
REL	relational adjective marker
S	subject
Sen	sentence
sg.	singular
Sp.	Spanish
s-stem	suffix-stem
subj.	subjunctive mood
temp.	temporal
TOP	topicalizing particle
trans.	transitive
Turk.	Turkish
V	verb
v	vowel
var.	variant
VP	verb phrase

Introduction
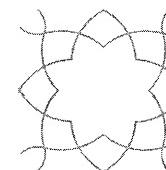
0.1 Where is Arabic spoken?

Arabic is the sole or joint official language in twenty independent Middle Eastern and African states: Morocco, Algeria, Mauritania, Tunisia, Libya, Egypt, Sudan, Djibouti, Somalia, Saudi Arabia, Kuwait, Bahrain, Qatar, the United Arab Emirates, Oman, Yemen, Jordan, Syria, Iraq, and Lebanon. It is the native language of Israel's Arab citizens and of the Palestinians who live in the occupied West Bank and Gaza. Since the end of the nineteenth century, there have been large communities of Arabic speakers outside the Middle East, particularly in the United States, and more recently in Europe.[1] Arabic is also the language of Islam's holy book, the Koran, and as such is the religious and liturgical language of all Muslims, regardless of origin. On the contemporary international stage, Arabic has been an official language of the United Nations alongside English, French, Spanish, Russian, and Chinese since 1 January 1971.

Recent estimates put the total number of native speakers of Arabic at about 250 million. In some of the countries listed above, however, Arabic is by no means the only, or even in some cases the first, language of some sections of the population. In the countries of North Africa—mainly Morocco and Algeria, and to a lesser extent Tunisia and Libya—there are scattered but large minorities, several million strong, whose mother tongue is one of a large

number of Berber dialects that are only distantly related to Arabic. Virtually all these Berber speakers, however, have at least a rudimentary knowledge of local spoken Arabic, and most are bilingual. As a result of more than a century of French colonization, some are trilingual in Arabic, Berber, and French. Further east, in the mountains of northern Iraq, there are several hundred thousand native speakers of Kurdish, an Indo-European language related to Persian. In southern Sudan, the southernmost country whose official language is Arabic, the indigenous population speaks a variety of mainly Bantu languages, quite unrelated to Arabic. In the "heartland" Arab areas of the Arabian Peninsula and the Near East, small pockets of speakers of other Semitic languages still exist here and there within the borders of the Arabic-speaking countries, although they have now all but disappeared: Aramaic in a few villages in Syria and northern Iraq, and a group of South Arabian languages (Mehrī, Ḥarsūsī, Jibālī) in the deserts and mountains of Oman. These linguistic minorities, large and small, are a vestige of the situation that existed in the Middle East and North Africa before the great expansion in the influence of Arabic, which began with the rise of Islam in the middle of the seventh century A.D.

In adjacent areas once but no longer under Arab hegemony, a linguistic flotsam was left by the receding imperial tide. Here, Arabic is still spoken as the first language by some of the population, although it has no official status. In Asia, the southern Iranian province of Khuzistan (or Arabistan) is demographically, and politically perhaps, the most important of these regions, but there are also Arabic-speaking minorities in Afghanistan (Balkh), parts of the former Soviet Union (Uzbekistan), northeastern Iran, and quite widely in southern Turkey. In Africa, there are mother-tongue speakers of Arabic on the fringes of the southern Sahara in northern Nigeria, Niger, Mali, and Chad. In the Mediterranean, a recognizably Arabic-based vernacular was still spoken until 1974 alongside Greek in the village of Kormakiti in northern Cyprus, and Maltese is undoubtedly structurally an Arabic dialect, though it has been heavily influenced by centuries of contact with Romance languages and, more recently, English.

But what exactly do we mean when we say that the inhabitants of such geographically separated, ethnically diverse areas "speak Arabic"? In what sense does an Arabic-speaking villager from Uzbekistan speak the "same" language as a northern Nigerian townsman from Maiduguri, a Waheibi tribesman from the sands of southern Oman, or a professor of Islamic law at the al-Azhar University in Cairo? An objective comparison of the varieties of Arabic found at the edges of the Arabic-speaking area might well lead one to the conclusion that they were historically related, but synchronically distinct and mutually unintelligible languages like English and Dutch.[2] But it is unlikely that this view would

be shared by ordinary speakers of these different varieties of Arabic themselves. Each would certainly aver that he or she personally "spoke Arabic" and would probably agree that the others did so too even though their normal vernacular might be very difficult to understand. How can we explain this apparent paradox?

It is well known that speakers' perceptions of closeness or distance between languages and dialects sometimes have as much to do with attitudinal, historical, and even political factors as with objective linguistic differences or similarities. The standard varieties of Hindi and Urdu, for example, although written in different scripts and differing from each other in literary writing in the quantity of Sanskrit-derived forms employed in the former and of Persian-derived ones in the latter, differ objectively only slightly from each other in the grammar and vocabulary of their nonliterary spoken forms. A foreigner who learns to speak the standard variety of either one of them will get virtually 100 percent comprehension from speakers of the standard variety of the other. Nonetheless, it is sometimes claimed by native speakers of these languages, despite their obvious structural similarity, that they are different to the point of mutual unintelligibility. In the Arabic-speaking world, we have the opposite phenomenon: considerable objective linguistic diversity[3] that can on occasion definitely hamper effective communication, allied to the homespun view that *kulluh ʿarabi:y*—"It's all Arabic." As with the Hindi and Urdu speakers who claim they cannot understand each other when observation suggests they can, this is more of a political and cultural than a linguistic statement.

0.2 Varieties of Arabic

The spoken Arabic dialects are the varieties of the language that all native speakers learn as their mother tongue before they begin formal education. Geographically, these dialects may be thought of as being distributed along innumerable sets of intersecting continua, from Morocco in the west to Oman in the east, and from the borders of southern Turkey in the north to Sudan in the south.[4] Within this vast area, the inhabitants of any given village or town will experience no difficulty in understanding the ordinary vernacular speech of the inhabitants of the next village or town in any direction. The greater the distance between any two points of comparison, by and large, the greater will be the differences between the ordinary vernaculars spoken in them. It is not then surprising to find that the varieties of Arabic spoken at the extreme peripheries of the area differ from each other considerably, and certainly to the point of mutual unintelligibility if we were to compare what might be called the plain uneducated vernaculars—say, that of an Omani nomad with that of a Moroccan

townsman from Marrakesh.[5] In practice, in the modern world, a number of factors work to reduce the effect of such geographical differences.

For decades, although less so today, expatriate teachers, technicians, and professionals of all kinds from Egypt and the Levant formed the backbone of education and technical services in less developed areas within the region, such as the Gulf states. Conversely, students from all over the Arab world, but especially from those countries that until recently lacked a system of tertiary education, have long studied in the universities of Egypt, Syria, and Iraq. Mecca, as the epicenter of the Islamic world, has for thirteen centuries been the goal of Muslim pilgrims from all over the Arabic-speaking lands. So interdialectal contact between different Arab populations is not new. But with the massive increase in inter-Arab contact occasioned by recent economic developments in particular, the scale and variety of types of interdialectal contact has become much greater. Nowadays, dialect contact occurs not just at the level of the well-educated business executive, flitting from country to country, or the expatriate teacher or technician on long-term contract, or the scholarship student. In Saudi Arabia, Iraq, and the modern city states of the Gulf that have developed in the last two decades, armies of semiliterate economic migrants—Egyptian, Sudanese, and Yemeni laborers, porters, doormen, and waiters, mostly from rural backgrounds and with modest educational attainments—have become a semi-permanent, highly visible feature of the landscape. Perhaps it is still unlikely that our Omani nomad and Marrakeshi townsman would bump into each other in a Kuwaiti supermarket and still less need to discuss the spiraling price of rice. But this is no longer because the opportunity could never arise. Large numbers of very ordinary Arabs, expatriate and host, are indeed nowadays faced with having to cope with the speech of others from very different geographical and educational backgrounds. How do they do this? As we shall see, they have a number of linguistic resources and "coping" strategies at their disposal.

From an early age, Muslim Arabs (that is, more than 90 percent of all Arabs) have some degree of exposure to the language of the Islamic scriptures: that is, so-called Classical Arabic (CLA). As the language of revealed scripture—in Muslim eyes the literal words of God—the Classical Arabic of the Koran is viewed as an immutable linguistic phenomenon fixed for all time. For most Arab children, it is the first kind of Arabic different from their mother tongue dialect to which they are exposed, and it leaves an indelible impression, reinforced throughout later life in the constant ritual of prayers and attendance at the mosque. This early exposure consists of the rote learning of verses or even whole chapters of the Koran and the rituals of communal prayer, and is traditionally begun at the age of five or six in special Koranic schools. Until recently, this was the only kind of education that many Muslims got; and while Classical Arabic is no more a functional linguistic idiom for Arabs than liturgical Latin is a living language for Roman Catholics, its rhythms and cadences are part of all Muslim Arab children's earliest conscious experience of language. Classical Arabic is revered by rich, poor, educated, and illiterate alike as the linguistic jewel in the Islamic cultural patrimony. It is regarded as the inimitable apogee of perfection, unsurpassable in beauty, an ethereal ideal of eloquence, perfect symmetry, and succinctness—however imperfectly, in practice, many Arabs understand it.[6] And although the syntax, vocabulary, and phraseology of Arabic have undergone considerable changes in the fourteen centuries since the Revelation, the common origins of Classical Arabic and all other contemporary varieties of the language are still plain for all to see. It is this fact, allied to the strong sense of shared cultural values that derives from the Islamic religion and its culture, that goes much of the way to explaining the claim made by speakers of apparently mutually unintelligible Arabic dialects that they all speak the same language.

Modern Standard Arabic (MSA), or Modern Literary Arabic (MLA), is the modern descendant of Classical Arabic, unchanged in the essentials of its syntax but very much changed, and still changing, in its vocabulary and phraseology. This unified, codified pan-Arab variety of Arabic is used for virtually all writing in the Arab world and nowadays, in its spoken form, also dominates the airwaves and the television channels of every Arab country. As the normal medium for formal discourse, it is used for all news broadcasts, political speeches, official announcements, and—most crucially—education in every Arab country. Ordinary Arabs themselves do not make a systematic terminological differentiation between CLA and MSA. Both are termed ʔalʕarabi:yatu lfusḥa: "pure/eloquent Arabic" or simply ʔalʕarabi:ya or ʔalfusḥa: for short, in opposition to ʔalʕa:mmi:ya 'the vernacular', which exists in innumerable varieties and is popularly thought to be a grammarless corruption of "real" Arabic (ʔalfusḥa:). Some purists maintain that the terms ʕarabi:ya and fusḥa: should be reserved for the sixth- and seventh-century language of pre-Islamic poetry and the Koran—that is, the language as it was supposedly spoken in Arabia before the Islamic conquests brought the Arabs into contact with non-Arabs who eventually (or so it is believed) corrupted it. In practice, however, there is no chronological point at which CLA turned into MSA, still less any agreed set of linguistic criteria that could differentiate the two. MSA is merely a handy label used in western scholarship to denote the written language from about the middle of the nineteenth century, when concerted efforts began to modernize it lexically and phraseologically. Most western scholars refer to the formal written language before that

date, and par excellence before the eclipse of Arab political power in the fif-
teenth century, as "Classical Arabic."

In symbolic terms, MSA is the language of power and control, as opposed to
the language of intimacy and domesticity (the dialect), and it impinges in mul-
tifarious and sometimes subliminal ways on the daily life of Arabs of all gen-
erations, backgrounds, and educational levels. Metaphorically, and often liter-
ally, given the amount of television viewing in the average household, MSA is
the backdrop against which the business of everyday life—itself invariably in
one form or other of the vernacular—is conducted. However imperfectly ordi-
nary Arabs may have mastered its rules, and however out of place they may feel
it sounds in nonformal, everyday, face-to-face conversational contexts, they
know that MSA is always there as a kind of communally owned linguistic reser-
voir that they can dip into when they need to—a word here, a borrowed phrase
there—in order to ensure that they make themselves understood to Arabs from
distant countries or outsiders such as Arabic-speaking foreigners. In normal face-
to-face conversation, as opposed to writing, however, a blanket switch from
dialect to "pure" MSA is rare indeed, even if it were within the ability of most
Arabic speakers, and is a strategy that is resorted to only when all else fails.

In practice, of course, it is rare for all else to fail. Some dialects—usually
those of large metropolises such as Cairo or Damascus—are more widely un-
derstood than others and have acquired the status of "prestige" national or even
international spoken standards that can be resorted to in cases of cross-dialectal
conversation. When speakers from distant parts of the same country talk to each
other—Egyptians from Lower and Upper Egypt, for instance—saliently 'local'
features not shared by the speakers may be neutralized in favor of "prestige
dialect" features, in this case Cairene, that do not form part of the native dia-
lect of either.[7] In cases of dialectal contact of speakers from more widely sepa-
rated areas, the matter is a complex one but depends basically on what the
participants perceive as the minimum degree of switching to "neutral" dialectal,
MSA, or even "hybridized" forms, which is necessary to ensure smooth com-
munication in an appropriate style. This is a question to which we will return
in chapter 9.

0.3 Aims of this book

My first aim in writing this book is to give in outline a linguistic descrip-
tion of the structure of modern Arabic as it is used by Arabs today. It will be
clear from what has been said already that this will involve describing two types
of Arabic: Modern Standard Arabic—the language of writing and formal speak-

ing—and dialectal Arabic, the language of normal conversation. Because, as has
been pointed out, the latter exists in innumerable varieties, an attempt will be
made to describe the structural characteristics that the majority of dialects share
in general distinction to the structure of MSA. This book is not, therefore, a ref-
erence grammar of MSA or any dialect, still less a pedagogical manual, but
rather a snapshot of a language undergoing rapid change.[8]

My second main aim is to illustrate, through the discussion of actual ex-
amples of language behavior, how these two types of Arabic are used by na-
tive speakers for different kinds of communicative purpose. In doing this, we
will, I hope, be able to breathe life into the somewhat idealized and static pic-
ture of the language that will inevitably, and for the purposes of clear exposi-
tion quite properly, have appeared in our snapshot. What kinds of Arabic do
Arabs choose to use on a range of social occasions, and why? How consistent
is their behavior? Are MSA and dialectal Arabic really as self-contained and dis-
crete entities as many available descriptions would lead us to believe? If they are
not, and some kind of hybridization of the two types occurs, what is the nature
of the mechanism that governs it? As well as describing the outline of language
form, this book addresses the question of the complex and evolving relation-
ship between structure and communicative function in the Arabic-speaking
world of today (and tomorrow).

I hope that this book will prove to be of interest to two different types of
reader: advanced students of Arabic who have a good practical knowledge of
the standard language and perhaps one dialect, and who wish to gain a more
principled understanding of the dynamics of the evolution of Arabic and the
detail of its contemporary use; and general linguists who have little or no knowl-
edge of it, but who are interested in how it compares structurally and sociolin-
guistically with other languages. The technique of statement throughout is
therefore conservative, because my aim is not polemical but descriptive; the book
aims not to support or refute claims advanced in favor of any particular linguistic
theory, but rather to describe and explain the status quo, and how it came to
be so, in what, it is hoped, is an insightful but as far as possible theoretically
neutral manner. Whether the book succeeds in these aims is left to the reader
to judge.

Notes

1. See the essays collected in Rouchdy 1992, 83–204, and Rouchdy 2002, 133–48, for data-
   based observations on the status and evolution of Arabic in the speech of immigrants to
   the United States. Between 1988 and 1990, some 60,000 Arab immigrants arrived in the

Detroit area alone, most of them seeking refuge from the Lebanese civil war (Rouchdy 1992, 173). Abu Haidar 2002 describes the types of interference from English on the Arabic of the large London-based community of expatriate Iraqis and the generational language shift that is taking place from Arabic to English. Boumans and de Ruiter 2002 give an overview of the language profile of the Moroccan Arabic-speaking diaspora in Western Europe, which began to arrive in the 1960s.

2. This is the position espoused in Kaye 1994.

3. According to Cadora (1979, 32), Syro-Lebanese and Casablancan Arabic share only 68 percent of their base vocabulary—less than the 70 percent that Swadesh's work indicates is roughly the point at which two dialects can be considered forms of the same language.

4. Fischer and Jastrow 1980 provide a useful overview of the modern-day Arabic dialects classified from a geographical perspective, with twenty sample texts, including one from Malta.

5. A particular and pervasive feature of the linguistic situation in the Arab world is that dialectal distinctions are by no means exclusively related to geography. Lifestyle—that is, whether a community was in its recent history nomadic, village based, or urban—and even religious or sectarian affiliation are additional important factors in the dialectal structure of many states in the region, and the linguistic fault lines that run between one region, community, or even neighborhood and another can be quite sharp. A good example is provided by Baghdad, where Muslim and Christian dialects of Arabic are still sharply differentiated. Until the early 1950s, and the mass emigration of Jews, there was a three-way dialectal split along confessional lines (see Blanc 1964).

6. The only indigenous attempt—and a schematic one at that—at simplifying the complex rules of CLA was Anis Furayḥa's *Naḥwa ʿArabīya Muyassara* [*Towards a Simplified Arabic*], published in 1955. There is a presumably apocryphal story that when some of Furayḥa's academic colleagues who earned their living teaching CLA saw the title of his book, they protested: *ya: ʔusta:z, xalli:na: n'i:s!* (liberally translated: "Oh, Professor, please don't take away our livelihood!").

7. Holes 1995b shows how, in Amman (Jordanian v. Palestinian), Baghdad (Muslim v. Christian v. Jewish) and Bahrain (Sunni v. Shiʿi), dialects associated with different national, religious, and sectarian communities, and originally spoken side by side in the same city, have assumed a pecking order over the last few decades as a result of one of the dialects in each case assuming the status of a national "nonstandard standard": the East Bank Jordanian dialect in Amman, the Muslim dialect in Baghdad, the Sunni dialect in Bahrain. A major factor in this has been increasing urbanization and the consequently enhanced importance of the city in national and political life. This has led to the marginalization of the other dialects in public spaces and their relegation to a mode of in-group speech.

8. As this book was going to press, a new data-based reference grammar of modern written Arabic, *Modern Arabic: A Comprehensive Grammar*, by Elsaid Badawi, Michael Carter, and Adrian Gully appeared, unfortunately too late for me to be able to consult it thoroughly.

I

A Brief History of Arabic
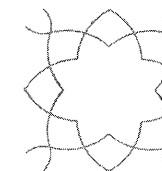
Although the purpose of this book is to describe the structure and use of modern varieties of Arabic, there are good reasons for beginning with a historical sketch of the development and spread of the language. The purpose of this is to highlight those factors that in the course of its long history have had a profound effect on its development and that in some cases—religion chief among them—remain crucial to an understanding of speakers' use of, attitudes toward, and beliefs about it. It would not be an overstatement to say that the study of the history of Arabic may also suggest the direction of future developments.

We shall make no attempt in this chapter to describe the minutiae of the structural differences between the various historically attested varieties of Arabic, but rather outline the chronology and typology of development that occurred in terms of general linguistics in order to elucidate how the present-day linguistic situation, which is the main object of this book, emerged. Inevitably, therefore, much of the detail of the story, which is in any case still obscure and controversial, is omitted. What remains is, it is hoped, a balanced and judicious account of the main lines of the language's development. Where reference is made to historical events or linguistic terminology that are unlikely to be familiar to the reader without a specialist background, these are explained, glossed, or footnoted as appropriate.

1.1 Arabic as a Semitic language

Arabic is a member of the Semitic language family, a term that designates a group of languages, some long dead, some still living, and some today having a marginal status as liturgical languages, that all show a sufficient degree of similarity of structure in their phonology, morphology, and basic lexical stock for a common origin ("Proto-Semitic") to be supposed.[1] The exact geographical homeland of the putative Proto-Semitic "mother language" is disputed, but the earliest texts written in languages that are presumed to have developed from it were composed in the ancient Near East—Syria-Palestine and Mesopotamia— and date from about 2500 B.C.

The Semitic languages are traditionally divided on grounds of both their structural properties and geographical provenance into three groups: northwestern, northeastern, and southwestern. These areas cover respectively what are now the Levant (modern Syria, Lebanon, Israel, parts of Jordan), central-south Iraq, and the Arabian peninsula and Ethiopia. Arabic is a member of the southwestern group. The earliest definite textual evidence we have for the existence of a distinct language identifiable as Arabic is an inscription found at 'En Avdat in 1986. This has been dated to the first century A.D. —recent by the standards of Semitic languages. This does not necessarily mean, of course, that Arabic was a latecomer onto the linguistic scene, because a datable written text merely shows that the language in question could be written at the time of its composition: it tells us nothing about the antiquity (or even necessarily the structure) of the spoken language that must have coexisted with it.

1.2 Arabic at the dawn of Islam

Arabic does not emerge into the full light of history until the sixth or seventh century A.D. The nomadic tribes who lived in the Arabian peninsula and contiguous areas before and at the time of the Islamic Revelation had, it seems, a thriving oral poetic tradition, the products of which have come down to us in the form of a large body of poetry. As has been pointed out above, there is no reason to believe that the Arabic in which this oral poetry was originally composed (and the vernacular form of the language that must have coexisted with it) does not reach back further in time than the relatively late period for which we have textual evidence. Indeed, some linguists point to the structure of this so-called Classical Arabic (CLA) as evidence that it does. Compared with other Semitic languages that were living at the same time (such as Aramaic), the CLA of pre-Islamic poetry shows a high degree of elaboration in its in-

flectional system, a richness in its derivational morphology, and, crucially, a markedly "synthetic" character.[2] This typological evidence has suggested to some philologists an archaic origin within the chronology of development of other known Semitic languages, although this is disputed.[3] The question turns on whether the structural characteristics of the language of the early poetry and other forms of elevated diction can be taken as the same as those of the contemporary vernacular. We will briefly consider this question, because the answer has obvious implications for what we take as the language that the Arab tribes exported to the Near East and North Africa during the conquests of the seventh and eighth centuries, which formed the basis for the development of the modern dialects.

1.2.1 Pre-Islamic poetry

The only direct evidence we have of the linguistic structure of Arabic before the time of the Prophet Muḥammad (570–632) is to be found in orally composed and transmitted poetry, the earliest specimens of which date from the early sixth century, but that began to be collected and committed to writing by the grammarians of Basra and Kufa (southern Iraq) only in the middle of the eighth century. The poems (Ar. qaṣa:ʔid 'odes') are generally short (rarely longer than 120 lines, and usually much shorter) and conform to strict prosodic rules of meter and rhyme. From the point of view of their type and purpose, several subvarieties can be distinguished: elegy, tribal or personal panegyric, and satire. The recurrent patterns of thematic structure, conventional imagery, and repeated linguistic oddities and archaisms found in this poetry point to an oral-formulaic origin of the type proposed for the Homeric poetic tradition of ancient Greece[4] and described in the modern period for certain tribal groups in southern India.[5]

As a source for the reconstruction of the contemporary nonpoetical language, spoken or written, this kind of material is problematical. In the first place, the fact that the poetry underwent a continual process of oral transmission by professional reciters, before commitment to writing, copying, recopying, and editing many years after the original composition must cast doubt on the reliability of the recensions we have as an accurate linguistic record. As Rabin comments:

> Both pre-Islamic and Islamic poems have been revised by editors, as can be seen not only from the extensive variants, but also from the not infrequent cases where verses are quoted by grammarians for some linguistic oddity, while on looking up the Dīwān [= "anthology"] we find the same line slightly reshaped so that the oddity is eliminated.[6]

These "oddities" may have reflected minor tribal dialectal differences in the speech of the poets that were ironed out at a later date in order to give the pre-Islamic poetry a consistency of grammatical structure and lexicon that it originally lacked. An even greater difficulty, however, even if we were to assume that the versions we have are exactly as they were when they left the poet's lips, lies in estimating the effects that the traditional mode of composition and the status of poetry in pre-Islamic society as "noncasual" communication had on the language used.[7]

There are two divergent views on the relationship between the language of oral poetry and the everyday speech of the poets themselves. The first, espoused by Fück[8] and, more recently, Versteegh,[9] sees the language of the oral poetry as being syntactically and morphologically identical with the Arabic vernaculars of the time. In this view, the pronouncing of case endings in nouns and mood markers in verbs (final short vowels), which is required for its scansion and, it is claimed, for its accurate construal, was a fully functional feature of nonpoetical written and spoken Arabic of the period. Although there is no body of written nonpoetic Arabic (and of course none of spoken) of a sufficiently early date, and sufficiently varied provenance, to allow this hypothesis to be tested, its proponents conclude that the linguistic situation must have been so from the fact that many grammarians of the ninth and later centuries were in the habit of resorting to unlettered Bedouin informants to settle disputed points of Arabic grammar, claiming that they alone still spoke the language in its pristine state. If the unlettered Bedouin of the ninth century were still speaking an Arabic with full case endings and mood markings, the argument goes, then the same must also have been true two centuries earlier. As Zwettler points out, however, this theory does not recognise "the possibility of a linguistic form separate from any spoken vernacular and peculiar to the non-casual verbal expressions of poetry,"[10] and further that:

> if the philologists still into the tenth century sought linguistic correctness and heard the case- and mood-endings among the Bedouin of their time, it must be recognized that what they were after was precisely the 'arabiːya (= Classical Arabic) that they knew from the Classical poetry and that the Bedouins alone could provide—because . . . they alone maintained . . . a tradition of oral poetry out of which the 'arabiːya had been generated and derived to begin with. And since it was the 'arabiːya that the philologists went out into the desert to seek, it was the 'arabiːya that they got. The products of formal elocution (poetry, oratory, etc.) were what interested them: so whether the colloquial speech of the Bedouin was

identical with the 'arabiːya or whether, as now seems more likely, it was substantially different in both morphology and syntax, we cannot definitely ascertain from the accounts and examples that have been recorded.[11]

If Zwettler's interpretation of the philologists' objectives and methods is correct, then one of the main planks on which the argument for the structural identity of poetical and nonpoetical Arabic at an early period rests is removed. Furthermore, a close study of the poetry, even in its later "polished" recensions, reveals diverse dialectal influences in some aspects of its morphology, occasional archaisms, and other forms that contravene what the later grammarians set up as normative CLA inflectional rules.[12] Far from representing a linguistic monolith, the language of the early poetry shows vestigial evidence of a heterogeneous origin. As a living, socially valued oral-formulaic art form that operated within strict prosodic rules and that was transmitted from one generation to the next, both its linguistic archaism and instability are readily understandable: these "retentions" and "borrowings" existed and were used in the poetic diction precisely because they adequately and quite satisfactorily functioned in specific prosodic contexts.[13]

The most plausible construction that can be put on the evidence we have is that the language in which the pre-Islamic poetry was composed represented an elevated "poetic style" that cannot be identified with any contemporary spoken dialect (of which in this alternative view there were a number) of pre–seventh-century Arabic. This poetic style is held to have had its origins in central and eastern Arabia and to be based on the dialect of this area, from which the dialects of the western part of peninsula, the Hejaz, differed slightly.[14] As a special register of Arabic, it may have gradually evolved in conditions, and then perhaps become fixed by constraints and conventions, that did not apply to everyday speech, although the question of the *extent* of the grammatical differences between the poetic idiom in its fully developed form and that used in the everyday spoken Arabic dialects of the seventh century is finally unresolvable given the nature and amount of the data available.[15] What cannot be in any doubt is that oral poetry, composed and recited by skilled poets, was an integral and important element in tribal life, fully comprehensible to the ordinary Bedouin of the time. The language in which this (and other elevated forms of diction) was composed, minus its lexical rarities and the grammatical frills required by its rules of prosody, cannot therefore have been very far removed from everyday speech. A parallel can still be observed in modern Arabia, where so-called Nabaṭī poetry is composed in a highly stylized, dialectally "Bedouin" form

of spoken Arabic quite different from the contemporary vernacular, but nonetheless fully comprehensible to speakers of it. It too is an integral part of the popular, nonliterate culture of its time.[16]

1.2.2 The Koran

Viewed as a purely linguistic artifact[17] from which we might make further deductions about the state of Arabic before it began to spread outside the peninsula, the Koran, the holy book of Islam, presents many of the same problems as the early poetry. Before considering these, let us first briefly review the historical record.

Muslim tradition states that the Koran (Ar. *qurʔaːn* 'reading, recitation') began to be revealed to Muḥammad, an apparently illiterate[18] Meccan of the locally important tribe of Quraysh, from approximately A.D. 610 when he was forty years old, and continued until his death in Medina in 632. Muḥammad professed to receive his divine messages from God through the Angel Gabriel, who commanded him, at the first revelation, to "Read!" (Ar. *ʔiqraʔ* (whence *qurʔaːn*)). According to tradition, Muḥammad's reply to this was *maː ʔana: bi qaːriʔin* 'I cannot read'. The Angel insisted and Muḥammad once more refused, so the Angel spoke the *suːra* 'chapter', called the '*suːra* of the blood-clot', for him. Tradition has it that Muḥammad received this revelation in a state of ecstasy or in a dream.

After an interval, Muḥammad began to have further revelations, and these continued regularly for the next twenty-three years. The early *suːras*, revealed at Mecca at the beginning of his prophethood, are very different in character from the later ones revealed at Medina, after Muḥammad had made the so-called flight (Ar. *hijra*) from Mecca with a small band of followers in 622. These early Meccan *suːras*, written in rhyming prose, are short, ecstatic exhortations to his fellow citizens to abandon idolatry and believe in one God, and to accept Muḥammad as His Messenger. They describe vividly the bliss that awaits the pious and the everlasting torment that will be the fate of wrong-doers when all are resurrected and called to account on Judgment Day. The Medinan *suːras*, on the other hand, have a quite different style. They are much longer and prosaic (in the nontechnical sense), dealing largely with matters of legal, social, and political import, and promulgating the laws and ordinances by which the nascent Muslim community was to live. The content of these *suːras* is the fundament of present-day Islamic law.

The process of collecting the text of the Koran and putting it into a canonical form was not completed until some years after the Prophet's death. Tradition has it that so many of the Muslim reciters were killed in the "Wars of Apostasy" that plunged the embryonic Islamic community into crisis immediately

following the Prophet's death that the early Caliphs entrusted with the leadership of the Muslim community feared that the Koran was in danger of being lost forever. After the Battle of Yamāma in 633, the collection began under Abū Bakr, the first caliph, or leader of the Muslim community. Collection from surviving Koran reciters continued during the next twenty years under Abū Bakr's successors, ʿUmar ibn al-Khaṭṭāb and ʿUthmān ibn ʿAffān. By 651, however, the political situation had been transformed: Syria, Iraq, Persia, and Egypt had all been conquered and the need to complete the establishment of a *textus receptus* had become more urgent. In different areas of the by now far-flung Islamic empire, variant Koranic readings were proliferating alarmingly, giving rise to fears that the substance of the scriptures would become a matter of dispute. A centrally "authorized version" was required to settle the matter once and for all. The collection of the Koran was finally completed in the middle decade of the seventh century under the third Caliph, ʿUthmān ibn ʿAffān, and disseminated to all parts of the empire with instructions that all Korans not conforming to it were to be burned. It is this ʿUthmanic recension of the Koran that has come down to us.

The Koran is the other main source of evidence, apart from the early poetry, for the state of Arabic during the premedieval period. We will now attempt to answer some questions bearing on both its formal linguistic aspects and the "textual status" it can be assumed to have had (or implicitly claimed) at the time of its revelation. Specifically, how similar to or different from the poetry that we discussed above is its language, particularly from the syntactic and morphological angles? Given that, like the poetry, its mode of transmission was oral, what relationship, similar or different, might its language have borne to the spoken Arabic of the time? Does it show any internal oddities or inconsistencies? And while, as prophetic utterance, it was *sui generis*, what genres of preexisting "oral text" might it have been akin to, stylistically speaking?

From the point of view of general textual structure, two sharp distinctions can be drawn between the Koran and the poetry. First, the Koran is written in rhyming prose and completely lacks "the isochronic metrical regularity that made up the fabric of Arabic verse rendition."[19] In other words, it is patently not the result of the careful crafting and fitting of words into preexisting structural schemata and would not therefore have been considered poetry in the sense that the poets themselves defined that term. Second, the rhyme schemes used at the end of the verse units that make up Koranic *suːras* depend on final pausal consonants rather than on final vowel cantillation, as is the case in most of the poetry. From these two formal points of view, then, the Koran is quite unlike the poetry. It did apparently bear some resemblance, however, in its use of rhyming prose to the utterances of the *kuhhaːn* ('soothsayers').

On the stylistic level, the sources and patterns of imagery of the Koran are far removed from those found in the poetry, in which a vividly physical, sensual, and profane world is evoked. More impressionistically, the packed, dense feel of the poetic line, with its flexible word order, dependence on inflection for correct construal, and frequently arcane vocabulary—summed up by Arberry[20] as "pregnant brevity and epigrammatic terseness"—contrast starkly with the simplicity of the Koranic ʔaːya ('verse'). It is small wonder that foreign learners of Arabic generally find the Koran easier to understand than the poetry: for all its sometimes ecstatic and exhortatory language, its directness and relative lack of syntactic and phraseological artifice are reminiscent of actual speech.

It can be argued, however, that there is an important functional parallel between the Koran and the poetry that has implications for any judgment of its narrowly linguistic aspects: both required the use of an "elevated diction," although for different reasons. The poetry as we see it in the seventh century is the end product of a fixed and fully elaborated tradition of ritual language use that we can only surmise must have evolved in peninsular Arabia over a long period.[21] The qualities required for language to be considered "poetic" inhered in its conforming to strict, learned conventions of prosody, skill in the use of which was acquired through practice and conferred on the user almost magical qualities.[22] But more significantly, as has already been suggested above, the traditions and physical circumstances of poetic composition had almost certainly resulted by the seventh century in a poetic language (ʿarabiːya = CLA) to some degree syntactically and morphologically archaic and not identifiable with any single contemporary vernacular. This conventionalized archaism must itself have been recognized by everyone as one of the defining characteristics that marked out poetic language as different from everyday usage. This probability—we cannot say "fact"—is of particular significance in assessing the language of the Koran. Muḥammad recited the Koran—a linguistic act—and thereby made his claim for the elevated, other-worldly status of Messenger of God: the only variety of Arabic appropriate for this in the sociolinguistic setup of Arabia at that time would have been the inflective ʿarabiːya.[23]

In its morphology, the language of the Koran shows almost complete conformity with the ʿarabiːya of the poets, although with sporadic examples of forms that we can recognize from the work of poets from the west of Arabia, the Hejaz, from where Muḥammad also came, as dialectal features not found in the east. In syntax, there is also some evidence of dialectal Hejazi influence in sentence construction.[24] The probability, then, is that Muḥammad chose an already existing type of Arabic suitable to the elevated diction required by his message that would have been familiar to his listeners but that was not his (or

for that matter their) normal vernacular—hence the occasional deviations from established ʿarabiːya usages.

The role of inflection in Koranic Arabic, which we will briefly consider, has been a matter of considerable controversy. Proof of the inflective nature of Koranic Arabic[25] (challenged by Vollers[26]) is sometimes adduced on the basis of verses that might theoretically be misconstrued if they had been recited without the final short vowel desinential inflection (Ar. ʔiʕraːb).[27] But on its own, this argument is unconvincing. The number of cases where the inflection is meaning-bearing and in which there would be a real chance of, say, subject being confused with object, compared to those where it is not, is minuscule. Whatever the arguments may be in favor of the Koran having been recited by Muḥammad with full ʔiʕraːb, they are not ones that depend on its syntactic importance. And even in the few cases where inflection *does* seem to carry a functional load in the canonical written text, comprehension of the Prophet's message may not have depended on it. Recitation is a species of *oral* performance, and, like any other type, Muḥammad's would have been marked by the use of sentence stress, intonation, and possibly paralinguistic gestures in order to make his meaning clear. In other words, Koranic ʔiʕraːb may well have been an appropriate *stylistic* feature, but it was by no means an indispensable *syntactic* one, required for the unambiguous communication of meaning.

1.2.3 Evidence from the modern Bedouin dialects

In the light of the conclusions about the functional dichotomy of ordinary language and elevated diction, poetic or Koranic, the fact that traces of one of the main systems of desinential inflection in the elevated diction (the ʿarabiːya)—that used for marking indefiniteness in the noun (tanwiːn)—still survive in some modern Arabic dialects of the "Bedouin" type[28] needs to be explained. In assessment of the evidence, it can in general be said that (i) the modern dialects in which tanwiːn occurs are always "Bedouin," never "urban," in the sense that the users either still pursue, or until recently their forebears pursued, a nomadic or pastoral way of life; (ii) as far as can be judged, those dialects that show the highest degree of retention of it are in central and eastern Arabia; (iii) the incidence is higher in formulaic language (folk poetry and epic narrative, riddles, proverbs, etc.) than it is in nonformulaic or otherwise unprepared speech;[29] (iv) the casual, as opposed to formulaic, use of tanwiːn is always optional; (v) the syntactic environments of dialectal tanwiːn, wherever it occurs, are a restricted subset of those in which CLA tanwiːn occurs, although these environments seem to differ from dialect to dialect,[30] and in some cases dialectal tanwiːn occurs where it never could have in CLA;[31] and (vi) the phonetic substance of dialec-

tal *tanwi:n* is in some areas uniformly *–in* and in others *-an*, in contrast to the *-un, -an, -in* case-determined system of CLA.

These observations hardly support the implicit but never fully articulated claim of some linguists[32] that *tanwi:n* in the modern Bedouin dialects is the direct descendant of the fully inflective CLA system that is claimed to have been used in ordinary spoken Arabic at the time of the Islamic Revelation and for several centuries after. Taken together, they suggest rather that modern *tanwi:n* is the continuation in the most conservative modern varieties of Arabic of what was certainly by the tenth and probably already in the seventh century an optional, syntactically redundant feature of normal spoken Arabic but that was, and still largely is, associated with formulaic phraseology and other kinds of ready-made language. Certainly, no one would reconstruct a universal three-vowel inflection system for the ancient dialects on the basis of the modern evidence of its syntactic and geographical distribution, were they not already convinced that CLA had such a system.[33]

1.2.4 Summary

At the time of the Islamic Revelation, the linguistic situation in Arabia can be summed up as follows. There seems to have been some degree of dialectal variation, with the main cleavage between the tribal groups living in the west (the Hejaz) on the one hand and those living in the central and eastern area (Najd) on the other. In the southwest, the Arabic of Yemen formed a continuation of the Hejazi type of dialect.[34] The grammatical and morphological differences between the western and eastern Arabic dialects were probably relatively minor, although we have no direct contemporaneous reports of what the everyday spoken Arabic of the time was like. A highly inflective poetic Arabic, or perhaps more accurately a poetic register, based principally on the speech of Najd, but no longer morphologically or syntactically identical to it, was in use throughout the whole of the area for the composition and recitation of oral poetry and other forms of elevated diction.

1.3 The spread of Arabic

The second half of the seventh century saw the founding of an Islamic Arab empire that by the beginning of the eighth century stretched from Spain to Persia. The history of how this empire was established is of course not the concern of this book. Our interest is rather in the nature and type of the social contacts that its establishment entailed between the Arabs and the conquered peoples, and the long-term linguistic results of contact and assimilation, inso-

far as these can be gleaned from a study of the Arabic used in written materials of the period and inferred from a comparison of the present-day language with what is known about pre-Islamic Arabic. There will be no attempt here to present these changes in the phonological, morphological, syntactic, and lexical substance of Arabic in any descriptive detail. The aim is rather on the one hand to outline the mechanisms by which Arabic began to replace the local languages in both speech and writing, ultimately completely supplanting them, and on the other to show in what ways the structure of Arabic evolved during this long process.

The first question to be addressed is what the linguistic situation was like in the conquered areas immediately before the conquests began. We are dealing here with the areas occupied today by the Levant states, Egypt, Iraq, and the coastal areas of North Africa as far as Gibraltar.

1.3.1 The language situation on the eve of the conquests

1.3.1.1 Syria

The ancient designation of Syria (Ar. *aš-ša:m*) covered the area now occupied by modern Syria, Lebanon, Israel, and Jordan. For many centuries until the Arab conquests, the whole of this area had been under the political control of the Byzantine emperors or their vassals. The degree of Hellenization at the grass-roots level, however, was slight: only in the cities do there appear to have been sizeable numbers of Greek-speaking government officials, merchants, and absentee landowners, whereas the vast majority of the indigenous population were peasants of Semitic stock who spoke various dialects of Aramaic.[35] For centuries, it appears that there had been some degree of contact between this sedentary, Monophysite Christian, Aramaic-speaking, and largely illiterate populace and the neighboring Arabic-speaking nomadic inhabitants of the Syrian steppe. Nomadic Arab tribes regularly summered in the vicinity of Syrian towns and settlements, and traded with local merchants. In some areas, however, there was a still more substantial Arab presence: notably in the Bekaa Valley and parts of what is now Israel, large groups of Arabs appear to have settled permanently and in such numbers that they may even have been in the majority in some areas by the sixth century. A Byzantine inscription in Greek, Syriac (the written language of the Eastern Church, in essence the same as Aramaic), and Arabic, found at Zabad, south of Aleppo, and dating to the early sixth century A.D., an era before Arabic had any known literature, shows that speakers of Arabic were frequent enough in this northerly location to warrant the effort that writing a hitherto unwritten language must have entailed.[36]
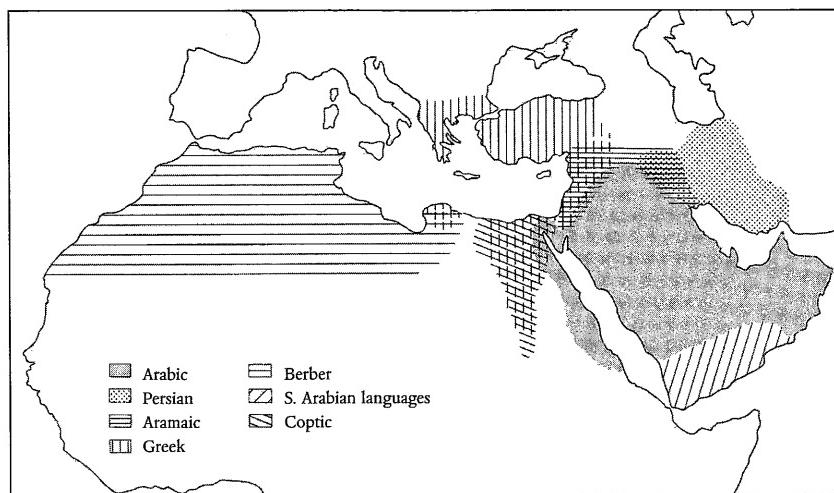
The language situation on the eve of the Islamic conquests

Thus on the eve of the Islamic conquests, there were three factors in the situation in Syria that would have helped the spread of Arabic as a spoken language: (i) there had been several centuries of trade-engendered contact between speakers of Aramaic and Arabic; (ii) some degree of permanent settlement by Christian Arabs had occurred, probably giving rise to mixed communities and some degree of bilingualism;[37] and (iii) whereas the "high" culture of Greece had failed to make much impact outside the cities and coastal ports, and had always been something imposed from outside, the common Semitic racial and linguistic roots of the Aramaic and Arab sections of the population were latent ancillary factors in the situation that might have predisposed the Aramaic speakers to accept Arabic as a language for communication with the conquerors and eventually, given favorable social conditions, as their first language.

1.3.1.2 Iraq

The linguistic situation in mid–seventh-century Iraq bore some similarities to that which we have sketched above for Syria. At that time, Mesopotamia, the fertile lowland plain of modern Iraq, formed the western border area of the Sasanian empire of the Persians, lying between the Zagros mountains to the east, a natural barrier that marked the beginning of Persia proper, and the deserts inhabited by Arab tribes to the west and southwest. Like Syria, Mesopotamia had long been something of a linguistic melting pot. The generality of the population was rural and sedentary, confessionally Nestorian Christian or Jewish, and

spoke Eastern Aramaic dialects, although there had always been sizable pockets of Zoroastrian Persian-speaking landowners and city dwellers throughout lowland Iraq, as well as Persian-speaking nomads in the foothills of the Zagros mountains. It seems also that by the mid-seventh century, the western edges of Mesopotamia had been settled by large groups of Arabic-speaking tribesmen who were in various stages of assimilation, social and probably linguistic, with the local Aramaic-speaking populace. Furthermore, there seems to have existed a loose economic relationship between the Aramaic and Arabic-speaking riverine sedentaries and the nomadic Arab tribes of inner Arabia that brought them into regular contact. The latter regularly visited the cultivated areas in order to obtain the agricultural produce and manufactured goods they could not obtain anywhere else. In Iraq as in Syria, then, an Aramaic-speaking peasantry formed the majority of the population, but with a significant admixture of Arabs of various provenances and in various stages of sedentarization. But unlike the situation in Syria, the number of speakers of the language of the imperial government—in this case Pahlavi (pre-Islamic Persian)—was probably large. It can reasonably be assumed that, as in Syria, some degree of Aramaic–Arabic bilingualism existed, although it would have been patchy owing to the nature of Arab settlement; in the cities, Persian and Aramaic would have dominated.[38]

1.3.1.3 Egypt

At this period in its history, Egypt was also polyglot, though here there seems to have been sharper ethnic and linguistic divisions than in either Syria or Mesopotamia. The population was made up of (i) the rural population of the Nile Valley and Delta, who formed the vast majority; (ii) the inhabitants of the towns and cities of the Delta and the Nile, with Alexandria as the capital; and (iii) the people who lived on the fringes of the cultivable areas to the east of the River Nile and Delta, in the deserts to the west of the Red Sea, and in Sinai. The densely populated Nile Valley and Delta were overwhelmingly Monophysite Christian and Coptic-speaking and pursued an agricultural way of life, under the nominal control of the Melkite Greek governor, army, administrators, clerks, and tax collectors. The latter lived in the cities, along with Greek traders and merchants and urban Copts. On the eastern periphery of the fertile valley and further into the deserts to the east and northeast, however, there had been a creeping process of arabization through the migration of disparate tribal elements from the peninsula, which had occurred over several centuries. By the time of the conquests, there were Arab communities already living in northwest Sinai, Sharqiya, Qena, in the towns and villages along the eastern edge of the Nile Delta, and along the Egyptian Red Sea coast down as far as Upper Egypt. Greek

historians state that some of these areas had been partially arabized by as early as 66 B.C. [39] Throughout Egypt, Coptic, as well as being the spoken language of the majority, was used in its written form as a liturgical and to some extent administrative language alongside Greek, up to and for a considerable time after the Islamic conquests.

1.3.1.4 North Africa

The linguistic, and indeed the political, situation in North Africa before the coming of the Arabs and Islam is known only in the sketchiest outline. At the time of the conquest, the North African littoral was, like Egypt and Syria, nominally under Byzantine control, imposed a century earlier in 531. Apart from in the cities, however, the Greeks had virtually no authority over, or contact with, the Berber tribes that inhabited this vast area. A narrow fertile coastal strip runs for virtually the whole length of the coastline, where the heaviest concentration of both urban and rural population has been since the earliest times. This is backed by an inhospitable hinterland that changes from steppe to mountain and then desert. These hinterland areas appear to have been wholly Berber-speaking at the time of the conquests, and, indeed, of all the earliest populations that were gradually arabicized and islamicized, the mountain Berbers have remained the most tenaciously conservative, culturally and linguistically speaking, up to the present day.

1.3.2 The early linguistic results of the conquests

By the early eighth century, the geographical boundaries of the Arab empire enclosed a vast and polyglot area stretching from Spain to Persia. Within these borders at that time, some ninety years after the initial conquest (of Syria), monoglot speakers of Arabic, whether of pure peninsula origin or mixed parentage, must still have been a small minority. It would take many centuries of gradual evolution—political, administrative, and sociocultural—before the central area was completely arabicized and islamicized, and in some of the peripheral areas neither or only the second of these processes was ever completed (and indeed, in some the first was later reversed).

In order to understand the spread of Arabic during the initial phase of the empire—a period for which we have little in the way of nonliterary linguistic evidence apart from a few chancery documents—we must turn to the historical record of the mode of colonization and the kinds of social relationship that the conquerors contracted with those they conquered. The details of this differ somewhat from one area to another, but we will try to highlight common factors and differences insofar as these may have had linguistic implications.

1.3.2.1 The pidginization hypothesis

In a controversial and thought-provoking polemic, Versteegh[40] recently put forward the theory that the Arabic that the tribesmen brought with them from Arabia (whatever kind that was—Versteegh assumes it was identical with the poetical ʿarabi:ya[41]) everywhere underwent a process of pidginization, creolization, and long-term decreolization during the first centuries of Islam. According to Versteegh's model, the conquered peoples would everywhere initially have communicated with the conquerors in a language that was neither Arabic nor their own native tongue, but a pidgin—a drastically reduced and simplified mixture of the two. In time this pidgin would have become a creole—that is, a natively spoken language—as the conquerors and their descendants began to intermarry with the women of the indigenous communities and raise children who would have acquired this creolized Arabic as their first language from their mothers. The greater the degree of mixing and intermarriage, the more rapidly would the Arabic creole thus formed have come to displace both the original languages and the Arabic of the conquerors. Decreolization, the reapproximation of the creole to the structure of the original "parent" language, in this case Arabic, would have eventually and gradually occurred through the dissemination of linguistic "models" of one kind or another—native speakers of "pure" Arabic, prescriptive school grammars and style manuals, the Koran, etc.

That a process of linguistic accommodation would have occurred in the initial phase of the conquests is indisputable. However, it is a huge leap from this to positing a full-blown "creolist" model of language acquisition. Neither the contemporaneous linguistic data, such as exists, nor retrospective extrapolation from later data support this hypothesis; and what we know of the stages and nature of the process of integration that the conquerors and conquered passed through also conjures up a social milieu different in character from those in which pidginization, creolization, and decreolization, in the generally accepted senses of these terms, typically occur.

There is no mention in the copious Arabic literature describing the social and political consequences of the conquests of the formation of any kind of Arabic that could be described as a pidgin. True, there are occasional references to linguistic errors committed by the non-Arab conquered peoples, and later there developed a separate genre within the indigenous lexicophilological tradition devoted to mistakes committed by educated Arabs in the use of the literary language (treatises on the so-called *laḥn al'a:mma* 'solecisms of the common people'). However, the examples of solecisms, culled from townsfolk's speech in the con-

quered territories, are given by writers who lived at least 100, and in most cases 200, years after the conquests and the first linguistic contacts. Where sentence examples are given,[42] they exhibit not the massively broken-down and remodeled morphology and syntax that typically result from the process Versteegh proposes, but rather the recognizable antecedents of today's modern dialects. The purpose of writing the treatises appears to have been purely prescriptive and educative: the mistakes are presented as illustrations of the extent to which certain of the norms of the 'arabi:ya—whose rules were themselves elaborated only at the end of the eighth century—were no longer being observed in the towns and settled areas at that time. In terms of Versteegh's model, what they would exemplify is a relatively late stage in the process—an Arabic creole in the process of decreolization, not of course the original pidgin. As we shall argue below, however, neither this evidence, nor the now copious and direct testimony available from early papyri, support the creolist hypothesis.

The main source of *direct* evidence for the structure of nonliterary Arabic in the early centuries of Islam is to be found in a small number of chancery documents and a vast quantity of papyri and paper documents of a mundane and ephemeral nature (personal letters, inventories, petitions, contracts, bills, etc.) discovered during the last century. A large collection of these documents, mainly from Egypt and Syria, some of which have been reliably dated to as early as A.D. 800 (still of course more than 150 years after the first contacts), has recently been analyzed linguistically by Hopkins.[43] The value of this material, in particular the ephemera, as evidence of the state of Arabic at the time, lies in the following: (i) much of the material was written before any prescriptive, codified system of written Arabic grammar had been disseminated, and, unlike the poetry, it is inconceivable that such material would ever have been systematically tampered with or edited after it was written, precisely because of the banality of its content; (ii) as nonliterary material for personal and domestic consumption, it is unlikely to have been written in any form of "elevated diction," but rather reflected everyday usage; and (iii) it exists in quantity, is of varied provenance, and shows several subgenres. The drawbacks are the extreme difficulties posed by deciphering the sometimes abominable scrawl, and in the orthography, at that time lacking diacritical points to distinguish many of the letters (let alone any vowel marks whatsoever). On top of this, there is a good deal of variability in spelling conventions, some of which may be because of scribal lapses, which makes it difficult to get a clear grasp of how some aspects of the morphology and syntax worked. Finally, it cannot be assumed that written Arabic of this kind, albeit written at a time before the standardization and codification of the written 'arabi:ya had taken hold, exactly mirrors the structure of

the spoken language of those who wrote it. Prima facie, it seems likely that the two would not have been very different, but a somewhat more formal style might have been used even in these kinds of ephemera.

Despite all these problems, Hopkins' cautious analysis leads him to one quite definite conclusion regarding the status of the language in which these documents are written, which is worth quoting at length:

> in almost every case in which the language of the Arabic papyri deviates from CLA, it deviates unmistakably in the direction of Middle Arabic, typologically akin to most of the modern colloquials. The 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 the 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 seen to lie very deep.[44]

This conclusion, based on a large amount of datable and authentic evidence, is difficult to reconcile with the pidginization–creolization hypothesis. Even allowing for the reading difficulties and scribal inconsistencies mentioned above, the material examined by Hopkins is written in a language full of morphological and syntactic variation. Variation and instability on this level suggest a language in a transitional phase, and it is apparent that the transition is from a variety of Arabic that has many of the features that we now exclusively associate with Classical Arabic (codified and fixed *after* many of Hopkins' texts were written) to one containing many, but not yet all, of the features typical of later medieval and modern varieties of informal written and spoken Arabic. Thus, while we note the total absence of some features that we associate with that variety of Arabic later codified and reified as "Classical Arabic" and the categorical presence of the corresponding "colloquial" feature (e.g., the complete loss of case marking in the sound masculine plural of nouns), in other subsystems of the language we find variation between what we now think of as the "dialectal" and the "Classical" system (e.g., in the morphology of numbers), and in yet other cases the dominance of the "Classical" feature and absence of the equivalent "dialectal" feature is clear (e.g., in the means for marking mood and aspect in the verb).

These late eighth- and early ninth-century ephemera are the linguistic "missing link" between early (ca. 640–800) spoken and nonliterary written Arabic, of which we have no extant unadulterated examples, but that we would argue

must have been derived from a form of Arabic structurally akin to, though not identical with, what later became codified as "Classical Arabic," and the later, fully documented medieval varieties of Middle Arabic that in turn evolved into the modern urban dialects. At no point during these twelve centuries of evolution from about A.D. 800 does there seem to have been any violent dislocation or change, but rather a long and gradual evolution toward the present dialectal situation.[45] What does this mean for the pidginization-creolization model?

Pidginization always involves a drastic breaking down and simplification of the structure of the input language. A modern example of the kind of thing that happens can be observed in the southern Sudan, where an Arabic pidgin, based presumably on Egyptian colloquial Arabic and variously known as Mongallese, Juba Arabic, or Bimbashi Arabic, was formed as a result of Egyptian military campaigns in the nineteenth century. It was later exported as a creolized variety known as (Ki)Nubi.[46] Juba Arabic and (Ki)Nubi are virtually unrecognizable as varieties of any kind of Arabic, ancient or modern. There has been a remodeling of the input language such that, to give but a few examples, the phonology of the pidgin/creole has lost all of the postvelar consonants (/ġ/, /x/, /ḥ/, and /ʻ/) so typical of all varieties of "heartland" Arabic in all periods, and, in the morphology of the verb, there has been a total loss of all inflection for tense and person and the collapse of all verb classes into a single paradigm[47]—again, completely unattested phenomena[48] for earlier periods of Arabic. The question then is this: if the Arabic of the conquering tribesmen in the middle of the seventh to the end of the eighth century was everywhere being pidginized and creolized, and presumably giving rise to drastically reduced varieties along the lines we today observe in Juba and some other parts of Africa, is it conceivable that, by the eighth/ninth century—the date of Hopkins' earliest ephemera—these pidginized/creolized Arabics could have been decreolized to the point where they have the morphological and syntactic structure Hopkins describes? Or, even if we assume that spoken and written Arabic were already different in the eighth/ninth century, is it conceivable that a creolized form of Arabic of the structural type exemplified by (Ki)Nubi could have coexisted as a spoken form alongside the very different variety of written Arabic that we see examples of in Hopkins' texts? The answer to both questions is almost certainly no. For the answer to be yes, a quite incredible series of coincidences would have had to occur in different parts of the new Arab empire, and a very different social setup from the one the historical sources describe would have had to exist.

Pidgins result from the need for communication, normally in a limited range of contexts between an indigenous people and outsiders. In this case, the in-

digenous people spoke a variety of languages—mainly Aramaic, Coptic, and Berber—with no close geneological relation to each other. In all cases, the incoming language was Arabic. In this situation, we should expect pidgins to be formed from a linguistic accommodation between Aramaic and Arabic, Coptic and Arabic, and Berber and Arabic. The resulting pidgins would all have been very different from each other: compare the differences, say, between a Chinese–English pidgin and New Guinea–English pidgin, both based on English but in which the other "parent" language is different. These pidgins would then in turn have given rise to a range of divergent creoles (compare, say the French-based creoles of West Africa with those of the West Indies), which would have been the earliest Arabic urban colloquials and all of which would later, following Versteegh, have been decreolized—that is, would have all moved in the direction of the imposed standard form of Arabic (in creolist terms, the "acrolectal standard"). The problem is that the linguistic facts as we know them do not support this scenario.

The earliest forms of urban non–Classical Arabic we know, of whatever provenance, from Hopkins' materials to those for the somewhat later period of Blau,[49] share a remarkably similar structure, and all differ in the same ways from CLA. These typological similarities have continued up to the present day, and far from being the result of creolization, the degree and pattern of similarity have led some linguists to propose an original non-Classical koinè as the most reasonable explanation.[50] Suffice it to say that all the early varieties, as well as their modern descendants, show a similar reduction of morphological distinctions and categories, the development of "analytic" features in syntax, a greater degree of symmetry in syntax, and (insofar as it is possible to judge from spelling) the loss of the same phonological distinctions. Given the different substrate languages in different parts of the empire, it is extremely difficult to explain these similarities in creolist terms. If there had been separate processes of pidginization in, let us say, Aramaic-speaking Syria and Coptic-speaking Egypt, the resultant creoles that we see a century and more down the line should (i) be much more different from each other than they actually are, (ii) show a much greater degree of local lexical and substrate syntactic influence, and (iii) show a much more reduced morphology and syntax, compared with CLA, than they actually do. Purely on the basis of a prima facie examination of the data, it seems very unlikely that the early non-Classical varieties could all be the result of the decreolising influence of CLA on disparate Arabic creoles with structures like that of (Ki)Nubi.

It is also unclear who or what would have been the purveyors of the model of "pure Arabic" to bring about this decreolization, and over what time scale.

For the first several centuries after the conquest, functional literacy in the ʿarabiːya was exceedingly rare in the population at large (including the conquerors themselves). Effectively, it was necessary only for a tiny minority of officials involved in tax collection and other administrative matters.[51] Rudimentary Koranic schools (Ar. kuttaːb) may have been gradually established from an early date for Muslim converts,[52] but with its complete reliance on rote learning and recitation of the scriptures, the seventh/eighth-century kuttaːb could no more be said to have provided an "acrolectal model" (to use the creolist term) to be approximated to in nonfrozen contexts of speech than its equivalent does today. Perhaps, then, we could look to the "pure" spoken Arabic of the largely illiterate conquerors themselves, their descendants, and later groups of immigrants from the peninsula to provide the model for decreolization. This explanation has attractions, because the Arabic of the immigrant groups, untainted by substrate influences, would have provided regular and relatively homogeneous infusions of "pure" Arabic into the developing linguistic melting pot of the cities and their immediate hinterlands.[53] It has already been argued above, however, that whatever this "pure" spoken Arabic was, it was unlikely to have been identical in syntax or vocabulary with the poetical ʿarabiːya, but was rather a group of mutually intelligible dialects with a high degree of structural similarity, in which some of the most characteristically "synthetic" features of the ʾarabiːya had already disappeared or were in the process of disappearing. The "acrolectal" standard would not therefore have been provided by an imposed ʿarabiːya but rather by the speech of "pure-blooded" Arabs—that is, various peninsula dialects. Only at a very much later date, when the ʿarabiːya had begun to be disseminated as the vehicle of Islamic culture in a fully islamicized society, could literate norms as such have exercised any decreolizing effect on the population at large. But even then, access to these norms through education was always severely limited; indeed, there were still until very recently many parts of the Arab world where the level of functional literacy in the ʿarabiːya (or, rather, Modern Standard Arabic [MSA], its modern equivalent) remained negligible. The pidginization model would thus work like this: (i) initial pidginization of the conquerors' dialect by the indigenous populace; (ii) long-term creolization of this pidgin as intermarriage and closer social contact between the two groups begins; (iii) decreolization through continuous contact with new immigrant populations from Arabia; and, very much later on (up to and including the modern era) (iv) the beginnings of the decreolizing effect of a prescriptive ʿarabiːya as the codified language of Islamic culture. However, even if this model or something like it is accepted, the problem still remains of explaining the remarkable and detailed similarities that exist between the results of this process—the early non-Classical varieties (= 'cre-

oles') and eventually the modern dialects—given the huge differences in the parent languages that were supposedly involved at the pidgin stage.

A more plausible explanation of the linguistic facts we have is simply to assume that the indigenous population learned Arabic from the conquerors *as a foreign language*, without the need to break down its structure. What we know about the immediate aftermath of the conquests is that the initial need was to set up an administrative and fiscal system in the abandoned towns, a task that the Arabs initially seem to have been content to leave to what remained of the local government after the Byzantines and Persians had left. This class of clerks was obviously literate and, in Egypt and Iraq, bilingual in the local language and either Greek or Persian; in Syria, Greek was the language of government. Such people, already accomplished language learners, were now facing a need to learn to communicate in speech (if not for some while yet in writing) with their Arab masters: why could they not have learned to do this directly, perhaps with the help of local bilinguals who knew Arabic? After all, as we have already noted, the circumstances were propitious: there had been contact with Arabic-speaking visitors and settlers for many centuries in all the conquered areas, although mainly outside the cities. It may well be that, immediately after the conquests, ephemeral forms of "kitchen" or "pidgin" Arabic arose as monolingual tradesmen and farmers struggled to do business with the new arrivals in the circumscribed contexts of buying, selling, and the daily round; but, in the towns at least, which is where the Arabs in Egypt and Syria were mainly concentrated and rapidly became settled in considerable numbers, there is every reason to suppose that, out of sheer self-interest if nothing else, the local townsmen would have set about learning to speak Arabic back to Arabs as it was spoken to them.[54]

1.3.2.2 The arabicization of Egypt, Syria, Iraq, and North Africa

The arabicization and islamicization of the conquered territories were separate but overlapping, long drawn-out processes that passed through a number of distinct stages in all parts of the new empire, although the speed and mechanics of how this happened differed somewhat from area to area. In Egypt, the invaders at first confined themselves to the towns and left the countryside in the hands of the Copts. There was no encouragement to mass conversion to Islam during the early centuries because this would have eroded the taxation base created by the Byzantines and taken over by the Arabs. There appears to have been a period of decades of peaceful coexistence in which Coptic- and Greek-speaking Egyptians continued to hold high office under the new dispensation and during which Arabic was not even used in official government

documents.[55] From the early eighth century, however, Arabic begins to appear in written documents and gradually becomes dominant over Coptic and Greek. It can be surmised that the same shift from Coptic–Greek to Coptic–Greek–Arabic multilingualism to Arabic monolingualism must have been happening concurrently in urban public life generally. However, certainly during the first Islamic century, the total number of native Arabic speakers compared to Copts was tiny. At the turn of the eighth century, there were probably about eight million Copts and no more than 80,000 Arabs, concentrated in Fusṭāṭ, the old section of what was later to become Cairo, and Alexandria.[56]

The period from the ninth to the twelfth century witnessed a number of developments that all tended to favor the spread of Arabic: a continuous flow of immigration into Egypt from Arabia; the embracing of Islam by the Copts on a much larger scale than hitherto in order for them to escape payment of the *jizya* ('poll tax'); a tightening of the conditions for employment by the state, so that adherence to Islam became a requirement; and the disbanding of the Arab army in 833, leading to a greater degree of mixing between the Arabs and Copts in civilian and social life in urban areas. It is interesting to note that from the thirteenth century, grammars and dictionaries of Coptic begin appearing, probably a reflection of efforts to keep the language alive among educated Copts in the cities who were fast losing it. In the countryside, however, the processes of arabicization and islamicization appear to have been much slower. In 1673, the European traveler Vansleb reports meeting the last Coptic speaker, although as late as the present century, there are reports of whole villages in Upper Egypt still speaking Coptic.[57]

To sum up, the arabicization of Egypt was a long process that began in the cities, where Arabic was learned initially as a (sometimes second) foreign language by the Coptic intelligentsia and used during the first Islamic century only as a language of government and administration, and, for the few early converts, of Islam. Over the succeeding centuries, the practical attractions of Islam, which increased as the Arabs introduced a more restrictive employment regime within the administration, led to increased conversion and gradual arabicization. These processes were no doubt aided by the continuous inflow of Arabic-speaking migrants into both the towns and rural areas and by the resulting greater social contact and mixed marriages between Arabs and Coptic converts to Islam.

The arabicization of Syria is less fully documented than that of Egypt, but it is thought that there were some important differences in how it came about. There certainly seems to have been a mass flight of the Greek population (much greater, it seems, than in Egypt) from the cities—Damascus, Homs, Aleppo, and Jerusalem—which were gradually filled by Arab settlers. As in Egypt, the num-

ber of Arab invaders was small in comparison with the indigenous population: probably no more than 40,000 men took part in the decisive Battle of Yarmouk in 636.[58] Unlike Egypt, however, in which the military cantonment of Fusṭāṭ later turned into a city in its own right, no new urban centers were created and the sources do not hint at the kind of early mass migrations in the centuries following the conquest that occurred into, and through, Egypt. After the end of the Umayyad dynasty in 750, tribal migration into Syria increased, but an educated guess would put the numbers of Arabs at the end of the eighth century (almost entirely confined to the cities) at between 200,000 and 400,000, as against 4 million non-Arabs.[59] The Syrian countryside, peopled at the time of the conquest by Aramaic-speaking villagers with little in the way of movable possessions to take with them—even if they had wanted to leave and had anywhere else to go—seems to have remained very much the same as it had been before the invasion, with little or no sequestration or settlement by the invaders.[60]

As in Egypt, the written language of administration remained at first Greek, with Arabic beginning to replace it from the beginning of the eighth century. As in Egypt too, mass conversions to Islam did not begin until the eighth and ninth centuries, and again the motives appear to have been at least as much ones of economic pragmatism—the desire to escape from the tax burdens of being a *ðimmi*: ('(Christian or Jewish) protected person')—as they were of religious conviction. Linguistically, the cities of Syria, because of the mass exodus of the Greeks and the inflow of Arabs, must have been polyglot from the earliest period of the conquest, with varieties of Arabic, Aramaic, and Greek all in use. We have already noted that there had been a considerable degree of contact with Arabic speakers from Christian Arab tribes in the centuries before the conquests, and parts of Syria–Palestine had been settled by Arabs from the fifth century, so some degree of familiarity with Arabic among the population can be assumed, at least among those sections in whose interest it would have been to know Arabic (e.g., traders and other providers of services). Not surprisingly, Arabic seems to have supplanted Aramaic in the cities relatively rapidly, where it is likely to have developed a role as a lingua franca between the various ethnic groups. In two to three centuries, its use had spread even to the extent of replacing Aramaic as a spoken and literary language (although written in the Hebrew script) among the Jewish population.[61] Similarly, a Psalm fragment of apparently Christian Syrian origin, written in Greek uncials but in which the language is (Middle) Arabic, and dated with reasonable certainty to A.D. 800, also bears witness to the use of Arabic among Christians, even in religious contexts.[62] The text appears to have been written for a readership (and perhaps by

a scribe) who could understand Arabic but not read it, or perhaps for whom written Arabic in a religious context was associated with Islam and hence inappropriate. In passing, it is worth mentioning that the pidginization hypothesis appears even less plausible in Syria than elsewhere, because the close similarities in morphology, syntactic structure, and vocabulary between Arabic and Aramaic would arguably have made it unnecessary even if the social conditions for it had existed. This is not to say that Arabic did not change as a result of Aramaic influence, of course, but these changes are more plausibly explained as the outcome of a protracted period of language contact and communal bilingualism.[63]

Aramaic–Arabic bilingualism, we can be fairly certain, lasted for a very long time outside the cities. This has been strikingly demonstrated by Garbell in a study of the non-Bedouin Arabic dialects of Syria, Lebanon, and Palestine.[64] She shows that these dialects underwent three successive stages of Aramaic influence on their phonology lasting eight centuries in all. Such developments could not of course have occurred unless both languages had been in continuous concurrent use over this period. Although the use of Aramaic at home and with other Aramaic speakers, and that of Arabic for public and interethnic communication, probably lasted for centuries across the vast area of Syria–Palestine outside the cities, the increasing prestige and use of Arabic as the language of government and state religion, together with the growth of mixed marriages[65] and the bringing up of the resultant offspring as Arabic-speaking Muslims, were factors that must have accelerated the disappearance of Aramaic. It is important to emphasize that the importance of Arabic was as a *spoken* language among the population at large; outside the cities, knowledge of written Arabic was probably virtually nil among native speakers of Arabic and foreigners alike. Even today, rates of illiteracy among rural communities remain high in all parts of the Arab world. The Hellenistic school system that the Arabs inherited in Syria played no part in the spread of Arabic. To the extent that it survived at all, it seems to have furnished a means by which the Arabs acquired the rudiments of science and philosophy, rather than as a means of teaching Arabic to foreigners.[66] The Koran schools, which developed from the late seventh century, can in no sense be considered to have constituted a means of fostering functional literacy in the population at large.

In Iraq, a name that at the time of the conquests designated only modern central and lower Iraq, and not the mountainous north, the linguistic results of conquest were in the long term the same as in Syria: complete arabicization. As has already been mentioned, there had been a large population of semisettled, mainly Christian Arab tribesmen living on the western edge of the fertile Iraqi

sawa:d ('alluvium') for decades, if not centuries, before the Arab conquest, although it is unlikely there was much penetration of Arabs into it before this time. The bulk of the ordinary population, rural and urban, were Aramaic-speaking Christians and Jews, no doubt the descendants of the ancient Semitic and non-Semitic populations of Iraq, with a sprinkling of Pahlavi-speaking land-owning Sasanian nobility in the countryside and a Pahlavi-speaking government administration in the towns. Again, as in Syria, arabicization must have been a long process, particularly in the countryside. There was no mass desertion of the land, and the conversion of Christians and Jews was not encouraged early on, in order to avoid erosion of the taxation base. However, in Iraq, in contrast to Syria, there were large migrations of Arabs from all parts of Arabia immediately after the conquest, and new, entirely Arab cities were founded on the base of the originally military cantonments at Kufa, in central Iraq, and on a smaller scale at Basra in the south.[67]

It is likely that Arabic would have become quickly dominant in the erstwhile Sasanian provincial capital of Iraq at Mada'in and in other central Iraqi cities as the language of government and administration, but, just as importantly, it would have served as the general lingua franca between Arabs, Aramaic speakers, and what remained of the Pahlavi-speaking population, which was augmented by mass desertions from the Sasanian armies to the Arabs after their defeat. In the succeeding centuries, the factors that ensured the disappearance of languages other than Arabic in the lowlands of central and southern Iraq were the continued massive flow of Arabic-speaking tribesmen into the towns and villages, mixed marriages and the consequent gradual linguistic assimilation of the indigenous communities over succeeding generations, and the eventual wholesale abandonment of other religions for Islam as this became economically attractive and politically encouraged. In the mountainous terrain of north and northeastern Iraq, however, Kurdish-speaking tribesmen accepted Islam but maintained their language intact. To this day, they constitute Iraq's largest linguistic minority.

The arabicization of North Africa proceeded at a much slower pace than anywhere else in the early centuries of the Islamic empire for a combination of demographic, political, and topographical reasons. The nature of the struggle was different in that the Arabs initially faced not only the Byzantines in their strongholds, dotted along the coast, but also fiercely independent Berber tribes, some partly Christianized, whom neither the Byzantines nor the Romans before them had ever subjugated. Only when the local Berber tribesmen in each area had been persuaded that it was in their interest to accept Islam and fight alongside the Arabs did the defeat of the Byzantines become a reality. This is of course a

simplification, and a full account of the vicissitudes of the historical record, in which the Arabs often found themselves facing mutinous Berber "allies" is well beyond the scope of this book; but it can be said in summary that there were several distinct stages in the early period of the arabicization and islamicization of the area that here, more than anywhere else except perhaps Persia, were separate processes.

Islam was initially accepted by the Berbers largely because enlistment in the Arab army guaranteed being put on the same footing as the Arab soldiers in the distribution of booty.[68] This was especially true of the conquest of Spain in the early eighth century. The earliest Berber converts were the lowland nomadic tribes of Tripolitania, Cyrenaica, southern Tunisia, and Morocco, rather than the montagnards of the Aurès and Kabylia ranges (Algeria) and the high Atlas (Morocco). However, it appears that these early conversions were nominal, and only after some degree of political stability had been achieved (in about 717–720) was a systematic effort made to convert the Berber population throughout the Maghreb. The concomitant process of arabicization appears to have been very slow, given the vast size of the territory and the sparsely populated nature of much of it. As had been the case with the previous imperial power, the Arab–Islamic power base was in the towns, either taken over from the Byzantines or, like Qayrawān and Tunis, founded by them. Everywhere else, the non–Arabic-speaking but islamicized Berbers remained the dominant group as, in the mountainous areas, they still do to this day.

The major demographic change that altered this balance did not occur until the middle of the eleventh century, when the first of a series of massive migrations of Arabic speakers from the east occurred. An estimated 50,000 warriors of the Banī Hilāl and Banī Sulaim, in addition to their women and children, moved west from Egypt into Cyrenaica. Some settled here, but many continued into the fertile plains of Tunisia and what is now Algeria. Further westerly migrations followed, reaching Morocco in the twelfth century. These nomads did not settle in the towns, but rather began to occupy the plains. The linguistic effect was the spreading of knowledge of the Arabic language in the countryside instead of it remaining limited to the towns. Before this invasion, the Berber dialects had formed the means of communication in the Maghrebian (= North African) countryside, but as groups of these Arabic-speaking nomads came from Egypt, Arabic gradually replaced the Berber dialects.[69]

1.3.2.3 Summary

Several common strands can be identified in the process by which Arabic spread in the newly conquered areas:

(a) Preconquest contacts with Arabs

With the exception of North Africa, the evidence is that the peripheries of the conquered territories had had extensive contact with Arab tribes and that there had been some degree of migration to and settlement on the fringes of cultivable areas by Arabic speakers over several centuries (west-central Mesopotamia in Iraq, east of the Nile Delta and Upper Nile in Egypt, some river valleys in Syria–Palestine). It can safely be assumed at the very least from this that there was some preconquest familiarity with Arabic in these areas.

(b) Islam

Initially, this was the *least* important influence in the spread of Arabic, although it came to acquire a central educative and hence linguistic role later on. The contrasting linguistic effects of the conquests as diffusers of spoken Arabic and that of Islam as a second-wave normative influence have been succinctly put by Wansbrough:

> Symbolically the contrast would be one between natural, uninhibited diffusion and artificial, consciously directed restriction. Linguistically arabicization is characterized by a concept of language as the most convenient means for meeting the demands of normal communication (*Mitteilungsbedürfnis*), and islamization by a concept of language as an instrument of education (*Bildungsprinzip*). An example of the first was the introduction of Arabic as official language into the Umayyad chancery; the example of the second was, of course, the Koran.[70]

(c) Urbanization

Everywhere, the conquerors occupied or established towns and cities that became centers of local or regional power. For administrative and economic reasons, as well as for ease of communication, Arabic became the language of the people who lived or migrated to these often initially polyglot towns and cities, and of trade between them. The rural peasant communities, on the other hand, as well as being more slowly islamicized, were also much slower to give up their original languages. Communal bilingualism must have been normal in the rural areas of all parts of the empire for very many centuries.

(d) Migration and assimilation

The size of the conquering armies was small—tiny, even, compared to the sizes of the indigenous populations. A major factor over several centuries in the arabicization of both town and country in North Africa, Egypt, Iraq, and (somewhat later) Syria was the infusion of new Arab blood through large-scale tribal migration from Arabia and remigration from already heavily arabized to

less arabized areas. The increased pace of conversion to Islam among indigenous populations from the ninth century must also have indirectly speeded up the process of social integration and linguistic change through mixed marriages.

The third and fourth of these factors, taken together, were without doubt the real engines of arabicization during the first two to three Islamic centuries in the eastern Mediterranean and Fertile Crescent, and in large parts of North Africa until the sixth Islamic century (thirteenth century A.D.). Preconquest contacts with Arabic speakers may perhaps in some places have been a linguistically predisposing factor, although they were not a necessary one, as is shown in the case of North Africa, where no such contacts occurred. Nor was Islam, in the absence of large infusions of Arabic speakers, a force that by itself could bring about wholesale linguistic change. These points are illustrated by the negative example of the linguistic history of Persia and other islamicized regions of central Asia, which were never subject to mass migration from Arabia, although they were islamicized, and where consequently Arabic never took over from Persian or the other Indo-European and Turkic languages spoken in these vast areas as the language of daily life. Although it was for a long period the language of government, religion, culture, and polite society, thereby influencing the lexical stock of the languages of the subject peoples, Arabic never became a general vernacular language here precisely because the social factors that would have been required to make it one never came into existence. Today, throughout modern Iran and Muslim central Asia, Arabic has receded to the point of being a scriptural language only.[71] The same thing applied until recently to parts of the heartland areas of the present-day Arab world that for topographical reasons were less attractive to tribal migrants from Arabia and hence never became settled by them and arabicized: the mountainous areas of northern Iraq, still populated by Kurdish-speaking Muslims, and the mountain ranges of present-day Algeria and Morocco, the home of Berber speakers. Only relatively recently, with the establishment of modern nation states in which Arabic has been adopted as the official language and propagated through a centrally controlled education system has functional bilingualism become widespread among these unassimilated non-Arab populations.

1.4 Middle Arabic, the modern dialects, and the evolution of Modern Standard Arabic

Having surveyed the reasons for the spread of Arabic in the early Islamic period, we will now trace its general lines of development through the Middle Ages and into the modern era.

We distinguished earlier between the 'arabiːya, defined as that variety of Arabic used in pre-Islamic poetry, the Koran, and "elevated diction" in general, and the tribal dialects (which we will henceforth refer to generically as Old Arabic, or OA) used in everyday speech. The chief distinction between them was that the 'arabiːya retained certain morphosyntactic features—most notably the final short vowel endings indicating mood and case—that OA had probably begun to lose by the late seventh century. This trend toward the loss of inflection in spoken Arabic accelerated in the circumstances in which Arabic was learned in the towns and cities of the conquered lands and was part of a general restructuring of OA involving, over time, a reduction in distinctions/categories and greater paradigm symmetry in phonology, inflectional, and derivational morphology; a more analytic and periphrastic phrase structure in syntax; and lexical borrowing from the various substrate languages. The dialects of the Bedouin who remained in inner Arabia were not affected by these contact-induced changes and retained many features of OA that were quickly lost in the conquered territories. This linguistic purity, later to become a popular and romanticized idée reçue, in part accounts for the common practice of the medieval Arab grammarians (q.v.) up to about the eleventh century of resorting to unlettered Bedouin informants during their investigations into and codifications of the 'arabiːya.

It is possible to trace many of the developments in Middle Arabic through time by studying surviving written ephemera and other kinds of text not subject to the normative influence of the 'arabiːya,[72] although the establishment of a precise comparative chronology for developments in different areas of the conquered territories is likely to remain an impossibility. For one thing, there are lacunae in the geographical and chronological coverage of the surviving texts, so that, for example, we have little or no material for Iraq or North Africa of the requisite type and almost none for the first one and a half Islamic centuries. Second, the degree to which the morphological and syntactic norms of the 'arabiːya may have influenced "informal" writing in the later Middle Ages remains an uncertain factor. The degree of influence depends partly on who wrote the text—Christian, Jew, or Muslim—because from an early period the different religious communities developed different conventions for what purposes Arabic was to be used for: secular or religious literature, religious ritual, everyday intercourse within the group, intragroup contact, etc., and consequently divergent communal norms and styles.[73] However that may be, it is possible to discern a clear line of development from the earliest written Arabic ephemera (ca. A.D. 800) through medieval Middle Arabic texts to the modern urban colloquials, at least for Syria–Palestine and Egypt.

In contrast to the relative obscurity that cloaks how spoken Arabic and informal written Arabic developed, the elaboration of the 'arabiːya as the linguis-

tic vehicle par excellence of Islamic culture and ultimately of all kinds of written communication in Arabic unfolded in the full light of history. In 750, a century after the first wave of conquests in the Fertile Crescent, the Umayyad dynasty, which had sedulously promoted specifically Arab political and economic interests at the expense of the *muwalladu:n*, the non-Arab converts to Islam, was overthrown. The center of power shifted to Iraq, where the Abbasid dynasty represented the aspirations of the heterogeneous polity of the eastern regions of the Muslim empire. Arabic had by this time firmly established itself in Iraq, as elsewhere, as the language of government and of Islam, although almost certainly not yet as the language of everyday intercourse outside the arabized towns. In these towns, as we have already noted, it was rapidly changing on account of its use as a spoken lingua franca. Not long after the rise to power of the Abbasids, we see the beginnings of philological activity in the new Iraqi towns of Kufa and Basra, first established about a century before as military cantonments for the armies of conquest. The motives traditionally imputed to those who began the codification of "correct" Arabic usage are summed up in the following passage from the fourteenth-century writer Ibn Khaldūn's *Prolegomena*:

> When Islam came, the Arabs left the Hejaz to seek the royal authority that was in the hands of [foreign] nations and dynasties. They came into contact with non-Arabs. As a result, their linguistic habit changed under the influence of the solecisms they heard non-Arab speakers of Arabic make, and it is hearing that begets the linguistic habit. Thus, the [Arab linguistic habit began to] incline towards adopting forms of speech at variance with it, because the Arabs became used to hearing them spoken, and their linguistic habit became corrupted. Cultured people feared that the Arabic linguistic habit would become entirely corrupted and that if the [process] went on for a long time, the Quran and the traditions would no longer be understood. Therefore, they derived certain norms for the Arabic linguistic habit from their way of speaking.[74]

The derivation of these "certain norms" and their codification can be seen as the second major step in the process of binding together and imposing order on the Muslim community, the first having been the fixing of a canonical version of the Koran about a century earlier. Both of these essentially "logocentric" events manifest the desire of the Arabs for Islam and the *'arabi:ya* in which it was made manifest to act as centripetal educative and political forces to counteract the centrifugal tendencies of arabicization, which, as we have seen, was creating communities in which uncontrolled linguistic innovation and dialectal diversity were bound to flourish. The grammatical tradition of the *'arabi:ya*, first elaborated in

the late eighth century by al-Khalīl and his Persian Muslim pupil Sībawayhi, had as its initial objective the extrapolation of the linguistic structure of the pure *'arabi:ya* from texts whose Arabian character was unimpeachable: the old poetry and the Koran, where necessary with contemporary supporting evidence from Bedouin informants. The need for such an enterprise was pressing: there is no reason to doubt Ibn Khaldūn's assertion that, from the earliest period of the empire, the *'arabi:ya* would have been, if not a foreign language, substantially different from the natively spoken Arabic of both pure Arabs and *muwalladu:n* living in the towns.

Over the succeeding seven centuries, the system of description devised by Sībawayhi was refined but never substantially changed.[75] It was puristic and prescriptive in attitude and taxonomic in approach, and sought to specify in the minutest detail what was and was not "correct."[76] Much of its early methodology and terminology appears to have been borrowed from the law,[77] and by the twelfth century there are treatises that directly equate the principles of grammatical reasoning with those of the law.[78] This nexus is not surprising, because the bases of law and government in Islam—the Koran and the recorded sayings of the Prophet—are linguistic events frozen for all time whose exegesis precisely requires grammatical and lexicographical skill. In public life in general, the ability to speak and write correct Arabic, to avoid *lahn* ('solecism') as defined by the grammarians, was both the condition of entry to and the distinguishing mark of the elite.[79] In particular, up to the end of the early Abbasid period (ca. A.D. 850) "the correct use of *ʔi'ra:b* (= case and mood endings in nouns and verbs) apparently was a necessary condition for an intellectual career, just as bad manners were betrayed by a wrong use of the endings."[80] In some respects, the linguistic situation in this period was analagous to that in the late Roman empire, with the *'arabi:ya* functioning in a similar way to Classical Latin as a language of scholarship and public life, and Vulgar Latin (= the Middle Arabic vernaculars) as the language of everyday intercourse.

By the end of the Abbasid period in 1258, we witness a gradual diversification of style in written Arabic. On the one hand, the *'arabi:ya*, as defined by the philologists and jurisconsults, gradually became a rigid system of formalized rules, forms, and vocabulary that at least theoretically could not be improved on. It remained in use in its pure form as the vehicle of all religious and doctrinal writing, as well as of philology itself. But on the other hand, interest in Hellenistic culture resulted in the establishment of a translation school in the newly founded city of Baghdad, where not only Greek but also Syriac works were translated into Arabic. Much of the material translated was on scientific subjects for which Arabic lacked any terminology—philosophy and medicine, for ex-

ample—and the work was often carried out by literate Christians and others who were familiar with the source language of the text but were less than complete masters of the 'arabi:ya. Many new scientific terms had to be coined, and the translators were faced with the task of developing an expository prose style for which there was no precedent in older Arabic literature and for which there were no available guides. At the same time as the translation movement was broadening the scope of subject matter that the written language had to cope with, there was increasing political fragmentation and regional autonomy throughout the empire, with the establishment of local dynasties that paid only lip service to central control. This political loosening up reflected itself linguistically in the proliferation of local substrate, dialectal, and nonstandard features in all kinds of writing except in the Islamic sciences proper and those traditional literary genres such as poetry that a few learned Arabs continued to cultivate. By about the end of the third Islamic century (= A.D. 912), however, the traditional 'arabi:ya had ceased to be used in the conversation of good society and in the law courts and colleges, and had ossified into a purely "high literary" idiom. To stick to the rules of ʔiʕra:b in speech was considered a sign of pedantry and affectation. In 955, the geographer al-Muqaddasī lists a large number of differences—phonological, lexical, and grammatical—between the varieties of Arabic used in different parts of the empire even by educated people.[81]

The final result of the political decentralization and exposure to outside linguistic influences occasioned by both the translation movement and the absorption into the Muslim community of disparate non-Arab elements can be appreciated from an examination of the wide range of styles used in Arabic literature from the late eleventh to the mid-thirteenth century. At the most conservative end of the linguistic spectrum, we have, for example, the maqa:ma:t ('assemblies') of the Basran al-Ḥarīrī (d. 1122), a kind of witty literary tour de force in ornate rhymed prose containing many lexical rarities and deliberate archaisms intended to show the erudition of the writer, and understandable only by the literary connoisseurs of the time. Still within the traditionalist camp linguistically, but with a more conversational cadence, we have the poetry of Bahā' al-Dīn Zuhayr (d. 1253). Further down the scale of formality and adherence to the canons of the 'arabi:ya come the anecdotal memoirs of 'Usāma b. Munqidh (d. 1188), a Syrian "officer and gentleman" at the time of the Crusades, in which the tone is often frankly conversational and the language smacks of the Levantine dialect that the author must have spoken. But even in the work of serious geographical and scientific prose writers of the period, such as Yāqūt (d. 1229) and al-Qazwīnī (d. 1283), "offence against grammar (sc. of the 'arabi:ya) is the rule rather than the exception."[82] At the bottom of the scale in observation of

the norms of the 'arabi:ya, we have the popular epic romances of the Banī Hilāl and 'Antar, written for the entertainment of the lower classes in a style that must have closely mirrored the urban vernaculars of the time.

Thus by the middle of the thirteenth century, the linguistic situation was one of considerable fragmentation and complexity. In the urban areas across the Muslim empire, spoken Arabic showed variation along both the horizontal (geographical) and vertical (social) axes. In the countryside, the substrate languages continued to survive, although by now challenged by Arabic following increasing islamicization and, in North Africa, large-scale Bedouin immigration. In contrast to the situation in the towns in the conquered territories, the spoken Arabic of inner Arabia remained structurally close to the original OA dialects of six centuries before, not having been subject to prolonged contact with other languages. As spoken Arabic had developed and become geographically diversified, it had become more and more different from the 'arabi:ya. The point of contact between the two was in certain kinds of popular literature, written in a style that avoided gross solecism and observed the syntactic rules of the 'arabi:ya insofar as this did not conflict with the need to communicate clearly with the intended (plebeian) audience. It is of course impossible to tell how far the linguistic forms used in this kind of writing were a conscious compromise between the norms of a superposed literary tradition and ordinary spoken usage. An answer to this question would require a detailed comparison of the language of this kind of literature with that of surviving nonliterary ephemera of the period that we can be sure would not have been subject to careful monitoring by the writer or later editing by someone else. For most ordinary inhabitants of the empire in the thirteenth century, the 'arabi:ya in its pure form, canonized and reified on the basis of ancient usage by the grammarians, had come to be an exclusively written, almost foreign language, even though venerated by all as the language of revealed scripture. It was the preserve of that minuscule proportion of the population that was literate and that devoted itself to the expounding of religiolegal doctrine or to "high" literature in its various forms.

The sociolinguistic situation described in the previous paragraph remained in its essentials unchanged up to the beginning of the nineteenth century, the point at which the Arabic-speaking world had its first close contacts with European civilization. The intervening period had been one of political, although not linguistic or cultural, subjugation to the Turks. Turkish had been the language of government throughout the Arabic-speaking areas of the Ottoman empire, but there was never any question of it replacing Arabic as the language of Islam, even for the Turks themselves, nor any attempt to Turkicize the ordinary people. Turkish linguistic influence on spoken Arabic was limited to lexical bor-

rowing, much of it in the military sphere, but also with sizable borrowings in the spoken terminology of agriculture, material culture, food, the household, and local administration.[83] There was no Turkish influence on the frozen norms of the 'arabi:ya, which, by the eighteenth and early nineteenth centuries, was used solely for increasingly sterile imitations of old literary genres.

The beginnings of a renaissance of literary Arabic in both Egypt and the Fertile Crescent came as a consequence of cultural contacts with the West. Despite the oppressive rule of the Ottomans, the Christians of Syria–Lebanon had had contacts with Europe since the seventeenth century, and these continued unbroken into the nineteenth because of the interests of the Papacy in extending its interests in the area. Originally for reasons of proselytization, Christian clerics and laymen established colleges and missions in the secluded mountain valleys of Lebanon and laid the foundation of the revival of literary Arabic as an instrument of general cultural transmission.[84] The most important single political event, however, was in Egypt. In 1798 Napoleon launched an expedition against Egypt and defeated the Mameluke Turks at Alexandria. The short occupation that followed—it lasted only two years until the French army was forced out in 1801—left a legacy of local unrest that led to momentous internal political changes. A young Albanian officer, Muḥammad 'Alī, who had been sent by the Turkish sultan at the head of an army to fight the French, was given the task of restoring law and order after their departure. He was successful and was confirmed as Pasha of Egypt by the sultan in 1807. He and his descendants ruled Egypt for the next 145 years until the military coup that brought General Muhammad Neguib to power on 23 July 1952.

Muḥammad 'Alī and his successors in government pursued a deliberate policy of cultural and educational rapprochement with Europe, one of the effects of which was to underline the degree to which Egypt had become technologically backward during the long period of Turkish domination. Missions were sent to France and elsewhere in Europe for training in modern administration, law, economics, medicine, science, diplomacy, etc. The training the members of these missions received using French, English, and German, and the difficulties they experienced on their return in attempting to apply what they had learned by means of a language that lacked both the necessary technical vocabulary and a living tradition of systematic word coinage impressed on the Egyptian administration the need for action. In the short term this meant that professional training inside Egypt would have to be done in foreign languages, but in the long term the Arabic language would itself have to be modernized to cope with the new demands that would be made on it. A School of Languages was estab-

lished in Cairo in 1836 under the direction of Rifā'a Rāfi' al-Ṭahṭāwī. The objectives of this institution, and other schools like it, were to prepare students for the professions and to train government officials and translators. A separate bureau of translation was added to the school in 1841 in which a great many works on geography, history, military matters, philosophy, and social issues were translated into Arabic, mainly from French.[85] Al-Ṭahṭāwī also became editor of the official Egyptian government gazette, al-Waqā'i' al-Miṣrīya, which later developed into the first Egyptian national newspaper. The Arabic printing press had been introduced by the French some thirty years earlier, and the effect of this was to increase popular access to written material of all kinds. De facto, the written Arabic language gradually started to cease to be the exclusive preserve of the religiolegal establishment and literateurs and to be used for a wider variety of purposes, educational and everyday, although there was not yet any attempt to modernize it systematically.

As the nineteenth century wore on, Arabic came to occupy an ever more important place in the thinking of Islamic reformist and nationalist political movements in Egypt and the Fertile Crescent. Its symbolic importance as the factor par excellence that united all Arabs and Muslims was redoubled after the collapse of the Ottoman empire and its dismemberment after the First World War. The significance of Arabic as a unifying symbol is summed up in this passage from Rashīd Riḍa (1865–1935), writing in 1922:

> One of the religious and social reforms of Islam was to bring about linguistic unity, by making its common language that of all the peoples who adhered to it. The religion preserved the language and the language preserved the religion. But for Islam the Arabic language would have changed like others, and as it had itself changed previously. But for Arabic, the different interpretations of Islam would have grown apart from each other, and it would have split into a number of faiths. . . . Thus the Arabic language is not the private property of the descendants of Qahtan, it is the language of all Muslims.[86]

In 1863, Arabic had been declared the sole official language of government in Egypt (previously this position had been shared with Turkish). But the use of Arabic in many arenas of public life was the exception rather than the rule: English, French, German, Greek, and other languages continued to be used in all the professions. There was no possibility of Arabic-medium training in Egypt because of the complete lack of modern educational institutions and Arabic-speaking teachers, and the poverty of modern Arabic terminology in these sub-

jects. After the British occupied Egypt in 1882, the cause of Arabic suffered another reverse, when English was declared the sole official language in 1898. But the pressure for the restitution of Arabic to public life continued unabated through the early part of the twentieth century.

One of the ways in which this pressure expressed itself in both Egypt and Syria was in attempts to establish an Arabic language academy to conduct research on linguistic problems and the development of technical terms. Starting in 1892, there had been a number of short-lived unsuccessful attempts in Egypt,[87] but in 1919 an academy was founded in Damascus, followed by another in Cairo in 1932. The Damascus Academy's stated principles are interesting in themselves, because they reflect the academicians' perceptions of where the roots of the problem lay: (i) the preservation of the purity of the language, (ii) making Arabic self-sufficient so as to meet the requirements of the arts and sciences, and (iii) rendering Arabic a suitable means of communication in the modern world.[88] The first principle is a defense against the encroachment of spoken Arabic on the one hand, and foreign languages on the other. It recognizes the *'arabi:ya* as the only "true" variety of Arabic and hence implicitly rejects the spoken dialects as corrupt. Nor should foreign elements be allowed to intrude into the regenerated *'arabi:ya*: the second principle enshrines the concept of self-sufficiency in generating new vocabulary. This meant that neologisms coined to fill terminological gaps had to conform in their morphology to roots and patterns already attested in the *'arabi:ya* through the application of the principle of *qiya:s* ('analogy'). Except in extreme circumstances, mere loan translation or, even worse, the transliteration of foreign terms into the Arabic script was to be rejected. The third principle hints at an admission that the *'arabi:ya* was not only deficient in terminology, but was also not a sufficiently flexible or widely understood medium to be usable for everyday purposes, but that somehow this needed to be remedied.

These desiderata strike one as at once prescriptive, grounded in conservative philological tradition, and based on an assumption that lack of technical terminology was the root of the problem. Indeed, much of the early work of the academies, and of others like them that sprang up later in Iraq and Jordan, was devoted to coining long lists of equivalents for words such as "microscope," "loudspeaker," "elevator," and "propeller."[89] Although some of these neologisms eventually came to be accepted, many did not, either because a transliterated foreign word had already gained such currency among concerned users that it could not be replaced or because the newly coined term, derived from an archaic meaning of an Arabic root, was simply semantically opaque to the intended users.[90] Worse, the proliferation of language academies, a consequence

of political fragmentation, often led to the coining of several terms for the same referent, making communication in Arabic between specialists from different Arab states even more difficult.[91] Small wonder that many continued to use English or French.

The question of how the *'arabi:ya* could be modernized so that it could become a means of normal communication for ordinary people was an even thornier issue and has to be seen in the context of the prevailing social and political climate of the early decades of this century. The background to the debate was the very high rate of illiteracy throughout the Arabic-speaking world and the political subjugation of the Arabs to western powers. There were basically two positions advocated: reduction of morphological and syntactic complexity, so that the simplified *'arabi:ya* that resulted resembled more closely the natively spoken dialects;[92] and promotion of the use of dialect instead of the *'arabi:ya* in secular contexts.[93] Both of these proposals can be seen as motivated by the desire to create a greater degree of social cohesion within the then-future nation-states of the Arab world by increasing the participation of the population at large in national life, which depended in many fields on literacy in the *'arabi:ya*. But neither ever had very much of a chance of being accepted, even if they had ever developed beyond vague statements of intent,[94] because both went against the grain of cultural history. In a politically fragmented Arab world, in which the western concept of the nation-state had in any case always been regarded with ambivalence if not suspicion, the Arabic language, along with Islam, the religion that was revealed through it, was *the* politically and culturally cohesive force. From this perspective, "Arabic" could only mean that which was common to all Arabs wherever they lived—the *'arabi:ya*. Because the spoken Arabic dialects differ from one area of the Arab world to another, attempts to argue the advantages of elevating them to the status of written languages could be construed only as playing into the hands of those outsiders who sought to divide and rule the Arabs politically and would have the effect of alienating all of them from their common cultural and religious patrimony.[95] Simplifying the grammar of the *'arabi:ya* might be less politically damaging in a pan-Arab context but would have the same culturally alienating effect.

From a western perspective, these objections to language reform smack of elitism. After all, the vast majority (in places up to 90 percent) of the rural and lower-class urban population in all parts of the Arab world was illiterate until thirty or forty years ago, so it was in effect culturally disinherited anyway. Such people would have had nothing to lose from a thoroughgoing reform. This is not to say that the maintenance of the *'arabi:ya* as the sole acceptable and legitimate form of written Arabic was a deliberate ploy to exclude the popula-

tion at large from participation in political and cultural life. It was rather some-
thing that followed almost inevitably from a combination of factors—the ir-
relevance of reading and writing to the way many, or even most, Arabs lived
until recently (on the land); the lack of a secular paradigm in which education
in general, and language education and planning in particular, could have been
elaborated; and the interests of the religious authorities in maintaining the hold
of Islam, and its vehicle, the ʿarabiːya over the population. Until well into this
century, all of these factors conspired to produce a general inertia.

1.5 The contemporary linguistic situation

The "democratization" of Arabic did in fact happen during the second half
of the twentieth century, although not in ways that the language academies or
other concerned instruments of government planned. There has been a grad-
ual but palpable narrowing of the gap between spoken Arabic and the ʿarabiːya
in its contemporary form, MSA. This has come about not through the imposi-
tion of an artificial norm from above, but rather through natural symbiosis: the
lexis and phraseology of MSA penetrate spoken Arabic, and, in somewhat less
obtrusive ways, syntactic structures common to most spoken dialects are recast
into a superficially MSA form in written Arabic.[96] The cause of this has been
massively increased education and literacy among all sections of the population,
the political impetus for which was the strongly populist policies of the social-
ist regimes in Egypt and the Levant in the 1950s and 1960s and in postcolo-
nial North Africa, underwritten by the improved economic situation in which
almost all Arab countries have found themselves since the mid-1950s.

Other changes have been externally motivated. Much foreign lexical, phrase-
ological, and even syntactic influence has been exerted on MSA in recent years
as a result of loan translation from European languages.[97] More than anything,
this is a consequence of the dominance of English, and to a lesser extent French,
in the international media. Arabic newspapers in particular are full of rapidly,
and often very literally, translated versions of press agency reports with what
Joshua Blau calls "Standard Average European" phrase structure.[98] The ad hoc
calques thus formed find their way into everyday use. A tremendous amount of
MSA stock phraseology in economics, politics, and popular science and tech-
nology seems to have arisen in this way, for example, al-ʿumla al-ṣaʿba 'hard cur-
rency', suyuːla naqdiːya 'cash flow', taʿwiːm al-junayh 'floating of the pound', al-
dibluːmaːsiːya al-makuːkiːya 'shuttle diplomacy', al-ʿadd al-tanaːzuliːy 'countdown',
ʿamaliːyat ziraːʿat al-qalb 'heart transplant operation', buḥayrat al-nifṭ al-zaliqa 'oil-

slick', as well as cliched, all-purpose metaphors, such as ḥajar al-zaːwiya 'cor-
nerstone', qimmat jabal al-θalj 'tip of the iceberg', tajmiːd 'freezing (e.g., of po-
litical relations, assets)', al-manaːx al-ʿaːtifiːy 'emotional climate', etc.

What of regional differences in MSA? Although, syntactically speaking, MSA
is relatively homogeneous across the whole of the Arabic-speaking world, there
are significant and systematic differences in vocabulary that differentiate the
Maghreb countries—principally Tunisia, Algeria, and Morocco—from those of
the Mashreq, or eastern Arab world. Some examples from Tunisian official signs
and hoardings, which I noticed on a recent (2001) visit to that country, will il-
lustrate the point. A modern European-style 'hotel' is typically nazl, rather than
eastern Arabic funduq—in the west, this latter term is not used to refer to the
modern hotel but variously designates certain premodern phenomena: a type of
old-fashioned caravanserai, or an entrepot for the storage of goods, or a kind of
guild house where artisans practicing a particular craft lived and worked. In
Tunisia, a small 'shop' is maġaːza (probably via a reborrowing of French maga-
sin, itself originally a borrowing into French from Arabic), which has now de-
veloped a further Tunisian calque in maġaːza rafiːʿa 'supermarket' (lit. 'superior
shop'). This is normally termed suːbarmarkit in the Arab east. 'Car rental' in
Tunisia is always kiraːʔ al-sayyaːraːt, compared with ʔiːjaːr al-sayyaːraːt in Cairo or
Damascus. Both words for 'rent' are equally acceptable as normative MSA and
appear in modern dictionaries. Shop signs advertising the sale of fruit in Tunisia
term it ġilaːl, whereas in the east the term is fawaːkih. There are hundreds of such
differences in the terminology of written Arabic for everyday material artifacts,
objects, and activities in which the western Arab world prefers one MSA syn-
onym or near synonym and the eastern another. These often reflect underlying
east–west dialect differences or, as in the case of maġaːza versus suːbarmarkit,
the different foreign languages with which each area has historically had most
contact. There are also differences in the "officialese" of government bureau-
cracy—for example, an administrative 'region' is jiha (adjective jihawiːy) in
Tunisia, whereas in the east jiha is most commonly used in the more general
sense of 'direction, perspective, point of view'; in the east, 'region' in an ad-
ministrative sense is minṭaqa. Potentially, and sometimes in practice, these dif-
ferences can lead to misunderstandings. In Tunisia, at least, 'the (natural) envi-
ronment' is termed al-muḥiːṭ, a word that in the eastern Arab world is used only
in the context of the natural world to mean 'ocean' (the semantic link between
the two is that al-muḥiːṭ literally means 'that which surrounds'); easterners use al-
biːʔa to denote 'the environment'. The Maghrebi term ḥariːf 'guest', pl. ḥurafaːʔ,
as in a notice to hotel guests who are addressed as ʔayyuhaː l-ḥurafaːʔ al-kiraːm

'Dear guests . . . ' sounds to someone from the Mashreq like an archaic word meaning 'artisan'. And in a Maghrebi taxi, the sign that specifies the *'adad al-biqa:'*, that is, the 'number of seats', causes puzzlement to visitors from the east, who would expect *'adad al-rukka:b* 'number of passengers'—for them *biqa:'* principally means 'spots, blots, or patches' rather than being the normal word for 'place, seat'.

In the spoken domain, the results of the spread of education and the exposure of the population to the broadcasting media are varieties of Arabic intermediate between "pure" MSA and "pure" dialect, in which there can be a greater or lesser mixture of MSA and dialectal elements, depending on the speaker's (or writer's) perception of the formality of the context. At the formal end of the spectrum, speakers select a basically MSA morphosyntactic and lexical base but modify it morphophonologically in the direction of their dialect; at the informal end, a fundamentally dialectal base is "classicized" by the insertion of MSA phraseology, lexical items, and the associated "prestige" pronunciation. In some circumstances speakers switch between using a "colloquialized" MSA and a "standardized" colloquial as they perceive the demands of the speech context to change. There have been a number of attempts to describe the varieties of Arabic produced in terms of a spectrum from the purest MSA through intermediate varieties to uneducated plain colloquial,[99] generated by the complementary stylistic tendencies of "leveling" (the elimination of very localized dialectal features in favour of more regionally general ones) and "standardizing" (the replacement of dialectal features by standard ones beyond the level demanded by the need to ensure clear communication), although so far none has been based on a sufficiently large data base to enable a complete "grammar" of the permissible combinatorial possibilities of MSA and dialectal elements to be written.

The interpenetration of MSA and the dialects is the linguistic concomitant of the spread of secular government-sponsored education, reinforced by the output of the ubiquitous news and entertainment media. Education has meant that a large proportion of Arab children now have a long and formative exposure to MSA at school. All school materials in all subjects are written in MSA, and a large number of hours of the curriculum are specifically devoted to mastering its rules and to reading its literature and that of its forebear CLA.[100] The media also provide a constant diet of MSA to the population at large in news broadcasts, religious programs, documentaries, and historical dramas. Without the speakers being aware of it, MSA words and phrases and pronunciations seep into their speech and in the long term may even replace some dialectal equivalents completely. But language contact is a two-way street: the fact that education has made MSA the "property" of a much greater proportion of the popu-

lation than was true even one generation ago means that MSA is itself now much more open to influence from the dialect than hitherto because of the more extensive concurrent use of the two varieties. This has already blurred, and will no doubt continue to blur, the distinctions between them.[101]

The contemporary sociolinguistic situation in the Arab world is thus a complex one, although perhaps no more complex than the situation at earlier but less well-documented periods of its history. The concept of Arabic as a "diglossic" language,[102] if it was ever accurate, is now an oversimplification: the behavior of most Arabic speakers, educated or not, is rather one of constant style shifting along a cline at opposite ends of which are "pure" MSA and the "pure" regional dialect, more accurately conceived of as idealized constructs than real entities. Most communication apart from the most "frozen," written as well as spoken, is conducted in a form of Arabic somewhere intermediate between these two ideals but is governed by rules nonetheless, even if we cannot yet capture the full complexity of the rules that control the combining and hybridizing of the two.[103] In many regions of the Arab world, the choice of what kind of Arabic to use on what kind of occasion is not confined to a one-dimensional cline. As a result of population movement, voluntary or forced, generational differences can develop within a community or even a single extended family. In Ramallah, in the occupied West Bank, for example, there are many families who have recently migrated from rural areas. The older nonliterate generation tends to know and use only a "ruralite" dialect, whereas the younger generation not only acquires a knowledge of MSA at school, but also an awareness of the "urbanite" dialect, which has local prestige. "Ruralite" features tend to be used by these younger speakers to their family elders, while "urbanite" features are preferred in most informal public contexts in which "ruralite" features are likely to be stigmatized or at best regarded as quaint; in more formal contexts of speech, MSA features variably replace "urbanite" features.[104]

The overall picture is thus one of kaleidoscopic variation. Within each modern Arab nation-state, the dialect of the capital city will usually carry some prestige and act as a kind of local "dialect standard,"[105] at least in everyday but nondomestic contexts within the national domain—for example, peasant farmers visiting government offices to sort out some work-related problem. In more formal but still "national" contexts, the influence of MSA will make itself more or less felt, depending on what is being talked about, to whom, and in what circumstances, such as an unscripted television discussion on a housing crisis in which all the participants are of the same nationality. In a supranational speech context, national dialects may still be used, although with some "leveling" and "standardizing." This kind of variation occurs in such contexts as a pan-Arab discussion on educational

cooperation. Again, depending on contextual factors (including the subject itself, which may have more or less strong MSA associations), the "leveling" and "standardizing" processes may elevate the language level to the point where most local features are zeroed out and replaced by standard ones. Finally, in speech contexts that are seen as outside any purely national framework (e.g., news bulletins, religious programmes par excellence) only "spoken MSA" is appropriate.[106]

In writing Arabic, there is less room for variation: basic grammar, morphology, and lexis (with some geographical exceptions noted above) are in theory the same for all. However, in practice there *is* variation, both natural and contrived. In some forms of nonfolkloric narratives and especially in drama, a "written colloquial" may be put in the mouths of characters or actors where the creation of a naturalistic atmosphere is being aimed at; and in the evocation of particular social milieu, it may be impossible, even were it felt desirable, to avoid the use of nonstandard terms. These points apply particularly to certain erstwhile leftist Egyptian writers such as Yusuf Idris.[107] Other kinds of marginal writing in which the colloquial is normally rendered into script are children's comics and political cartoons. It can be inferred from this that "written colloquial Arabic" conventionally conveys multiple associations of the "domestic," "homely," "amusing," and "nonserious." The same associations attach to the "colloquialized written Arabic" that occurs in certain kinds of informal writing not for public consumption in which a nonstandard written Arabic is produced less consciously. Examples are handwritten personal letters to friends, draft typescripts, memoranda, and other kinds of informal documents and ephemera.[108] Here there may be much less care taken over excising the influence of the colloquial, or a "colloquialized standard" may even be deliberately aimed at for affective reasons, as in letters to friends. It will be remembered that it is precisely this kind of informal material from the early medieval period that has enabled us to infer what the structure of spoken Middle Arabic was like.

Notes

1. Moscati et al. 1964, 15–16.
2. The three key "synthetic" Proto-Semitic features that have reflexes in CLA are: (i) a set of final short vowel case inflections that are suffixed to the noun, e.g., *?al-malik-u* (nom.), *?al-malik-a* (acc.), *?al-malik-i* (gen.) 'the king'; (ii) the suffixation of *-n* (so-called nunation, in Arabic *tanwi:n*) to designate indefiniteness in most types of noun, e.g., *malik-un* (nom.), *malik-an* (acc.), *malik-in* (gen.) 'a king'; (iii) a set of short vowels suffixed to imperfect verbs to indicate mood, e.g., *yaktub-u* (indic.) 'he writes, is writing', *yaktub-a* (subj.) '(that) he may write', *yaktub* (juss.) 'let him write'. This marking of final short vowels to designate case in the noun and mood in the verb is called by the Arab grammarians *?i'ra:b*. Only in two of the

most ancient Semitic languages, Akkadian and Ugaritic, is (i) attested in as complete a form as it is in CLA, whereas in no other do we find a system of equivalent completeness and consistency for (ii) and (iii). In the Semitic languages contemporaneous with sixth-/seventh-century CLA for which we have data, e.g. Aramaic, these features have all been lost.
3. For example, see Moscati et al. (1964, 94–96, 134–36) on the case and mood endings in CLA as evidence of its antiquity. For a contrary view, see Wansbrough 1977, 106ff.
4. See Zwettler 1978.
5. Emeneau 1964, 330–40.
6. Rabin 1955, 21.
7. Zwettler 1978, passim, but especially chapter 3.
8. Fück 1955, 3.
9. Versteegh 1984, 4–5.
10. Zwettler 1978, 132–33.
11. Ibid., 134.
12. Ibid., 110–11, 113, 115–16.
13. Ibid., 112.
14. Rabin 1951, 3–4.
15. Old Arabic dialect differences ascribed to various tribes, as they are reported in Arab grammarians' works beginning in the eighth/ninth centuries A.D., have been collected and classified in Kofler 1940–42. Zwettler (1978, 103–10) argues, via an analysis of irregularities in the patterns of vocalic assonance at rhyme ends and the skewed incidence of final consonant clusters in the same position, that, in all probability, the desinential inflection system in the noun was breaking down in the spoken dialects at the time of the composition of the poetry.
16. See Sowayan 1985, especially 147ff., in which it is argued that modern folkloric poetry in Arabia is the direct lineal descendant of the classical oral-formulaic tradition.
17. "Artifact" is not of course a description that would ever be applied to the Koran by an orthodox Muslim, to whom it is the literal word of God, revealed through his Messenger.
18. The epithet *?ummi:y*, applied to Muḥammad in the Koran itself, in modern Arabic simply signifies the inability to read and write. Its original, and Koranic meaning, however (see Noeldeke 1909–19, 14) seems to have been "one unfamiliar with the ancient scriptures" (i.e., a Gentile) and only by extension one who could not read.
19. Zwettler 1978, 158.
20. Arberry 1957, 250.
21. According to tradition, the first Arabian ode was composed in about A.D. 500, but prototypes were certainly in existence well before this date (see Nicholson 1969, chapter III).
22. It was believed that poets were possessed by spirits (*jinn*) and had access to the occult. The word for 'poet' *ša:'ir* means literally "knower."
23. Similarities between the language of the early Meccan *su:ras* and that of the oracular utterances of soothsayers, and the compositions of the poets seem to have led Muḥammad's detractors to accuse him of being one of either or both, to judge by several denials in the Koran itself, such as Sura XXXVI, 69: "We have not taught him [i.e., Muḥammad] poetry; it is improper for him. It [i.e., the Koran] is only a reminder and a

clear recitation." Given that the textual structure of the early *su:ras* and the poetry are so different, one can only assume that these accusations were based on the fact that the *linguistic system* used in the poetry and the Koran was the same—the inflective *'arabi:ya*—and different from that of everyday Arabic speech (see Zwettler 1978, 157–61).

24. Rabin 1951, 3–4.

25. See Fück 1955, 3.

26. Vollers 1906.

27. Short vowels, wherever they occur in an Arabic word, are not normally written, hence the difficulty in interpreting how a consonantal skeleton would have been pronounced in material that , unlike the poetry, is metrically irregular.

28. See, e.g., among many others, Johnstone 1961, 264–66, Matar 1981, 233–39, Holes 1983a, 27–28.

29. See, e.g., Johnstone's computations based on his Dōsiri data (1961, 264).

30. Contrast the data for Dōsiri Arabic given by Johnstone with that for nearby Bahraini Arabic given by Holes in the references cited in note 28.

31. Johnstone 1961, 264.

32. See, e.g., the lengthy discussion of *tanwi:n* in the modern dialects in Blau 1981a, 167–222.

33. Cf. the comments of Corriente 1976, 87–91, Zwettler, 1978, 122.

34. In Yemen, another Semitic but non-Arabic language, Himyaritic, was still in use alongside Arabic. See Rabin 1951, 25ff.

35. Donner 1981, 92–94.

36. Versteegh 1997, 33.

37. See Versteegh 1984, 66–67.

38. See Donner 1981, 167–73 for more detail on the social composition of Iraq during this period.

39. 'Umar 1970, 12–13.

40. Versteegh 1984. See the generally critical reviews by Holes in *BO* 43, no. 1/2 (1986): 218–22; Goodman in *JPCL* 1 no. 1 (1986): 165–70; Hopkins in *ZAL* 18 (1988): 98–99.

41. Versteegh 1984, 2.

42. A typical example is quoted in Fück 1955, 8.

43. Hopkins 1984.

44. Hopkins 1984, xlvi. Hopkins (personal communication) now feels that "Middle Arabic" as the term is being used here is confusing, because it fails to distinguish that mix of Classical, Neo-Arabic, and pseudocorrect elements typical of certain types of medieval written text, and the Neo-Arabic component of such texts.

45. Geoffrey Khan (personal communication) states that early dated Judeo-Arabic material from the Cairo Geniza written in voweled Hebrew script has allowed him to reconstruct a diachronic grammar of Cairene Arabic that traces its gradual evolution over several centuries. Publication of this grammar is forthcoming.

46. See Versteegh (1984, 117–19) for references on this and other modern Arabic pidgins and creoles in Africa.

47. Pasch and Thelwall 1986.

48. The only report of a medieval Arabic pidgin/creole was discovered in Egypt in 1982 in a manuscript written by the Hispano-Arab geographer al-Bakrī in the mid-eleventh cen-

tury and relating to the speech of the black population in the town of Maridi in Mauritania, N.W. Africa (Thomason and Elgibali 1986, discussed by Holm 1988 [2:568–74]).

49. See Blau 1981a and 1988 passim.

50. Ferguson 1959.

51. In Egypt, for example, Coptic (and according to some, Greek ['Umar 1970: 30–31]) continued to be used as the sole written language of government and administration until A.D. 705, some sixty-five years after the conquest of Egypt was completed. It is only at this point that bilingual Coptic/Arabic records and protocols begin to appear alongside monolingual Coptic ones.

52. The islamicization of the conquered realms, especially the peasantry, was also a gradual process: for example, most of the Egyptian rural population remained non-Muslim (and almost certainly non-Arabic speaking) until at least the middle of the eighth century, some one hundred years after the conquests (see 'Umar 1970, 43).

53. For example, between about A.D. 760 and 870 there was a continuous flow of tribal migrations into Egypt from the peninsula for a variety of political and economic reasons. Migrations by communities from no less than thirty-three different tribes have been identified as occurring during this period ('Umar 1970, 40).

54. The details of what kind of Arabic was brought by the conquering tribesmen and the early history of its implantation in the conquered territories are controversial. See Miller 1986 for a summary of the arguments.

55. 'Umar 1970, 30–31.

56. Ibid., 33–34.

57. *EI* 5:92, article '*Kibṭ*'.

58. Donner 1981, 221.

59. Versteegh 1984, 64.

60. Donner 1981, 245–50.

61. Blau 1981a, 21–22.

62. Violet 1901–2.

63. Blau 1988, 288–90 demonstrates Aramaic substrate influence on the morphology, syntax, and lexicon of south Palestinian Christian Arabic of the first millennium.

64. Garbell 1958.

65. Donner 1981, 222.

66. Versteegh 1984, 74–75.

67. Donner 1981, 229–45 passim.

68. Abun-Nasr 1975, 71ff.

69. Ibid., 85–86.

70. Wansbrough 1977, 89.

71. There are a few small Arabic-speaking *Sprachinseln* [speech islands] that are an exception to this: in Uzbekistan, described in Fischer 1961, and in Afghanistan, described in Sirat 1973.

72. See Hopkins 1984.

73. Literate medieval Muslims, in all but the most mundane types of writing, seem to have been more influenced than non-Muslims by the norms of the *'arabi:ya* on account of its by then fully elaborated, disseminated, and specifically Muslim religiocultural signifi-

cance. See, for example, Blau 1981a, 19–50 on the significance and use of different "levels" of Judeo-Arabic (Arabic written in the Hebrew script) in the medieval period.

74. Ibn Khaldūn 1986, 3:321–22.

75. See Owens 1988 for a general overview.

76. The notions of "correctness" overlap with those of "eloquence" and "purity" in Arabic grammatical writing. Sībawayhi makes use interchangeably of a variety of positive ascriptions when approving of the sources of particular linguistic forms, e.g., *turḍa: ʿarabi:yatuhum* '(those) whose Arabic is satisfactory', *al-ʿarab al-mawθu:q bi ʿarabi:yatihim* 'Bedouin whose Arabic can be relied on', and *fuṣaḥa:ʔ al-ʿarab* 'the eloquent among the Bedouin'. See Fück 1955, 45.

77. Haarman 1974.

78. Carter 1983, especially 77ff.

79. Ibid., 71.

80. Versteegh 1983, 157.

81. Fück 1955, 163ff.

82. *EI* 1:571, article "*Arabīya*."

83. For example, see the comments in the introduction of Hinds and Badawi 1986 on the foreign elements in Egyptian Spoken Arabic. A comprehensive list of the specifically Ottoman Turkish lexemes in the same dialect can be consulted in Prokosch 1983.

84. Hourani 1967, 55–56; Suleiman 2003, 81–82.

85. Hourani 1967, 71.

86. Rashīd Riḍā *Al-Khilāfa* (*The Caliphate*) Cairo, 1341 A.H. (1922–23), quoted in Hourani 1967, 300.

87. Chejne 1969, 104.

88. Hamzaoui 1965, 9.

89. Stetkevych 1970, 17.

90. Ibid., 31–33.

91. See, for example, the inconsistencies in the Arabic geological terminology proposed by the language academies in Chalabi 1984, 287–88.

92. The proposals of Anis Furayḥa 1955, 183–96 for the adoption of a simplified written language based on the speech of the educated are the classic statement of this position. Furayḥa even goes as far as to propose the use of the Roman alphabet.

93. An Egyptian perspective on how the movement for the use of colloquial Arabic in Egypt was first promoted by the British and then taken up by a group of Egyptian reformers in the late nineteenth and early twentieth century can be found in Saʿid 1964, 43–71 and 75–149. The attitude of the Arab language academies to the idea of adopting the colloquial has been universally hostile. Fairly typical is the comment of academician Muḥammad Riḍā al-Shabībī in an article titled "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 (*dasi:sa istiʿma:ri:ya*) which has only been accepted by a few people terrified by the call of Islam." The identification of the Classical language with religion and of the colloquial with a divisiveness that can be exploited by outsiders has wide popular currency throughout the Arab world.

94. Stetkevych 1970, 83.

95. A typical example of views of this kind is that of the Lebanese leftist physician G. Hanna, cited in Chejne 1965, 465.

96. Blau 1973, 193–200 and 1976, 162–71.

97. Blau 1981b, 60–141 passim.

98. This term was originally applied in a Semitic context with reference to Israeli Hebrew. See Blanc 1957, 400.

99. Blanc 1960 proposes five quasidiscrete levels of Arabic on the basis of a short sample of interdialectal conversation; Badawi 1973, 89ff. also suggests five levels of Arabic within Cairo, basing his analysis on media output. See chapter 9 for a detailed discussion of the descriptive adequacy of such models.

100. Al-Toma 1974, 291–92. For the rather different situation and history of educational language use in Algeria, North Africa's most populous country, see Holt 1994; and Boucherit 2002.

101. Mitchell 1986, 20–27. In the written sphere, Rosenbaum 2000 describes the phenomenon of what he terms *fuṣḥa:mmiyya*, a style found in modern Egyptian short stories, novels, and especially magazine articles, in which there is frequent and seemingly unpredictable switching back and forth between MSA and Cairene colloquial Arabic. See 9.5.3.

102. Ferguson 1971.

103. Mitchell 1986, 9–10. See chapter 9.

104. Cadora 1970. A similar "switching" phenomenon is reported in Grotzfeld 1983, 90–92 for rural Lebanon. Cross-dialectal switching in a sectarian milieu in Bahrain is discussed in Holes 1983b, 1987 passim.

105. See Holes 1995b for data on Jordan, Iraq, and Bahrain; Haeri 1996 for Egypt.

106. But see now Al-Batal 2002.

107. Vial 1983 provides a representative glossary of common "Egyptianisms" in contemporary Egyptian Arabic literature.

108. Meiseles 1979.

2

Phonology
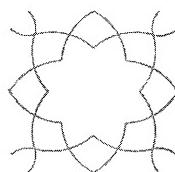
The organization of this chapter, and of those on morphology and syntax, is intended to facilitate comparison of the structure of MSA and dialectal Arabic as they are used today. In reality, MSA in its written or spoken form and the natively spoken dialects are frequently employed in tandem in many social and occupational contexts, and the result is mutual influence at every level of linguistic structure. This symbiotic relationship between MSA and the dialects, as it expresses itself in phonology, will be exemplified in 2.3 below in some detail, following a description of some salient aspects of the phonological differences between the two in 2.1 and 2.2. The relationship between the phonology of Arabic and its script is described in 2.4.

The problem in describing the phonology of MSA is that MSA is not natively spoken by any group of Arabs, so no contemporary regional or social grouping can reasonably claim that its habits of performance represent a model of what is "correct." In the native grammatical tradition, *faṣaːḥa* (= 'purity' and, by extension, 'correctness'), whether in phonology or any other aspect of the language, ultimately derives from historical precedent: that is, the CLA as elaborated by the grammarians. Although we have a clear picture of what Arabic morphology, syntax, and vocabulary were like from an early period from the copious textual material that survives, we have little or no information on many aspects of how it was pronounced. The doyen of Arabic grammarians, Sībawayhi, devotes a short section of his voluminous *Kitāb*

(*Book*) to describing the place and manner of articulation of the consonants and the long vowels, and also deals with topics such as consonant assimilation that, while nonphonemic, were nonetheless important to correct Koranic recitation; but on many other nonphonemic matters and suprasegmental features—positional variation in the quality of the short vowels, word stress, juncture, and intonation, for example—there is total silence.[1] Today, where a "high-flown" spoken MSA is required by context—such as television dramatizations of momentous events in Arab-Islamic history—the actors' performance usually aims at reflecting the age-old pronunciation conventions of Koranic and Classical poetry recitation.[2] But in less culturally loaded contexts that also demand scripted speech—media news bulletins, for instance—there is a variable tendency for the performance of MSA to reflect the phonology of the speaker's dialect. In unscripted contexts that require speakers to sustain an educated style of spoken Arabic, their dialectal background is likely to become even more obvious.[3]

Arabic has a relatively large consonantal inventory and a small vocalic one. In MSA, there are twenty-eight distinct consonant phonemes but only three short vowels, plus a phonemic quality of length, applicable equally to consonants and vowels. Two of the twenty-eight consonants, /w/ and /y/, are best considered semivowels rather than full consonants, because at the phonological level (reflected in the writing system) they are distinguished from their corresponding short vowels /u/ and /i/ only on grounds of syllable structure rules: /w/ and /y/ may be both syllable-initial and syllable-final, whereas /u/ and /i/ may not be syllable-initial.

Perhaps the most salient acoustic impression created by spoken Arabic is its "guttural" quality. There are no less than eight consonantal phonemes articulated in the velar and postvelar region of the vocal tract (compared to three in English), including a uvular plosive and two pairs of voiced–voiceless fricatives, one of which is velar and the other pharyngeal. The other characteristically Arabic phonological feature, evident from the chart below, is four pairs of plain/emphatic (or "velarized") dental-alveolar plosives and fricatives. The acoustic impression created by the emphatics is one of hollow resonance as opposed to the sharper sound of the corresponding plain consonants. Three coarticulations characterize the emphatics: the retraction of the tongue tip from the position of the plain consonant, the raising of the back part of the tongue toward the soft palate, and the constriction of the top of the pharynx caused by the retraction of the tongue root. Emphatic consonants have phonetic effects on contiguous segments to the left and right: nonemphatic consonants with which the emphatic is in contiguity become emphatized, and vowels are lowered and backed. In some cases the whole word may become emphatized.[4] This fact, along with

Table 2.1   Consonantal Inventory of Spoken MSA

Place of articulation		Plosive		Fricative		Affricate	Liquid	Nasal
labial			*b*		*w*			*m*
labiodental				*f*				
dental	plain	*t*	*d*	*s*	*z*		*l*	
	emphatic	*ṭ*	*ḍ*	*ṣ*	*(ẓ)*			
interdental	plain			*θ*	*ð*			
	emphatic				*ḏ̣*			
alveolar				*š*	*(ž)*	*j*	*r*	*n*
palatal					*y*			
velar		*k*	*(g)*	*x*	*ġ*			
uvular		*q*						
pharyngeal				*ḥ*	*ʻ*			
glottal		*ʔ*		*h*				

Manner of articulation spans the Plosive, Fricative, Affricate, Liquid, Nasal columns.

the observation that four nonemphatics—/r/, /x/, /ġ/, and /q/—also exercise a backing effect on contiguous vowels similar to that of the emphatics, has led some linguists to propose that emphasis is in fact not the property of the four individual consonants that are conventionally treated (and graphologically represented) as such, but rather a prosodic or suprasegmental feature[5] whose minimal domain is a consonant and an adjacent vowel.[6] The phonological scope of emphasis spread differs from one dialect to another, however,[7] and hence the scope of its realization in "spoken MSA" will, like other normally unmarked graphological features (stress, intonation, and short vowel quality) tend to reflect the phonological facts of the speaker's dialect in the absence of any accepted MSA norm.

2.1 The phonology of MSA

As will be apparent in 2.1.1., the pronunciation conventions, even for formal MSA, are to some degree dependent on the region and usually reflect the background dialect of the speaker. Because little is known about how ancient forms of spoken Arabic (generically referred to as OA or Old Arabic) were pronounced, none of the reading conventions currently in use for the pronunciation of MSA has any historical validity and reflect a compromise that is not identical to any known Arabic dialect, ancient or modern. The lay Arab view is that the con-

temporary pronunciation of MSA continues that of CLA, the scholarly language of the medieval period, and ultimately that of Quraysh, the Prophet's tribe. But again, we have no means of knowing how the Arabic of the Prophet's time was pronounced, beyond the statements of the grammarians, themselves writing two hundred years later, that there was considerable dialectal diversity in the Arabic of the Prophet's time.

2.1.1 Segmental phonology

Table 2.1. of MSA consonants given above reflects the norms of scripted formal speech (e.g., in radio and television news bulletins) in the Arab world as a whole, including the few regional differences (in brackets, and discussed below).

The bracketed regional variants, which occur even in the most formal styles of "secular" MSA but are normally standardized in the "religious" style of Koranic recitation and formalized religious discourse, are geographically distributed as follows:

(a) Invariably in Egypt, and to some extent in Oman, MSA /j/ is replaced by its Cairene/Omani dialectal analog /g/. In Syria, Lebanon, Palestine, and Jordan, on the other hand, /j/ is (variably by some, unconditionally by others) replaced by a fricative /ž/, a voiced counterpart of /š/, again reflecting the local dialectal usage of the educated population in cities such as Damascus, Beirut, Haifa, Jerusalem, Nablus, and Gaza. The same /ž/ is also normally heard in North African spoken realizations of MSA and is also the pronunciation of the dominant city dialects of Tunis, Algiers, and Rabat.

(b) In Cairene MSA (and to some degree in Levantine), the emphatic interdental fricative /ḏ̣/, normal in Iraqi, Gulf, and Arabian pronunciations of MSA, is replaced by /ẓ/, with a dental-alveolar point of articulation, thus forming a plain–emphatic dental fricative contrast /z/–/ẓ/.[8]

(c) Again in Cairene MSA, even in relatively careful scripted speech, the interdental fricatives /θ/ and /ð/ are sporadically replaced by their corresponding dialectal phonemes /s/ and /z/. This tendency is not nearly so pervasive at this formal level, however, as is the substitution of /ẓ/ for /ḏ̣/.

Three vowel qualities are distinguished in MSA:[9]

/i/ close, front, lip-spread ——————— /u/ close, back, lip-rounded

/a/ open
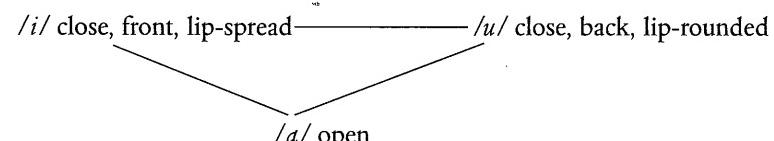
The phonetic realization of these vowels, whether in their short or long form, depends primarily on consonantal context. For example, /a/ is variously: (i) a front half-open [æ] in contiguity with plain consonants articulated forward of the soft palate, for example, [fænnæːn] 'artist', [sædd] 'dam', [θæːliθ] 'third'; (ii) a lower and more centralized vowel, symbolized here as [å], in the presence of a pharyngeal, for example, ['åːlæm] 'world', [ʔinsåħåbu] 'they withdrew', and often /r/, for example, [råjåʕå] 'he returned'; (iii) backed, when in contact with a velar fricative or /q/, for example, [qaːla] 'he said', [xaːmis] 'fifth', [baġdæːd] 'Baghdad'; and (iv) invariably backed (and somewhat rounded if a labial is also present) when an emphatic consonant is in contiguity, for example, [ṣɒbr̩] 'patience', [bɒṭṭ] 'ducks'. /i/ and /u/, short or long, also exhibit a range of allophonic variation that depends on consonantal environment.

The examples above illustrate the conventionally "correct" way of realizing /a/ in MSA. However, once again, the regional background of the speaker is an important factor in performance because, although there are phenomena that are universal in the dialects—the backing and lowering effect of the emphatics, for example—there are considerable and subtle dialectal differences in the quality and positional distribution of the vowel allophones (as there are in stress assignment and preferred syllable structure, for which see below), and these tend to show through in spoken MSA. Along with the consonantal variations noted above, it is these differences that enable the sophisticated native speaker spinning the dial of his radio to spot in a few seconds the probable origin of an anonymous news broadcaster who is in midsentence. An educated Arab hearing a snippet such as the following (+ = word boundary, {} = pause):

> [fæjjårå + riẓɒːlu + lmuqaːwæmæti + ħwaṭaniːyɒti + llubnæːniːyæ + lyawm + {} 'åb-wåtɒn + nɒːsifɒtɒn . . .]
> 'Men of the patriotic Lebanese resistance today detonated an explosive charge . . .'

would conclude that he was not listening to an Egyptian station because [j] and [ž] occur where Egyptian would have [g] and, despite the potentially North African [ž], that the speaker was unlikely to be broadcasting from Morocco or Algeria either—if he or she were, /a/ in final position would be likely to be realized as [ɑ] or even [ɑː]. Lebanon or Syria would, however, fit the phonological facts.[10] In reality, of course, the political slant of their content makes the origin of news broadcasts very obvious, but where this is less easy—in more neutral pan-Arab cultural material, say—the phonological clues usually give the game away to the experienced ear.

2.1.2 Syllable structure

All syllables in MSA begin with a single consonant and are either open (Cv or Cv:) or closed (CvC, Cv:C, CvCC, or Cv:CC). There are positional restrictions: Cv:C occurs internally (that is, within a word) only in a restricted set of forms, for example, 'a:mma (= Cv:C + Cv). Otherwise, in the connected "spoken prose" of reading style, only Cv, Cv:, and CvC occur. Where Cv:C would occur across word boundaries, as it does in certain common types of junctured phrases, it becomes CvC. For example, fi: + lma:ḍi: 'in the past' would be syllabically (where / = syllable boundary) fi:l/ma:/ḍi:, so the vowel of fi: is shortened: fi + l/ma:/ḍi:. A form such as ba:ħaθu: + lʔazma 'they discussed the crisis' is realized as ba:/ħa/θu + l/ʔaz/ma for the same reason. There are, however, a few cases where a Cv:C or a CvCC syllable is admitted in nonfinal position so as to maintain a grammatical distinction. For example:

> mu/maθ/θi/lu: + š/ša/ri/ka
> 'the company representatives'

contrasts with

> mu/maθ/θi/lu + š/ša/ri/ka
> 'the company representative'.

As the result of a reading convention whereby final short vowels are omitted in prepause position, Cv:C, Cv:CC, and CvCC, may occur utterance-finally. The sentence in the previous paragraph (here in broader transcription) would be divided syllabically as follows:

> fáj/ja/ra/ + ri/ja:/lu + l/mu/qa:/wa/ma/ti + l/wa/ṭa/ni:/ya/ti + l/lub/na:/ni:/ya + l/yawm/{} 'ab/wa/tan/ + na:/si/fa/tan/ . . .

As is clear from this example, syllable boundaries in connected MSA speech are not necessarily coterminous with word boundaries.

2.1.3 Neutralization

In carefully enunciated MSA, long vowels and consonants are maintained except in prepause position, where they are shortened. Thus ʔadduwalu lkubra: 'the superpowers' is realized as [ʔadduwalu lkubra] in utterance-final position and ha:ða lħaqq 'this right' as [ha:ða lħaq]. However, a distinction needs to be drawn here

between doubled final consonants that are phonetically short as a result of neutralization and those that are phonemically short. In many dialects, forms such as [*sákat*] < *sakat* 'he was silent' and [*sakát*] < *sakat+t* 'I was/you were silent' both have a phonetically short final [*t*], yet the stressing of these forms (stress being automatic and nonphonemic in MSA and the dialects) suggests a difference in the substance of the final syllable that is difficult to hear. Acoustic experimentation might reveal whether there really is no difference between such forms; if there is not, one would have to allow stress a phonemic role in distinguishing just these types of forms or regard the stress in forms such as [*sakát*] to have been assigned at the "deep" level (i.e., *sakat+t* before a surface rule of final consonant neutralization (e.g., $C_3 \rightarrow \emptyset$ / — $C_2 C_3$ # where $C_2 = C_3$). Similar pausal forms also occur in MSA, for example, *ʔájal* 'indeed' and phrase-final *ʔajál* < *ʔajállu* 'more sublime'.

The main exceptions to the shortening of long vowels and consonants in prepause position are a small number of pronominal suffixes that are generally kept long in pause: -*i:* 'my', -*ni:* 'me', -*ha:* 'her', -*na:* 'our, us', -*kuma:* 'the two of you/your' and -*huma:* 'the two of them/their'. These are conventions accepted throughout the Arab world. However, once again, local dialectal influences may make themselves felt in less careful but still highly formalized MSA speaking. In Cairene dialect, only one long vowel per phonological word is allowed: *ʔa:lu* 'they said' becomes *maʔalu:š* 'they didn't say', rather than *maʔa:lu:š*. In MSA, by contrast, words containing two or three long vowels are common. There is a tendency among Egyptian speakers speaking MSA to follow the dialectal rule and neutralize the length of all but the final long vowel in such words, for example, *ʔištira:ki:ya* 'socialism' is realized as [*ʔištiraki:ya*], *lisa:ni:ya:t* 'linguistics' as [*lisænyœ:t*].

2.1.4 Stress

Stress in both MSA and the dialects is nonphonemic, and the possible exceptions to this, as we have just seen, are marginal and artificial. The conventions for word stress in "secular" MSA reflect the rules for stress placement in the urban dialects spoken in the Mashreq, that is, Egypt and the countries to the east of it. And just as between certain of these dialects there is a degree of variation in stress placement in certain syllabic concatenations (in particular between Egypt and the rest), so there is a corresponding fluctuation in regional realizations of these same concatenations in MSA. It is convenient to divide syllable types into "light" and "heavy" on the basis of the number of segments they contain. "Heavy" syllables in MSA generally occur only in prepause position (see 2.1.2) and contain four segments (length counting as a separate segment): *Cv:C*, *CvCC* (where *CC* may be two different consonants or two the same). An excep-

tion is *Cv:C* in certain forms such as *'a:mma* 'public' and *duwaybba* 'small beast'. "Light" syllables contain three or less: *Cv*, *Cv:*, *CvC*.

In MSA, no more than one heavy syllable may occur per word, and if one occurs, it is stressed, for example, *rijá:l* men', *billáwr* 'crystal', *sijíll* 'register'. In words of "light" syllables only, stress assignment depends on the number of syllables in the word:

(a) In words containing three syllables or less, the accent is on the penultimate, for example, *málik* 'king', *gúrfa* 'room', *madrása* 'school', *θallá:ja* 'refrigerator', *ba:ḥáθu:* 'they discussed'. The two types of exception to this are that (i) some speakers accent antepenultimate *Cv:* if the penultimate is *Cv*, viz. *bá:ḥaθu:* (and there are a small number of common words in which this stress pattern is applied by all speakers, e.g. *há:ðihi* 'this (f.)' and *ðá:lika* 'that (m.)'); and (ii) if penultimate and antepenultimate are both *Cv*, the antepenultimate is stressed, for example, *ḥáraka* 'movement'.

(b) In words of four or more syllables containing *Cv:*: (i) *Cv:* is stressed (the *Cv:* nearest the end of the word if there are two), for example, *muná:saba* 'occasion', *muḥtawayá:tuhu* 'its contents', *ba:ḥaθú:hu* 'they discussed it'. If there is no *Cv:*, then (ii) the antepenultimate is stressed: *katábtuhu* 'I wrote it', *muškilátukum* 'your problem'.

Some Cairene MSA speakers, however, in line with the stress rules of their dialect, stress forms like those exemplified in (b) above differently. Cairene has no long vowels in antepenultimate position. Long vowels such as those in *muna:saba* are neutralized, and stress falls on the penultimate syllable.[11] This explains the variable tendency for such MSA forms to be realized as *muna:sába* by Cairenes when, for example, reading aloud from an MSA text. Cairene stress rules also stipulate that a nonfinal syllable following *CvC* is stressed.[12] Thus in forms such as *muḥtawaya:tuhu*, and those exemplified in (b) (ii) above, these rules may lead speakers to say *muḥtáwaya:tuhu, katabtúhu, muškílatukum*. But dialectal influence may not always be "negative." In the case of *katabtuhu*, the corresponding dialectal form is *katábtuh*, so stressed by the dialectal rules that, if directly applied to MSA syllable concatenations, gives "incorrect" outputs such as *katabtúhu*. However, the fact that the stress in dialectal verb forms such as *katábtuh* falls where it does may cause some dialect speakers to stress the corresponding MSA form at the same point in its morphophonological structure and (accidentally as it were) produce the "correct" MSA *katábtuhu*.

2.1.5 Pausal and junctural phenomena

Pause is defined (rather vaguely) as an audible break in delivery. The incidence of pauses in spoken MSA is variable and depends very much on the con-

tent of what is said, the actual or presumed audience, and the speaker's intentions. For obvious reasons, public orations at which there is a live audience present, and for some of whom at least MSA in its more "high-flown" styles will be a "foreign" language, tend to be delivered more slowly and with more frequent pauses to facilitate understanding. The following is an extract from a speech of Gamal Abdul Nasser, president of Egypt from 1954 to 1970, during the Suez Crisis of 1956. {} marks a pause; words between {} are in juncture:

ʔinna miṣr{} ḥi:nama: tuda:fiʕu ʕan ḥurri:yatiha:{} wa ʕan ʔistiqla:liha:{} wa tadʕu: li ssala:m{} fa hiya taʕlamu ʕilma lyaqi:n{} ma: huwa lfarqu bayna ssala:mi wa listisla:m {} . . . ʔaʕlanat miṣr{} wa gami:ʕu ʔabna:ʔi miṣr{} ʔinnaha:{} fi: sabi:li lmuḥa:faẓati ʕala: ha:ðihi lʔahda:fi lkubra{} satuqa:tilu fi: sabi:li ssala:m{} wa tuqa:tilu fi: sabi:li lḥurri:ya{} wa tuqa:tilu fi: sabi:li listiqla:l{}

'Egypt{} when it defends its freedom{} and its independence{} and calls for peace{} knows for certain{} what the difference is between peace and surrender{} . . . Egypt has made it clear{} as have all its sons{} that{} in order to safeguard these great aims{} it will fight for peace{} and fight for freedom{} and fight for independence{}'

Compare this slow delivery with the speed and less frequent pauses of a typical radio broadcast (Syrian television news, 9 April 1989):

fajjara rija:lu lmuqa:wamati lwaṭani:yati llubna:ni:ya lyawm{} ʕabwatan na:sifatan fi dawri:yatin ʔisra:ʔi:li:ya{} ʔawqaʕat biha: ʔiṣa:ba:tin ʕidda{} wa ʔaḍra:ran ma:ddi:yatan kabi:ra{} wa ʔafa:dati lʔanba:ʔu lwa:ridatu min janu:bi lubna:n{} ʔanna lmuqa:wimi:n fajjaru lʕabwata ʔaθna:ʔa muru:ri ddawri:yati lmuʕa:diya{} fi: minṭaqatin taqaʕu ma: bayna ma:rja:yu:n wa lxayya:m{}

'Men of the Lebanese national resistance today detonated{} an explosive charge under an Israeli patrol{} causing numerous casualties{} and considerable material damage.{} There is a news report from the south of Lebanon{} that the resistance fighters detonated the charge while the enemy patrol was passing{} in an area lying between Marjayoun and al-Khayyam{}'

In prepause position, the following occur:

(a) Final short vowels that appear in prejuncture position are omitted in prepause: so, for example, a grammatically definite noun such as *ssala:mi* 'peace' in line 2 of Nasser's speech is realized as *ssala:m* before the pause, whereas later in the same line the word is in prejuncture, not prepause position and so appears

as *ssala:mi*. Similarly, *lyawma* 'today' is prepausal in line 1 of the Syrian text and so appears as *lyawm*.

(b) The inflectional suffixes that mark indefiniteness in the noun, -*un* (nom.), -*an* (acc.), -*in* (gen.), are not all treated alike: -*an* and -*in* are in some cases maintained prepausally, whereas -*un* is always omitted. The details, based on a more extended analysis of recent Syrian news broadcasts, are as follows:

(i) -*in*: *Pace* the comments of Beeston,[13] it seems that there is a distinction to be drawn between the treatment of -*in* as a marker of a noun in the genitive case and -*in* as a marker of an attributive adjective that is in agreement with a genitive head noun, when in phrase-final prepause position. In the Syrian broadcasts examined, the adjective in structures like *dawri:yatin ʔisra:ʔi:li:yatin* 'an Israeli patrol' is invariably realized without the -*in* when such an adjective occurs in prepause position, in line with Beeston's description, but nouns that are in a genitive relationship with another noun (except for mensurative expressions such as *θala:θata ʔayya:m*) or that are governed by a preposition, retain -*in* (shown in bold below) in prepause position. For example:

(1) genitive:
   *wa qa:la ʔinna lʔinsiḥa:ba ssu:ri:y{} yartabiṭu bi jala:ʔi ʔa:xiri ʔisra:ʔi:li:**yin**{} ʕan ʔarḍi lubna:n{}*
   'He said that Syrian withdrawal{} was linked to the departure of the last Israeli{} from Lebanese territory.{}
   . . . *yawma ġad**in**{}* (three-second pause)
   '. . . tomorrow{}'
(2) governed by a preposition:
   *wa ʕani lqawa:ʕidi lmaʔlu:fa{} bi nnisbati li muʕfad**in**{} ja:ʔa kama: qa:l{} . . .*
   '(and a departure) from the normal rules{} that apply to a delegate{}, who had come, as he said{} . . . '

(ii) The prepausal behavior of -*an* also seems to depend on its functional and grammatical status. When it marks an adverbial accusative, for example, *taqri:ban* 'approximately', *sa:biqan* 'previously', *kulli:yan* 'completely', some speakers (the vast majority?) retain it, while a minority realize it as -*a:*. This latter realization is archaic and (it is my impression) occurs in "secular" MSA mainly in the more "conservative" Arab countries: for example, it is regularly realized in this way by the newsreaders of Oman television. But when prepausal -*an* is grammatically a marker of an object noun or of an adjective that agrees with one, it is sometimes retained and sometimes omitted. In the Syrian broadcasts examined, mas-

culine nouns, masculine noun phrases consisting of a noun–adjective phrase in the accusative, and predicative adjectives that are the object of *ka:na* 'to be', all retain the -*an* prepausally (shown in bold in the examples below). For example:

(1) object noun alone:
*wa ʔanna su:ri:ya lan taddaxira juhd**an**{} fi musa:nadati ssu:da:n{}*
'and that Syria would spare no effort{} in supporting Sudan{} . . . '

(2) attributive adjective:
*laqad ḥaqqaqat ha:ðihi zziya:ra tafahhuman ka:mil**an**{} li lmawqifi fi ssu:da:n{}*
'This visit has brought about a complete understanding{} of the situation in Sudan{}'

(3) predicative adjective:
*wa qad ka:na ha:ða ttafahhumu ka:milan wa ša:mil**an**{} min na:ḥi:yati lʔaxi rraʔi:s ḥa:fi ẓ al-ʔasad{} . . .*
'This understanding was complete and total{} on the part of the President, Hafiz al-Asad{} . . . '

But prepausal -*an* is dropped from feminine nouns and adjectives in identical structures in all of (1), (2), and (3); for example (from the text quoted above), *ʔiṣa:ba:tin 'idda* (for *'iddatan*) 'numerous casualties' and *ʔaḍra:ran ma:ddi:yatan kabi:ra* (for *kabi:ratan*) 'considerable material damage' (the loss of the -*t* in these forms is discussed in (c)).

Much turns in the cases quoted here, of course, on what counts as a "pause." In all the examples quoted, there is a perceptible silence of at least a second where {} is marked, but it is however the case that what I have noted as the apparently exceptional forms retaining prepausal -*in* do not occur at the end of a sentence (where there tends to be a slightly longer pause) but at the boundary of clauses or phrases within a sentence. In other words, the decision to use pausal or nonpausal forms may depend on the type of grammatical and textual boundary with which the pause coincides. Harrell interprets some parallel examples from Egyptian radio Arabic of the 1950s of prejunctural -*in* occurring in prepausal, phrase-final position as "an attempt by the announcer at especial clarity or seriousness."[14]

(c) The -*t* morpheme that marks feminine gender in certain nouns and adjectives disappears in prepause, together with the -*u*/-*a*/-*i* (definite) and the -*un*/-*an*/-*in* (indefinite) case inflections suffixed to it, as described in (a) and (b) above. Prejuncturally, -*t* and these endings are all retained. This is illustrated in the Nasser speech by the phrase *fi: sabi:li lḥurri:ya{}* '(lit.) in the path of freedom', which would be *fi: sabi:li lḥurri:yati* in prejunctural position, and by *dawri:yatin*

ʔisra:ʔi:li:ya 'an Israeli patrol' in the news broadcast, which would be *dawri:yatin ʔisra:ʔi:li:yatin* prejuncture. In "religious" contexts, final -/*a*/, which results from the prepausal deletion of -*t*, is pronounced -/*ah*/.

(d) Some nouns and adjectives derived from "weak" roots (see 3.5.1) have an -*in* suffix in the nominative and genitive cases, graphologically identical to that which marks the genitive case in nouns from "strong" roots. This -*in* is normally realized as -*i:* in pause.

The rules for the treatment of words in prepause and prejuncture position in spoken MSA are of course at odds with the rules for their treatment in all forms of dialectal Arabic. And just as dialectal stress patterns tend to show through in the performance of MSA, so too do dialectal speech habits in the matter of pausal and junctural phenomena. In the dialects, the contrast between contextual and pausal forms was lost—pausal forms became universal. There is therefore no variation in the realization of words that depends purely on position in the utterance that can be compared with the examples in (a)–(d) above. The case inflections (exemplified in (a) and (b)) have also been lost in the dialects, and final long vowels have either been lost or, where they do occur at surface level, have become morphologized and are therefore not liable to positional neutralization, for example, Gulf Arabic *smá'u* 'they heard' contrasts in all positions with *sma'ú:* 'they heard him', and likewise Cairene Arabic *btíddi* 'you give' with *btiddí:* 'you give him', and Palestinian/Syrian *bíddi* 'I want' with *biddí:* 'I want it (m.)'. The result is that, even in formal "spoken prose" contexts such as news bulletins, there is a tendency for speakers to use pausal forms in certain commonly occurring phrase types, no doubt because of dialectal influence, where they should theoretically use contextual ones.[15] An example occurs in the noun phrase at the end of the first line of the extended extract from the Syrian broadcast quoted above: *. . . lmuqa:wamati lwaṭani:yati llubna:ni:ya lyawm{}* ' . . . the Lebanese national resistance today', which should read *lmuqa:wamati lwaṭani:yati llubna:ni:yati lyawm{}* because *llubna:ni:ya*, which is a prepausal form, is not in prepausal position. The reverse tendency—the hypercorrect use of contextual forms in prepause position—has already been exemplified in the examples given in (b) above. In news broadcasts, it is statistically much rarer than the use of pause forms in context[16] and may be just as much the result of wishing to avoid inflectionless forms that "sound dialectal" as "an attempt at especial clarity and seriousness" as Harrell claims.

In a recent experiment, Parkinson tested how acceptable as MSA a group of ordinary Egyptians judged different tape-recorded performances of the same MSA texts.[17] Certain segmental and nonsegmental (junctural, pausal) phono-

logical features, as well as some in the domain of morphophonology, were de-liberately varied in different combinations to produce versions of the text that diverged more or less markedly from normative MSA. Two main conclusions emerged. First, most Egyptians, if the experiment was at all representative of ac-tual behavior, will accept as MSA (in their terms *fuṣḥa:*) a wide range of styles in which dialectal influence may appear quite marked to the nonnative observer. In particular, mistakes in the application of the junctural and pausal rules de-scribed above apparently count for little in the average native speaker's evalua-tion of a performance as "MSA" or "not MSA." As long as the basic grammatical structures and vocabulary are perceived as *fuṣḥa:* (or, putting it negatively, non-Egyptian), an erroneous realization of such phonological features as these seems to have little importance. Second, it nonetheless emerges that there is a peck-ing order of features that speakers apply when ranking various performances of the same text on a cline of "MSA-ness" and that primary among these are seg-mental features. Thus, although the realization of MSA /θ/ as /s/ in words such as /θa:liθ/ 'third' does not in most speakers' eyes disqualify the text in which it occurs from being considered *fuṣḥa:*, Parkinson's research suggests that such re-alizations seem to be relevant in speakers' perceptions of lowered acceptability.

2.1.6 "Secular" and "religious" pronunciation

Native speakers' preparedness to accept phonologically nonstandard per-formances of MSA texts as, in their terms, *fuṣḥa:* does not extend, however, to scripture or other religious materials—at least, if they are Muslims (see note 2). What Badawi terms "heritage" *fuṣḥa:* (*fuṣḥa: al-tura:θ*), and *a fortiori* Koranic recitation (Ar. *tajwi:d*), requires the meticulous observation not just of the pausal and junctural phenomena described above, but also of certain phonetic and phonotactic conventions not usually observed in "secular" MSA. The following is a summary[18] of the most salient differences between this "heritage" style and, at the other extreme, the relaxed reading style that most ordinary Egyptians would employ in reading aloud nonscriptural material:

RELIGIOUS	RELAXED READING
heavy emphatization of /ṭ/, /ḍ/, /ṣ/, /ð̣/ (and spread to contiguous segments) with very pronounced lip-rounding	backness, merely, compared to the corresponding plain consonants
—	spread of backness to neighboring syllables
final -/ah/ in feminine nouns	final -/a/ in feminine nouns

performance of long vowels in line with orthography	use of long vowels only in accented and/or open syllable
interdental fricatives maintained	interdental fricatives replaced by corresponding dentals
/j/ maintained	/j/ replaced by /g/
/q/ maintained, with very retracted point of articulation	/q/ more front, sporadically replaced by /ʔ/
other postvelars pronounced with a very retracted point of articulation	postvelars with a more forward point of articulation
-/ay/ and -/aw/ maintained	pronunciation relaxed, so that they take the form of diphthongs proper -/ai/ and -/au/

2.2 The phonology of the dialects

In general, the dialects differ from MSA in that they have a reduced and re-structured consonant system but more complexity in the vowel system. The rules of syllable structure and accentuation are also different, and, as we have already noted, the contrast between contextual and pausal forms typical in the formal oral performance of MSA has been lost, although that is not to say that the dia-lects lack pausal phenomena. We will now review some of the more important features of dialectal segmental (2.2.1) and suprasegmental (2.2.2) phonology.

2.2.1 Segmental phonology

The dialectal phonological systems, when compared to those of Old Arabic (henceforth OA = the putative ancestor of both MSA and the colloquials) and its modern descendant MSA, have lost some phonemic distinctions and partially or completely restructured others. The overall effect is that the dialects show a reduced phonemic inventory and considerable rearrangement of the system, although the details vary from one region to another. Cross-cutting this primary geographical variation, there is a further important covariation of phonologi-cal type with socioeconomic lifestyle. The dialects of those Bedouin groups that have only relatively recently become sedentarized or are still seminomadic show a typological similarity even across large distances and national boundaries. Those from such widely scattered areas as the Negev in Israel, the Syrian desert, or the recently established city-states of the Gulf, speak dialects that have more in common with each other—at least in phonology and morphology, on which external influences have tended to work more slowly—than they do with the

dialects of neighboring towns. In the same way, the dialects of those with a continuous line of descent through urban populations are also typologically akin. A third socioeconomically defined dialect type can also be distinguished in some areas (especially the Levant) that differs from both the "city" dialects on the one hand and the "Bedouin" ones on the other. This type, which might be termed "ruralite," is spoken by long-established settled farming communities in villages. The overall effect is of a dialectal patchwork quilt. So, for example, despite the fact that they live some distance apart from each other and the intervening space is filled with communities of "ruralites," Jerusalemites, Damascenes, and Aleppines all speak a single type of urban dialect, phonologically speaking. The historical reasons for this situation lie in the different origins and patterns of settlement of the waves of migrants into these areas over many centuries from the Arabian Peninsula. The long-term maintenance of these distinctions is a consequence of the endocentric social structures of the Bedouin and ruralite groups, and the nature of the contacts, social, economic (and therefore linguistic) that have only recently started to change.

In general, it can be fairly said that the Bedouin-type dialects, such as are spoken in southern Iraq including Baghdad, the desert areas of Syria and Jordan, central and eastern Saudi Arabia, and the Gulf States, preserve a greater number of archaic (in the sense of nearer to OA) features than do the dialects of the old-established cities of the Fertile Crescent and Cairo. This is undoubtedly because of their historical isolation and consequent conservatism. But the Bedouin dialects of today are not of course the same as OA. Like any language, they have been undergoing changes throughout their history, if perhaps at a slower pace than is true of the urban vernaculars. In recent decades, the perceived "prestige" of speakers of the urban vernaculars has also had a dramatic effect, as we shall exemplify below.

We will now look briefly at three areas of salient difference in segmental phonology between MSA and the dialects of the eastern region of the Arab world, and between these dialects themselves. Of the many differences that could have been chosen, these particular ones are important because ordinary Arabs from the area are themselves aware of them and, as a consequence, they have acquired the status of markers of geographical/social identity and speech style.

2.2.1.1 The interdentals

In the eastern Arab world, Bedouin-type dialects preserve the OA/MSA interdentals /θ/, /ð/, and /ð̣/, whereas the city dialects have merged them with the dentals, thus:

OA/MSA	GULF	JERUSALEM	
θa:nin	θa:ni	ta:ni	'second'
ḥa:ð̣a:	ḥa:ð̣a	ḥa:da	'this'
ʔað̣allu	að̣all	adall	'I stay'

Such correspondences as these between Bedouin-type dialects and OA/MSA are the source of much lay Arab opinion on the supposed "purity" (and hence cultural superiority) of these dialects. However, the other side of coin is that, in the Bedouin dialects, /ḍ/ (that is, a voiced, emphatic, dental stop, the normal MSA realization of orthographic ḍa:d) does not exist (or more probably never existed as a separate phoneme), merging with /ð̣/, so the following typical contrasts occur:

OA/MSA	GULF	JERUSALEM	
ḍaraba	ð̣arab	ḍarab	'he hit'
ʔabyaḍu	abyað̣	abyaḍ	'white'

In fact, what probably happened historically was that the Bedouin and city dialect systems restructured the asymmetrical OA system in different ways, as illustrated in the diagram below. The city dialects merged the interdental fricatives with the corresponding set of dental plosives, and OA ḍ^l, a lateralized emphatic, became the voiced counterpart of t. The Bedouin dialects, on the other hand, reduced the asymmetry of the OA system by merging ḍ^l with ð̣, creating three-member sets of dental plosives and interdental fricatives in which there is a three-way voiced–voiceless–emphatic set of contrasts as the diagram on page 72 illustrates.

In the city dialects of Syria, Jordan, and Egypt, a more recent trend in the speech of educated speakers has been the performance of interdental fricatives that occur in neologisms imported from MSA (and now more generally in "dialectal" words also) as the corresponding dental fricatives /s/, /z/ and /ẓ/. Among urbanites, a cline of formality/"correctness" seems to have developed such that dental fricative realizations are perceived as more "educated" than the native urban dialectal plosives /t/, /d/ and /ḍ/, yet not so formal as "correct" MSA interdental fricatives.[19] So, for example, in Jordanian city dialects such as that of Amman, variation between the following might typically occur, both within one person's speech (depending on the context and topic of what is being said) and from one speaker to another in analogous speech contexts (depending on the speaker's relative level of education):

STYLE

| "informal/uneducated"	"semiformal"	"formal/educated"	
(dental plosive)	(dental fricative)	(interdental fricative)	
maṭal	*masal*	*maθal*	'example'
ḍihin	*zihin*	*ðihin*	'mind'
ḍill	*zill*	*ðill*	'shadow'

Of course, recently sedentarized Bedouin who move into expanding towns like Amman rapidly become aware of the different social and stylistic significances of such variants. But despite the fact that their native dialectal /θ/, /ð/ and /ð̣/ are, from the historical (and popular) point of view, the "original" forms and indirectly promoted as "correct" through the use of MSA in education and the media, these Bedouin nonetheless tend to switch to nonstandard variants when faced with interlocutors from the city. This kind of dialect-to-dialect switching, in which there is a clash between the linguistic prestige system as-

(a) OA system:
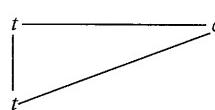
```
stops                 t ——————— d
                      |        ╱
                      |       ╱
emphatization         ṭ      ḍ
lateral

fricatives            θ ——————— ð
                       ╲        |
                        ╲       |
emphatization            ð̣
```

(b) "city" system:

```
stops                 t ——————— d
                      |         |
                      |         |
emphatization         ṭ ——————— ḍ
(fricatives)         (θ)       (ð)
(emphatization)               (ð̣)
```
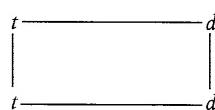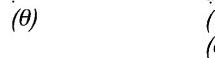
(c) "Bedouin" system:

```
stops                 t ——————— d
                      |        ╱
                      |       ╱
emphatization         ṭ

fricatives            θ ——————— ð
                       ╲        |
                        ╲       |
emphatization            ð̣
```
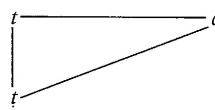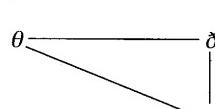
sociated with pan-Arab MSA and that of a locally dominant social group has been frequently commented on,[20] and described in detail for other areas of the Arab world.[21] It is not then the case that normative MSA forms, prestigious though they are, are the sole influence in phonological variation and change: the perceived prestige of competing local or regional dialectal usage is also a potent force. As far as its operational scope is concerned, segmental variation of this type seems to depend primarily on the connotational and denotational meanings of a word as "local," "nonlocal," or "potentially both local and nonlocal." Its status in these regards determines whether it is liable to variation in its phonological form at all, and, if it is, the particular choice of phonological variants that are permissible. Thus, in Cairo, the /t/ of *taːni* 'again' in plain dialectal adverbial usages such as *matiʔulʃⁱ kida taːni* 'Don't say that again!' cannot be simply replaced by /θ/ or /s/ to give *θaːni* or *saːni* if this sentence is to be performed in a more formal style; the /t/ is an invariant component of the lexical identity of *taːni* as the Cairene way of saying 'again'. But *taːni* as an adjective in the semantically identical but less local phrase *marra taːniya* 'again, one more time' could and would be realized as *marra saːniya* or (more formally still) *marra θaːniya* if for some reason an educated or, at any event, less Cairene-sounding style was being aimed at. Conversely, a completely nondialectal word such as *θaqaːfa* 'culture' cannot be "dialectalized" by being realized as *ˣtaqaːfa* but could be made somewhat more informal sounding, if need be, by being realized as *saqaːfa*. /θ/ (or at a pinch /s/) remains just as much an unalterable part of the identity of the nonlocal word *θaqaːfa*, whoever is pronouncing it, as /t/ does of the local adverbial usage of *taːni*. A more extended treatment of the lexical conditioning of phonological variation can be found in 2.3.1 and 2.3.2 below.

2.2.1.2 Velars and uvulars

The OA phonemes /k/ and /q/ developed differently in the city, rural, and Bedouin dialects.[22] Basically, the city dialects in the east (Cairo, Jerusalem, Damascus, and Beirut) preserved /k/ and generally backed /q/ so that it became a glottal /ʔ/, with the original glottal in these dialects being replaced in some positions by /y/, in some by /w/, and in others disappearing completely. In many areas of rural Palestine, /q/ was fronted to /k/ and original /k/ unconditionally affricated to /č/ (these are the classic features of "ruralite" dialects in the Levant). In the dialects of Bedouin and Bedouin-descended populations (central/eastern Arabia and southern Iraq including Muslim Baghdad), /q/ developed differently again, being generally fronted and voiced to a velar plosive /g/. Some Bedouin dialects took the fronting process further, affricating this /g/ to /j/ (and further to /dz/) in front vowel environments only and /k/ to /č/ (and

further to /ts/) in those same environments. These historical developments gave rise to the following typical correspondences in the modern eastern dialects:

OA/MSA	(A)	(B)	(C)	(D)	(E)	
qa:la	ʔa:l	ka:l	ga:l	ga:l	ga:l	'he said'
ṭari:q	ṭari:ʔ	ṭari:k	ṭari:g	ṭari:j	ṭari:dz	'road'
samak	samak	samač	simač	simač	simats	'fish'
-kum	-ku/kum	-čim	-kum	-kum	-kum	'you (pl.)'

(A) Cairo/Damascus/Jerusalem ("urban" dialects)
(B) Central Palestinian villages ("ruralite" dialects)
(C) Eastern Jordan/Muslim Baghdad ("Bedouin" dialects Type 1)
(D) Southern Iraq/Gulf coast ("Bedouin" dialects Type 2)
(E) Central Saudi Arabia ("Bedouin" dialects Type 3)

So much for history; synchronically, just as in the case of the different reflexes of OA /θ/, these dialectal reflexes of /q/ and /k/ have come to acquire distinct shades of social and stylistic significance in different areas.

In Jordan, which lacked any large urban centres until the 1940s, speakers of Type (A), (B) and (C) dialects have all settled in the cities of Irbid and Amman. The original "Jordanian" dialect—that spoken on the east bank of the River Jordan—is designated here Type (C). As a result of the political upheavals of 1948 and 1967, large influxes of speakers of dialect Types (A) and (B) have had to be absorbed. The results of this mass immigration have been the creation of a linguistic melting pot in the cities and the emergence of an urban koinè. But this koinè is not (at least not yet) a stable dialectal variety in the same way as Cairene. Subtle differences distinguish the language of men and women, and one generation from another. According to Abdel-Jawwad, the /g/ pronunciation in words that had OA/MSA /q/ seems now to have acquired associations of "Bedouin" masculinity and toughness, whereas the /ʔ/ realization is associated with softness, femininity, and sophistication.[23] Men who come from backgrounds in which the (C) /g/ dialects are spoken retain /g/, while those originally from (A) /ʔ/ dialect backgrounds (e.g., Jerusalem) are tending to drop /ʔ/ in favor of /g/. Conversely, women from all dialectal backgrounds strongly favor /ʔ/.[24] The loser in this is the (B) /k/ dialect, whose younger speakers seem to be switching strongly to /g/ or /ʔ/, because of the perceived "ruralness" (and hence stigmatization) of the /k/ pronunciation. The /č/ reflex of OA /k/ also seems to be universally stigmatized as "rural" and is dying out rapidly in the speech of the younger generation.[25]

The same dialectal reflexes in Kuwait have acquired quite different local values and provide a striking example of what Ingham[26] aptly calls the "multivalency" of dialect features in Arabic. Johnstone noted that, while walking along

the seafront in Kuwait City in 1958, he was rebuked by two small boys for using the Kuwaiti pronunciation simač ('fish') instead of samak, the pronunciation shared by both MSA and the dialects of the urban areas of the Levant and Cairo. Young Kuwaitis found it difficult or even embarrassing to speak "pure" Kuwaiti in his presence. He concludes that "there is no real doubt that the Kuwaiti dialect is doomed to disappear in a relatively short time and that it will be replaced by a local version of the pan-Arabic koinè."[27] In fact, this has not happened. The demographic policies adopted by the Kuwait government, which accompanied the country's rise to economic dominance and which can be seen as a defense against the dilution of the Kuwaiti identity by the huge numbers of expatriate Arabs and others who flocked there, were resolutely separatist and nonassimilatory. From the 1960s onward, Kuwaitis were encouraged to live in special areas and to marry only other Kuwaitis, and tended to engage in certain kinds of employment only. A by-product of this has been to strengthen those aspects of culture and self-representation—notably dress and dialect—that are distinctively Kuwaiti. Educated Kuwaitis do not of course go around talking to non-Kuwaiti Arabs in extreme varieties of their own dialect, but within the family and the social milieu—which tends to be exclusively Kuwaiti—the dialect survives and indeed thrives as a badge of social identity and solidarity, and not, as far as I can tell, in a debased or koinèized form. Kuwaiti popular culture, in the form of dialect poetry, film, theater, and radio and television soap opera, flourishes and is even exported to neighboring states. So the /ʔ/ of the Egyptian, Syrian, and Palestinian Gastarbeiter shows no sign of replacing the Kuwaiti /g/ or the /k/ of the Kuwaiti /č/—at least, in everyday domestic and even public contexts. In fact, quite the reverse: recent political events have, if anything, served to strengthen any element, including the dialect, that tends to distinguish Kuwaitis from other Arabs. On occasions when a rather more formal style is required, especially if it involves interlocutors or listeners from other Arab countries, Kuwaitis switch in the direction of MSA standards rather than those of other dialects.

2.2.1.3 The alveolar affricate /j/

It was noted in our description of MSA segmental phonology that OA /j/ (Arabic letter ji:m) is usually rendered in MSA as an alveolar affricate /j/, except in Egypt where it is pronounced in all readings of nonreligious material as a velar plosive /g/. This is, of course, also the normal Cairene dialectal realization, and, when asked, Arabs from outside Egypt invariably cite it as the most salient feature of "Egyptian" speech—although it is in reality, apart from Cairo, a feature of Lower Egyptian dialects only and specifically those of certain Nile Delta towns. In the dialects of many parts of the Levant, especially the major cities of Beirut, Damascus, and Jerusalem, /ž/ corresponds to Cairene /g/ and

MSA /j/. Just as /g/ is universally regarded as "Egyptian," so /ž/ is thought of as "Levantine," even though it occurs much more widely than in this area. To a certain degree, as already noted, /ž/ has also infiltrated into Levantine pronunciations of MSA, in which it varies with /j/. Ruralites and Bedouin-descended groups in the Levant areas generally have an alveolar affricate /j/ identical to the MSA phoneme. Further east, in Jordan and the desert terrain beyond the River Jordan as well as in Arabia proper, the predominant reflex is also /j/, but in many areas of Lower Iraq and in the Gulf States in general, /j/ varies with /y/. Again, with the increasing ease of international travel and the dialectal contact that has followed as a consequence, /y/< OA /j/ has come to be regarded by Arabs from outside the area as a marker of "Gulf speech." In the case of reflexes of OA /j/, just as in the case of the reflexes of the interdentals and velar and uvular consonants already discussed, it is the dialectal pronunciation heard in the major center of population, political power, and economic strength in each country or area that has acquired the status of regional dialectal "standard" for those subject to its influence, and the label of national dialectal stereotype for those from outside who come into contact with it.

2.2.2 Aspects of suprasegmental dialectal phonology

In this section we will look briefly at some important aspects of dialectal suprasegmental phonology and contrast them with the position in MSA.[28] Two different dialect types will be exemplified: Cairene, an example of a stable urban type with classic "sedentary" dialect features, and Gulf Arabic as exemplified by the dialect of Bahrain (as spoken by Sunnis), which shows many classic "Bedouin" ones.

2.2.2.1 Syllable types and consonant clusters

As we saw above, all syllables in connected MSA speech are consonant-initial, and the basic types are Cv, Cv:, and CvC. Other types—$CvCC$, Cv:C, and Cv:CC—generally occur only in prepausal position as a consequence of the convention that deletes word-final short vowels in this position (see 2.1.2, 2.1.5 above).

The situation in Cairene is superficially similar. As in MSA, Cv and CvC are by far the commonest syllable types. Cv: is much less common than in MSA, and the types that are positionally restricted in MSA—$CvCC$ and Cv:C—are also uncommon.[29] However, unlike the situation in MSA, $CvCC$ and Cv:C may occur word-finally in juncture. When they do (a not uncommon phenomenon, obviously), there is a strong tendency to add, for rhythm's sake, a final short high vowel that breaks $CvCC$ into $CvC + C$. This "rhythmic" vowel is a noted feature of Cairene pronunciation. Compare the following examples, in which the first

line is a broad transcription with morpheme boundaries (= –) and word boundaries (= +) marked, the second line is a gloss of the morphemes, and the third is a syllabic transcription (boundary = /). Pause is marked in the third line ({}).

$CvCC$ prepause:

(1) MSA   *sawfa* + *yu-'allimu-kumu* + *d-dars*
            FUT   3msg-teach-you   DET-lesson
            saw/fa/yu/'al/li/mu/ku/mud/dars {}
            'He will teach you the lesson'

(1a) Cairene *ḥa-y'allim-ku*   + *d-dars*
             FUT-3msg-teach-you   DET-lesson
             ḥay/'al/lim/kud/dars {}
             'He will teach you the lesson'

$CvCC$ in juncture:

(2) MSA   *ha:ða* + *d-dars-u*   + *ṣa'b*
            this   DET-lesson-NOM difficult
            ha:/ðad/dar/su/ṣa'b {}
            'This lesson is difficult'

(2a) Cairene *ʔid-dars*   + *da* + *ṣa'b*
             DET-lesson   this + difficult
             ʔid/dar/si/da/ṣa'b {}
             'This lesson is difficult'

So in Cairene "citation," $CvCC$ syllables often resolve themselves in juncture into the more basic types CvC and Cv. Further, when in Cairene word-final Cv:C occurs in juncture, or when it ceases to be word final as a result of morpheme suffixation (again, a common occurrence), the long vowel is shortened. Thus word-final Cv:C in the citation form *bitʔu:l* (consisting of an aspectual prefix *b(i)* and a verb *tʔu:l*) 'you say' becomes in word juncture (3), and before suffixation (4):

Cv:C in word juncture:

(3) *bi-tʔul* + *ʔe:h*
    bi-2msg-say what
    bit/ʔul/ʔe:h {}
    'What do you say?'

Cv:C with morpheme suffixation:

(4) *ʔilli + bi-t̠ʔul-luh + ṣaḥḥ*
    which *bi*-2msg-say-to-him true
    ʔil/li/bit/ʔul/luh/ṣaḥḥ {}
    'What you say to him is true.'

The result in both cases is the resolution of the long syllable, *Cv:C*, into the more basic type *CvC*. Finally, *Cv:* is also subject in Cairene to positional reduction to *Cv* in a number of positions: when it occurs in antepenultimate position (see the description of the instability of MSA stress in 2.1.4 above), or when, as a result of suffixation, the vowel in the syllable following is also long (*ʔimtiḥa:n* 'examination' but *ʔimtiḥana:t* 'examinations'). These various reductional processes conspire to produce the observed predominance of light syllables, *Cv* and *CvC*, in Cairene. It should be clear that although, in a sense, there is a superficial resemblance in the syllabic structure of MSA and Cairene in that *Cv* and *CvC* are overwhelmingly the most common types and that the heavier syllables *Cv:C* and *CvCC* are in both cases positionally restricted, the motivations for these restrictions are different. In the case of MSA, the characteristic patterning of long and short vowels is produced by consciously learned reading conventions that require short-vowel case marking to be observed in juncture but not in pause; in Cairene, vowel length is governed by natively acquired rules of dialect prosody that give Cairene its particular rhythm.

In Bahrain ('*Arab* (= Sunni) dialect), the main syllable types are *Cv* and *CvC* as in MSA and Cairene. But, unlike the position in these latter two varieties, *Cv:C* occurs without restriction word-initially and medially, and is stable in juncture. For example:

initial: *ja:lbu:t* (*ja:l/bu:t*) 'type of boat'
medial:

(a) within a word: *yiso:lfu:n* (*yi/so:l/fu:n*) 'they chat'
(b) in juncture: *yigu:l lina* (*yi/gu:l/li/na*) 'he tells us'

CvCC occurs word-finally as in Cairene and MSA (where of course it is restricted to prepausal position) but unlike in the latter two varieties, it occurs regularly word-initially, medially, and in juncture. For example:

(a) initially: *šiftha* (*šift/ha*) 'I saw her'
(b) medially: *kallamtha* (*kal/lamt/ha*) 'I spoke to her'
(c) in juncture: *šift hind* (*šift/hind*) 'I saw Hind'

Table 2.2 Syllable Structures in Bahraini and Cairene

OA/MSA/LOAN	BAHRAINI		CAIRENE	ENGLISH
	Morphemic	Syllabic	Syllabic	
rakiba-hu	*rčib-a*	(*rči/ba*)	(*ri/ki/bu*)	'he rode it'
takallam-tu	(*ʔi*)*tkallam-t*	(*tkal/lamt*)	(*ʔit/kal/limt*)	'I spoke'
ta-xa:f-u:na	*t-xa:f-u:n*	(*txa:/fu:n*)	(*ti/xa:/fu*)	'you fear'
('hospital')	*sbe:tar*	(*sbe:/tar*)	—	'hospital'
ta-qu:l-u	(*ʔi*)*t-gu:l*	(*tgu:l*)	(*ti/ʔu:l*)	'you say'

Other syllable types that do not occur in MSA or Cairene or are marginal in them—*CCv, CCv:, CCvC,* and *CCv:C*—arise frequently in Bahraini as a result of various deletion and affixation processes and also in arabicized loan words. In sequences of more than two *Cv* syllables, in certain sequences of *Cv* and *CvC* syllables, and where *Cv* precedes *Cv:C*—in all these cases the vowel of the first *Cv* is normally deleted, especially if it is /*i*/ (whether this /*i*/ continues historical /*i*/ or developed as a result of the raising of historical /*a*/).[30] The results are word-initial *CCv, CCvC, CCv:,* and *CCv:C* syllables, typical also of northern and eastern Arabian dialects in general and of other "Bedouin" dialects from much further afield (e.g., Sinai and eastern Libya). In utterance-initial position, *CC*-clusters are freely formed if one of the elements is a stop and the other a continuant, or both are continuants. Certain foreign loan words are also fitted to this pattern. In cases of a two-stop initial cluster, a prosthetic vowel (devoiced /*i*/) is optionally prefixed. Examples (with Cairene forms for comparison) are given in Table 2.2. above.

CCvCC syllables also commonly arise in Bahraini as a result of the affixation of bound pronominal or interrogative morphemes to words with a *CCvC* or *CvCC* structure or through the arabicization of loan words. For example:

CCvCC *sma'-t* (*sma't*/) 'I heard'
        š-gil-t (*šgilt*/) 'What did you say?'
        š-rang (*šrang*/) 'What color?'
        trinj (*trinj*/) 'type of citrus fruit' (< Pers. *turunj*)

2.2.2.2 Stress

Three principles govern word stress in Arabic dialects virtually without exception:

(a) Stress is predictable and automatic, and determined by syllable structure.
(b) Stress is not syllabically fixed within a given word, but is assigned ac-

cording to whether the word is morphologically unadorned or has bound morphemes attached to it (i.e., stress assignment is determined by the syllabic shape of the phonological word, which may be different from that of the lexical word).

(c) Stress is nondistinctive.

In dialectal Arabic, preferred syllable structure is the main factor at the level of suprasegmental phonology that differentiates one dialect from another. Stress plays only a minor role, and the dialectal rules that express it are in fact remarkably similar to each other across vast geographical areas—it is the input to them that is different. For instance, in the dialects of Djidjelli, eastern Algeria, and Bahrain in the Persian Gulf, areas separated by 2,000 miles and in many other respects typologically different from each other, the stress rule is identical: if the last syllable of a polysyllabic word is long (*Cv:*, *Cv:C*, or *CvCC*), stress it; if not, stress the penultimate.[31] Where apparent differences in stress patterns between dialects occur, they are usually the result of differences in syllable structure. So, for example, in the dialects of Damascus and Cairo (typical eastern urban dialects from this point of view) this same "common denominator" stress rule applies to those syllable concatenations to which it applies in Djidjelli and Bahrain. However, sequences of short open syllables as in *katabu* 'they wrote' may occur in the Damascene and Cairene dialects (but not in those of Djidjelli or of (Sunni) Bahrain) and are normally stressed on the antepenultimate.[32] In Sunni Bahraini, the vowel of the first syllable in such sequences is deleted; but in Djidjelli, an open syllable may never appear in penultimate position, so the vowel of the second syllable is deleted. The effect of these syllabic constraints is to reduce all underlyingly *CvCvCv(C)* trisyllabic forms in these two dialects to disyllabics (in this case *ktíbu* and *kátbu* respectively), which are then stressed in accordance with the general stress rule given above.

There are, however, a few conditions in which stress does play a marginal distinctive role in the dialects, in apparent violation of principles (a) and (c) above. In some of the eastern dialects of the Arab world (Cairo and the Gulf among others), in the past tense forms of verbs whose stems end in -*t*, such as *sakat* 'he was silent', the affixation of the 1sg./2msg. enclitic -*t* results in a final homophonic cluster, for example, *sakatt* 'I/you were silent'. As we saw in 2.1.3, such final clusters of homophones are reduced to a single consonant in rapid speech, but stress is still assigned as if the final syllable was a long -*CvCC*. So [*sákat*] < *sakat* 'he was silent' contrasts with [*sakát*] < *sakat* + *t* 'I/you were silent'. In some Maghrebi dialects, principle (a) is not absolute in that there is one category of morpheme, p-stem verb prefixes indicating person: *nə-* (1st), *tə-* (2nd),

and *yə-* (3rd), that are never stressed even when they occur in syllabic positions that would attract stress according to the general stress rule noted above. As a result of the coincidental loss of initial and final semivowels in verbs from some roots, a few homophones result that are distinguished only by stress placement, for example, in Djidjelli,[33] [*nə́ḥla*] 'bee' (root consonants *NḤL*) and [*nəḥlá*] < *nə* + *ḥla* 'I become sweet' (root consonants *ḤLW*), [*tə́qwa*] 'piety' (root *WQY*) and [*təqwá*] < *tə* + *qwa* 'she is strong' (root *QWY*).

2.3 Phonological variation in the dialects

When describing the phonological characteristics of formal, "read" MSA, we noted that in both segmental and nonsegmental aspects of performance, the dialectal background of the speaker tends to make itself felt. In some cases, for instance in the Cairene /*g*/ realization of the Arabic letter *ji:m* (our /*j*/), this process has reached the point where the dialectal feature has become an accepted alternative MSA pronunciation. In other cases, segment replacement intrudes more sporadically and unconsciously, as in the tendency of Levantines to oscillate between standard /*j*/ and their own dialectal /*ž*/ and in the geographically widespread substitution of the dialectal midvowels /*o:*/ and /*e:*/ for the MSA /*aw*/ and /*ay*/.

In this section, we look at phonological influence in the opposite direction—from MSA to dialect. This is a complex subject, and to deal with it without also considering co-occurrent variation at levels of analysis beyond the purely phonological would tell only one part of the story. A separate chapter (chapter 9) is devoted to the covariation of linguistic elements across levels, whereas here we shall confine ourselves to word-level phenomena—principally the conditions under which a native dialectal segment may be replaced by its MSA analogue.

2.3.1 /*q*/

We saw in 2.2.1.2 that, historically, OA /*q*/ (however it was pronounced, and there is no certain evidence) developed differently in different parts of the Arabic-speaking world. Today, as a voiceless uvular stop, it exists in relatively few places—in northern Iraq, parts of Oman, Yemen, and North Africa. Everywhere else it is no longer a uvular. It became variously /*ʔ*/ in the big cities of the east, /*g*/ (sometimes with conditioned affrication to /*j*/) among the nomads, and /*k*/ in rural areas. We saw also that there is nowadays a pecking order of prestige between these competing dialectal variants in countries like Jordan where for social, economic, or political reasons speakers of dialects with different dialectal reflexes of OA /*q*/ have been thrown together: "urban" dialectal features are

variably replacing "Bedouin" or "rural" ones. But this interdialectal variation occurs against a backdrop of increasing literacy and awareness of the norms of MSA that exert a continual influence through education and the media. One of the salient features of spoken MSA in all countries, as heard in news broadcasts, secular public speaking of all kinds, mosque sermons, and the like, is the preservation of OA /q/ as a uvular plosive. Because this realization is so generally and distinctively nondialectal, it has become a kind of badge of formality when used in the context of extempore dialectal speech. By using it, speakers signal a deliberate raising of their speech style, for whatever reason.[34]

It is important to realize, however, that not all /q/s, so to speak, are equal—the use of /q/ in some types of word carries more significance, sociolinguistically speaking, than it does in others. Many words and phrases now in everyday use in all sections of society were not part of the dialects at all until recently and remain largely unassimilated to dialectal structural patterns, phonological, morphological, and syntactic. They are, as it were, implants, on "permanent loan" from written Arabic. Many of these words and phrases originated as loan translations from English or French, coined to express concepts for which there were no, or no exact, equivalents in Arabic. They are particularly common in areas of life and discourse in which western models and influence are pervasive: the media, academe, material culture, and science. To the extent that new artifacts, social structures, and ways of thinking about the world are gradually replacing old ones—and of course this varies with the education, social class, and even the sex of the speaker—the gap between the vocabulary and phraseology of MSA and the dialects is thus becoming narrower. MSA neologisms may simply *replace* dialectal words; for example, in the Gulf states the very general word MSA *minṭaqa* 'area, region' is supplanting a variety of local words such as *wa:di* 'village and its surrounding area', *di:ra* 'tribal homeland' and *firi:g* 'quarter of a city' as the social worlds to which these concepts belong change and disappear; or they may denote novel artifacts or concepts, for example, *tilivizyu:n naqqa:l* 'portable television', *taswi:q* 'marketing'. Either way, the direct borrowing into the dialects of MSA words and phrases is one of the main ways by which nondialectal phonological elements, including /q/, enter the dialects. Here are some further examples of MSA words and locutions containing /q/ that are now commonly heard even in informal, relaxed dialectal speech:

Abstract/process nouns: *θaqa:fa* 'culture', *ʔiqtiṣa:d* 'economy', *ʔistiqla:l* 'independence', *ʿaqli:ya* 'mentality', *taḥqi:q* 'checking, verification', *mura:qaba* 'censorship', *tabaqa* 'class, level'.
Concrete nouns: *miθqab* 'drill', *na:qila* 'oil-tanker', *qunbula* 'bomb'.
Adjectives: *mutanaqqil* 'mobile, peripatetic' (as in *ʿiya:da mutanaqqila* 'mobile

clinic'), *raqmi:y* 'digital' as in *sa:ʿa raqmi:ya* 'digital watch', *wa:qiʿi:y* 'realistic' (as in *mawqif wa:qiʿi:y* 'realistic position').

Set phrases: *al-naql al-ʿa:mm* 'public transport', *al-quwwa:t al-musallaḥa* 'armed forces', *qa:ʿida jawwi:ya* 'airbase', *al-qiwa al-ʿa:mila* 'workforce', *al-šira:ʔ bi l-taqsi:ṭ* 'hire purchase'.

Phonologically, words and phrases such as these are only marginally assimilated to local dialectal norms when used in extempore speech. Although the quality of the vowels usually reflects dialectal pronunciation, other features of their structure remain unaffected. Syllable structure, internal syntax where relevant (e.g., noun–adjective agreement and the MSA synthetic genitive in phrases) and certain consonantal segments (such as /q/) are examples of "resistant" features. So, for example, the long vowel in *θaqa:fa* may be realized as relatively front in dialects where /a:/ normally has this realization (most Levant dialects) or relatively back (Sunni Bahrain); its final vowel may be relatively high and front [æ] (Levant cities) or relatively low and back [a] (Arabia in general). But its CvCv:Cv syllabic structure remains unaffected whether this concatenation is normally allowed in the dialect or not, and its /q/ is never replaced by any of its dialectal analogs /ʔ/, /g/, or /k/. Thus there are no forms *θqa:fa, *θaʔa:fa, etc. However, a hybrid form such as *saqa:fa* is possible (indeed likely) in the speech of those for whom /θ/ is not a dialectal phoneme (Cairenes and Damascenes) and occurs in this form, as we have already seen, in more informal/less careful performances of MSA written texts (see 2.2.1.1). In other words, /θ/ is at the formal end of a /t/→/s/→/θ/ stylistic continuum in which /s/ covers the area of both "informal" MSA and "more formal" dialect; /q/, however, is not part of a continuum in the same sense and is a "frozen" element in the MSA words and phrases borrowed into the dialects.

This last statement has to be qualified, however. Some MSA lexical elements containing /q/ have become variable in some dialects. For example, a number of MSA neologisms, or words that have a specialized meaning, such as *mustaqbal* 'future', *mawqif* '(political) position', *ḥaqq* '(legal) right' may nonetheless be realized with dialectal /ʔ/ in Cairene: *mustaʔbal, mawʔif,* and *ḥaʔʔ.* Gamal Abdul Nasser, president of Egypt from 1954 to 1970, frequently pronounced these words in this fashion in passages of his speeches in which he was using dialect, but in other formal items of political vocabulary, such as *mi:θa:q* 'charter', *taqaddum* 'progress', *ʔiqta:ʿ* 'feudalism', he always maintained the /q/, even when the passage in which they occurred was in other respects dialectal. The different treatment is explicable in terms of the degree of specialization of meaning of the words involved: the more specialized (here in the field of politics), the less

likely to be variable; the more general, the more likely to become "dialectalized" (because of more frequent usage in a wider variety of contexts) and variable.

This relationship between semantic scope and susceptibility to phonological variation can also be illustrated by the distribution of meaning in words derived from the same triconsonantal roots. In Cairene dialectal Arabic, the verb 'to fight' is *ʔa:til* (historical root *QTL*), and 'one who fights' in a general sense (whether against another person or, metaphorically, for instance, sleep) is *muʔa:til*. In both these words, */ʔ/* is invariably the reflex used in dialectal speech. But the word for 'private soldier' in Cairene is *muqa:til*, and 'fighter plane' is the feminine form of the same word (*muqatla*), in both of which */q/* is invariable when these words are used in these technical senses, even in broad dialect. Under the root *QRR* we find a similar correlation of phonology and meaning: *ʔara:r* means 'bottom part' in Cairene, either in the sense of 'bottom of the sea' (*ʔara:r il-baḥr*) or the lowest octave an instrument can play; but *qara:r* means a decree or decision in a political sense, as in the typical collocation *qara:ra:t gumhu:ri:ya* 'presidential decrees'. Similarly, in Jordan, a radio presenter, when speaking broad dialect to a semiliterate peasant farmer nevertheless consistently uses */q/* in the participle *mutanaqqil* 'peripatetic, constantly on the move' in the MSA phrase *'iya:da mutanaqqila* 'mobile clinic' but uses */ʔ/* in the verb from which the participle *mutanaqqil* is derived when his meaning is less specialized: *lamma batnaʔʔal . . .* 'When I move around . . .'

Just as there are certain MSA words and phrases that are not susceptible to dialectalization by simple substitution of a dialectal reflex for */q/,* so too there are dialectal words that cannot be standardized by the reverse substitution process, for very similar reasons. In Sunni Bahraini,[35] where the dialectal reflex of OA */q/* is */g/,* the factors that block replacement of */g/* by */q/* are:

(a) lack of morphosemantic congruity with MSA lexical analogs: Bahraini *ḥalg* means 'mouth', but MSA *ḥalq* 'throat'; Bahraini *riyu:g* 'breakfast', but MSA *ruyu:q* 'spit'. In some cases the dialect morphologized an OA lexical item, for example, Bahraini *ga:'id*, an active participle derived from the verb 'to sit' but used with a following p-stem verb to indicate continuous action, for example, *hu ga:'id yizra' baṣal* 'he's in the middle of planting onions'; substituting MSA */q/* for */g/* to give *qa:'id* in the same sentence could only mean 'he is sitting planting . . .'

(b) archaisms: some dialectal words, although they appear in MSA or Classical Arabic lexica, have dropped out of active use. Bahraini *ba:g* 'to pilfer' has a "theoretical" Classical analog *ba:qa* 'to cheat, swindle', but *ba:g* cannot be style-raised by substituting */q/* for */g/* because *ba:qa* is not recognized as part of the MSA lexicon.

(c) local technical terms: many terms connected with local industries that have */g/ <* OA */q/* are so closely associated with the local culture that their phonological form is an essential part of their identity; for example, *galla:f* 'shipwright', *nagi:la* 'seedling'. Substitution of */q/* here would result in a nonce word.

(d) specialization of meaning: some items, although most of their meanings are shared by the dialect and MSA, have certain specialized local uses. An example in Bahraini is the verb *ga:m*, MSA *qa:ma*, whose basic meaning is 'to rise, get up'. When *ga:m* is used with the preposition *bi* in the dialect, it means 'to look after' (a person), but *qa:ma bi* in MSA means 'to undertake' (work, a visit, etc.).

We could summarize the position by saying that the more specialized in reference and connotation a term is, whether it is standard or dialectal, the less susceptible it is to phonological adjustment in the direction of the other kind of Arabic. What about the rest of the lexicon? Where phonological variation between */q/* and its dialectal reflex *does* occur—and where it has particular stylistic significance—is precisely in those words in which there is high degree of semantic overlap and at the same time morphological congruity between MSA and the dialect. These are words that are not closely associated with a specific field of knowledge or cultural context and that are rather general in their meaning, for example (in their Bahraini dialectal form—for Cairene substitute */ʔ/* for */g/*): *ga:bal* 'to meet', *gabal* 'to accept', *ga:l* 'to say', *wa:fag* 'to agree', *farg* 'difference', *gali:l* 'few, little', *gadi:m* 'old', *gari:b* 'near', *'agl* 'mind'. Because speakers have a freer choice in how such words are realized, the use of the dialectal reflex on the one hand or */q/* on the other becomes stylistically significant in a way that it does not in those words and phrases in which */q/* and */g/* are treated by all speakers as invariant elements. The use of */q/* in such culturally neutral words as these is a token of the speaker's desire to sound "correct," "nonlocal," and (often) "authoritative." To recapitulate what was said earlier: not all */q/*s in extempore Arabic speech carry the same degree of stylistic significance.

The figure at the top of the following page illustrates schematically the relationship between word meanings and potential for dialectal/MSA phonological variation posited here. As noted earlier, the "dialectal" variant in this figure is not necessarily a single phoneme: for some urbanizing social groups in Jordan, */ʔ/* and */g/,* and for others */ʔ/* and */g/* and */k/,* are all possible dialectal reflexes varying with MSA */q/.* Switches between these dialectal variants are governed by local, in-country pressures—the desire to sound "sophisticated and urban" versus the desire to identify with a Bedouin "manly" culture (or in some interpretations, to sound "Jordanian" and avoid sounding "Palestinian").[36] By contrast, switches from any of these variants to */q/* has no *purely local* motivation.[37]
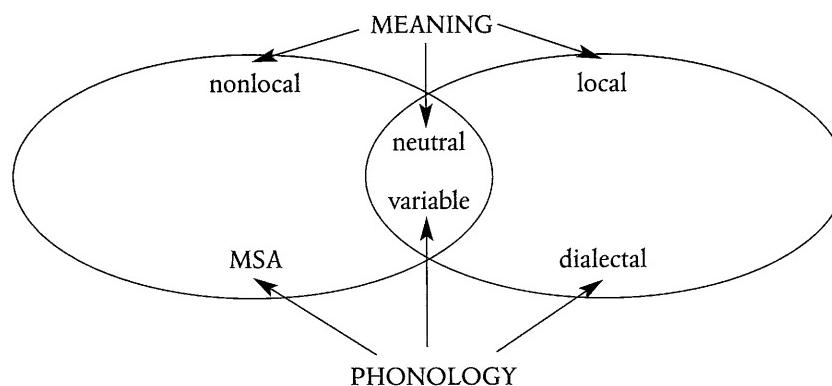
2.3.2 Hybridization and language level

Variation is a phenomenon that knows no neat linguistic boundaries: phonology, morphology, and syntax are all implicated, and our example of /q/ substitution is just one small but important index of stylistic shift that it would be oversimplifying things to consider without regard for other covarying elements. Segmental substitution of the type described is often a contributory process to a more general and complex one that might be termed hybridization: the tendency, which began among the educated but is now general, to compound dialectal and MSA elements—phonological, morphological, and syntactic—at word and phrase level in the "neutral" area of the lexicon in which we saw that the conditions for /q/ substitution are met. In order to amplify our study of /q/, we will look at some actual examples of multilevel covariation within individual words and phrases, leaving until chapter 9 the question of syntagmatic covariation across word, clause, and sentence boundaries.

In educated Cairene Arabic, in different speech contexts, the following phrases could each be used to mean 'he/it is not accepted/admitted (e.g., a person to a hospital, a student to a university; an action or principle)':

(1) *la: yuqbalu*
(2) *la: yuqbal*
(3) *ma: byuqbalš*
(4) *ma: byit?abalš*[38]

(1) and (2) represent combinations of linguistic elements drawn from the MSA side of the linguistic divide, (4) a combination of purely dialectal elements. (3), the linguistic "hybrid," is a mixture. The contributing elements can be arranged in the form of binary choices at different levels of structure:

MSA	CAIRENE
phonology:	
/q/ (as in (1), (2), (3))	/?/ (as in (4))
mood:	
mood ending -*u* (as in (1))	no corresponding ending
negation:	
negative *la:* (as in (1) and (2))	negative *ma: . . . š* (as in (3) and (4))
aspect:	
no aspectual prefix	prefix *b-* (as in (3) and (4))
passivization:	
passive pattern *yuCCaC-* (as in (1), (2), and (3))	passive pattern *yitCaCaC-* (as in (4))

It is clear that these binary choices are constrained with respect to each other. The choice of one element in some cases *implies* categorically the concurrent choice of another element of the same dialectal or MSA variety at another level of structure but does not do so in all cases. So for instance, the choice of the MSA negative does not imply choice of the MSA mood ending; the choice of the dialectal negative does not imply choice of dialectal passive pattern, nor imply anything about the choice of /q/ or /?/; but the choice of the MSA passive pattern does imply the choice of /q/, just as choice of the dialectal passive pattern implies the choice of /?/, and so on. Native speakers reject as "strange" forms such as *la: yu?bal* and *ma: byitqabalš*, which seem to violate unwritten combinatorial rules: one element is "too dialectal" to be combined with another that is "too high-flown."

Table 2.3 illustrates the relationship between (1) and (4) in the form of a scalar arrangement in descending order of formality and "correctness." (1) is the most "correct": the contextual MSA form with full mood endings, such as is normally heard only in the spoken prose of news bulletins, Koranic recitation, or quotation from written text; (2) is the MSA pausal form without mood endings, typical of formal but more relaxed speech, for example, much political speechifying and attempts at public speaking by ordinary people; (4) is plain Cairene colloquial as used by everyone in casual speech; and (3) as the position of the lines might suggest, is the hybrid, typical of careful, educated but extempore speech, for example, an unprepared, unscripted radio interview.

Table 2.3   Levels and Hybridization

	Neg.	Asp.	Verb	Mood	Neg.	
(1)	*la:*	ϕ	*yuqbal*	*u*	ϕ	MSA
(2)	*la:*	ϕ	*yuqbal*	ϕ	ϕ	MSA
(3)	*ma:*	*b*	*yuqbal*	ϕ	*š*	Hybrid
(4)	*ma:*	*b*	*yitʔabal*	ϕ	*š*	Colloquial

The co-occurrence of /q/ with an MSA passive stem on the one hand ((1), (2), and (3)) and /ʔ/ and a dialectal passive stem on the other ((4)) suggests that choices at the level of derivational morphology may be closely linked with phonological ones. In other words, the verb *yuqbal*, which is a fusion of the non-past template *yuCCaC* and the root *QBL*, functions as a unitary morphemic "chunk" in that the choice of constituents that may be fitted into this particular template is restricted to the MSA side of the divide. Similarly, choice of the dialectal template *yitCaCaC* restricts the choice of phonemic elements that may fill the C slots to dialectal ones. (Note, however, that there are some morphological templates that are common to both MSA and dialect—just those cases, in fact, in which simple phonemic substitution, as described in 2.2.1.1–3, is what seems to be going on.) But there are also selectional restrictions on the syntactic frames that these chunks may slot into, which are asymmetrical: the MSA chunk *yuqbal* is free to fill the verb slot in a dialectal syntactic frame—in this case, the negative p-stem *ma: b-* . . . *š*, but *yitʔabal* is not free to slot into an MSA one: **la: yitʔabal* for 'he/it is not accepted' sounds "strange" to native speakers. True hybrids such as *ma: byuqbalš* seem then to arise within discourse whose syntax is dialectal but in which nondialectal lexical elements (because of what is being talked about or who is being talked to) necessarily intrude. They are not random mixes of elements from different self-contained systems but the outcomes of ordered choices that at some points in linguistic structure are relatively free and at others relatively restricted, in circumstances where there is pressure to speak in an "educated" but at the same time "natural" way.[39]

From this single but representative example it is clear that the mechanism and motivation for phonological variation in extempore Arabic speech are quite different from those that work to produce variation in the "read MSA" of news bulletins, which we looked at briefly in 2.1.1. In "read MSA" the dialectal features that sporadically intrude are minor unintentional phonological interferences from the speaker's own dialect. But it is difficult to see how the very different pattern of phonological variation in unscripted educated spoken Arabic can be convincingly explained except in relation to the morphosemantic structure and status of the larger linguistic units (verb stems, nouns) in which it occurs. This should come as no surprise: it is just another kind of conditioning environment, no different in principle from, for example, the rules described by Labov in the 1970s for the reduction of final consonant clusters in the Black English vernacular,[40] which were partially *morphologically* conditioned (whether the final consonant was a marker of the past tense or not).

2.4 Phonology and script

The Arabic alphabet, orthographic conventions, and major script styles are described in the appendix. Here, we are concerned with the degree of fit between Arabic orthography and the phonological structure of spoken MSA. We shall also consider how, in certain kinds of writing, by no means all of which are marginal, the alphabet and orthographic conventions are adapted to reflect dialectal phonology.

2.4.1 MSA phonology and orthography

Compared to a language like English, the match between the phonologically distinctive features of spoken MSA and the alphabet is close and consistent. The Arabic alphabet and the system for marking vowels is, in other words, phonemic. Of the twenty-eight named letters that make up the conventional alphabet, all but one represent consonantal phonemes. The twenty-eighth, *ʔalif*, has no consonantal value of its own but has a number of orthographic uses, one of which is to carry the sign for the twenty-eighth consonant, the glottal stop *hamza*, which is phonologically a fully functional consonant even though it is not normally considered a separate letter of the alphabet.

Arabic writing, which goes from right to left, is cursive, even when printed in books and newspapers. Many of the letters modify their shape, depending on whether they occur at the beginning, in the middle, or at the end of a word.[41] Orthographically, the written Arabic word is essentially a consonantal skeleton: the three qualities of short vowel, which are just as much phonemes as the consonants that make up the written word, are not normally marked in writing except in the Koran and children's reading primers, where they are written above (/a/, /u/) or below (/i/) the letter that they follow. These three vowels are always written in the same manner, whatever phonetic value they have. There are two other superscript orthographic signs that have a purely phonological function (as well as others whose function is primarily syntactic). One of these indicates that a consonant is not followed by a vowel and the other that a consonant is to be pronounced long. Like the short vowels, neither is normally written.

This means that words with quite different meanings, such as *darasa* 'he studied', *durisa* 'it was studied', *dars* 'lesson', *darrasa* 'he taught', and *durrisa* 'it was taught' are homographic in normal handwriting or print. Long vowels are, however, reflected orthographically in the body of the word by the use of *ʔalif* to indicate /a:/, and the letters *ya:ʔ* (= semivowel /y/) and *wa:w* (= semivowel /w/) to indicate /i:/ and /u:/ respectively, so that the long vowels in *mudarrisa:ni* 'two teachers (nom.)', *mudarrisu:na* 'teachers (nom.)' and *mudarrisi:na* 'teachers (acc. and gen.)' are distinguished in spelling as well as in pronunciation. There are one or two words, principally demonstrative adjectives/pronouns and the word *ʔalla:h* 'God' in which /a:/ is not written as a long vowel but is pronounced as one. This is a vestige of the spelling inconsistencies of pre–eighth-century Arabic.

It can be seen from this brief exposé that Arabic writing is almost entirely phonemic. However, the practice of not writing vowels or other orthographic signs with phonological significance (a particular target of educators and script reformers, both Arab and non-Arab) makes it difficult even for educated Arabs to read with complete accuracy unless they have had a thorough grounding in normative MSA.[42]

There are one or two peculiarities in the relationship of MSA phonology to its orthography that will be described here:

2.4.1.1 Hamza, the glottal stop

The anomaly in the convention for writing the glottal stop, *hamza*—that it is a fully functioning consonant yet is not orthographically an independent letter, as the other consonants are—is one consequence of the fact, as Beeston, like many others before him, has observed,[43] that the script partly reflects the phonology of the Arabic spoken in western Arabia at the time of the revelation of the Koran, in which it seems the glottal was no longer a fully functioning phoneme, and partly that of eastern Arabia (which happened to be the dialect of the early codifiers of the language and much of the early poetical canon), where it was a phoneme that could occur in all positions. The orthographic compromise was, and is, that the glottal stop does not appear on the line in the consonantal body of the word—reflecting its absence in ancient western Arabian phonology—but is written above or below other letters that carry it—reflecting the pronunciation with *hamza* that was imposed by the philologists and that has continued to be a feature of spoken and written MSA to the present time. Depending on its function and the surrounding vowels, *hamza* is written on a so-called chair (*kursi:y*) that is (i) sometimes *ʔalif* (and always so when it occurs at the beginning of a word), sometimes *ya:ʔ* (/y/) minus its sub-

script dots, and sometimes *wa:w* (/w/); or (ii) on its own in the middle or at the end of a word.

2.4.1.2 Final -/ay/

Words ending in -/ay/ were pronounced with a final glide in ancient west Arabia,[44] whereas in the east they were pronounced with final -/a:/. Present day orthography, as in the case of *hamza*, mirrors the ancient western pronunciation in spelling such endings with final /y/, but the eastern pronunciation -/a:/ of such words is universal.

2.4.1.3 Case endings

The indefinite case endings *-un* (nom.), *-an* (acc.), and *-in* (gen.) in the commonest declension of the noun are marked in fully voweled script by writing the /u/, /a/, or /i/ vowel marks twice and suffixing an *ʔalif* in the case of *-an* except in the case of grammatically feminine singular nouns morphologically marked as such. In unvoweled script, only the *ʔalif* of the accusative indefinite is obligatorily written. The ancient usage was to pronounce pausal *-an* as -/a:/, a practice that is still followed today in MSA read by more conservative speakers. The other case endings are not pronounced in pause position, as we have seen. Orthographic convention here, as elsewhere in the system (see 2.4.1.4 below), is to ignore the rules for syllabic concatenation and write each word as if it were a prepause form whether or not it in fact is, leaving the reader to supply the correct junctural or pausal features.

2.4.1.4 The article

The Arabic article is /l/ is prefixed to the noun to make it definite. Because, it will be recalled, syllable structure rules demand that all words start with a consonant, the resulting initial cluster needs to be broken up. In MSA, this is done when a defined noun is in postpausal position by prefixing the vowel /a/ with a glottal onset. Thus *bayt* 'house', *ʔalbayt* 'the house'. Orthographically, this glottal is carried by *ʔalif*. However, in juncture, the glottal onset /ʔa/ is redundant: if the preceding word ends in an open syllable, it is closed by the /l/ of the article, and if it ends in a closed syllable, it is supplied with a final vowel. Notwithstanding this, the article is written with *ʔalif* (i.e., as if it were postpausal) wherever in the utterance it occurs.

If a noun begins with a dental, interdental, or alveolar consonant (with the exception of /j/, this exception probably to be explained by the fact that in OA /j/ was pronounced as a palatal, not an alveolar) the article /l/ is assimilated to it, doubling its length. The /l/ is nonetheless still written as /l/.

2.4.1.5 The feminine marker in nouns

Prepausally, the large category of nouns that in the orthography carry the feminine marker are today pronounced with final -/a/, for example, *šarika* 'company', although ancient orthographic practice spelled such words with a final -/h/. In juncture, such words are always pronounced with a final -/t/ before the addition of the case ending: *šarikatu nnafṭ* 'the oil company'. Today's orthography is a compromise that points to both the pausal and junctural phonetic value of the feminine ending: the letter shape used to represent the feminine ending is /h/ (mimicking the original final -/h/, and the -/ah/ pronunciation) but it is written with two dots above it, like the normal letter used for /t/ (in recognition of its phonetic value in juncture).

2.4.1.6 Certain particles and prepositions

Although by any criteria separate words, *wa* 'and', *fa* 'so', *bi* 'in, by, with' and *li* 'to, for' are written as if they were part of the word that follows them. Functionally similar words that are "heavier" monosyllables or bisyllabic, for example, *ʔaw* 'or', *fi:* 'in', *ʿala:* 'on', are not so written. The reason seems to be that *Cv* syllables (which of course consist of a single consonant orthographically) are not viable in Arabic, at least if they are classified according to the indigenous grammatical tradition as *ḥarf* (= prepositions, conjunctive, interrogatory and negative particles, and other function words). Monoliteral verb forms, however, such as *qi* 'protect!' and *ʿi* 'take care!', being verbs, are written separately (although it must be admitted that they are "theoretical" forms not actually used in today's MSA).

2.4.1.7 Redundant ʔalif

Although it is included here, this is a purely orthographic point. At the end of plural verbs ending in -/u:/, orthographically -*w*, a so-called protecting *ʔalif* is written, which has no phonetic value. The reason, it is claimed,[45] is to avoid the possibility of the final -*w* being taken as the *w*- of *wa* 'and', which (see above) is conventionally prefixed to the word that follows it. This is a consequence of the orthographic convention of not marking short vowels. This ensures that *shrw llyl* is read /sahiru: llayl/ 'they stayed up all night' rather than as a homographic phrase, such as *shr wllyl*, i.e., /sahira wa llayl/ 'he stayed up, and the night. . . .'

2.4.2 Dialect and orthography

Native speakers of Arabic often say that their dialects are not written languages. It would be more accurate to say that dialectal Arabic is not felt by native speakers to be an *appropriate* variety of the language for writing in most contexts. However, it has always been true, certainly since the early Middle Ages, that Arabs write some kinds of texts—personal letters, informal notes, and other ephemera not for public consumption—in a style closer to their ordinary speech than the standard Arabic of the day.[46] And there are some kinds of writing for which, if not "written dialect," then a "colloquialized standard" was and is the only appropriate vehicle—popular literature.

The earliest versions of popular literary texts—the *Thousand and One Nights* and the folk-heroic literature of Egypt and North Africa—were written in a style that often diverged markedly from the accepted "literary" syntax and phraseology of the time. In later recensions, however, editors have had a tendency to tidy this up. In the modern period, text types in which dialectal Arabic is used to lend realism or in which the dialectal medium is itself an indispensable part of the message have multiplied: newspaper cartoons and political caricature, children's comics, the scripts of plays, the rendering of dialogue in prose works otherwise written in MSA, and, first and foremeost, collections of folk stories, poetry, and popular literature in general in which, very often, the point is to satirize and ridicule.

In different areas of the Arab world, different traditions for transliterating the local dialect have grown up, and this will be briefly illustrated by examples taken from the Bahraini dialect poet ʿAbdul-Raḥmān Rafīʿ.[47]

Segmentally, Rafīʿs *Qaṣāʾid Shaʿbīya* (*Popular Odes*) employ an eclectic orthography in order to render the phonology of the dialect in as direct and unambiguous a manner as possible. The dialectal consonants, regardless of how their MSA analogs are spelled, are rendered by straightforward phonetic transliteration into the normal script: so /y/ < OA /j/ is rendered by *ya:ʔ* in words such as *yidi:d* (MSA *jadi:d*) 'new', *yi:ra:n* (MSA *ji:ra:n*) 'neighbors', and /j/ < OA /q/ by *ji:m*, for example, *jiri:b* (MSA *qari:b*) 'near', *ʿajil* (MSA *ʿa:qil*) 'wise, clever'. Other OA/MSA phonemes such as the glottal stop *hamza* that have disappeared or changed their function are omitted or spelled to accord with dialectal phonology, for example, *ru:s* 'heads' is spelled without the *hamza* that appears in its MSA analog (*ruʔu:s*) and with a *ya:ʔ* (/y/) in forms such as *ga:yil* 'saying' (MSA *qa:ʔil*). Sounds for which there are no letters in the normal Arabic script are written by using the modifications introduced into Persian and Urdu orthography: /g/ < OA /q/ is written as a "Persian k" (*ka:f ʿajami:y*) in words such as *yigu:l* (MSA *yaqu:lu*) 'he says', /č/ < OA /k/ as a "Persian j" (*ji:m ʿajami:y*) in words such as *sača:či:n* 'knives', and /p/ in borrowings as a "Persian b" (*ba:ʔ ʿajami:y*), for example, *panka* 'fan'. The familiarity of many Bahrainis with the Persian script, if not the language itself, ensures that this purely local practice causes no problems.

Most of the phono-orthographic anomalies of the MSA script noted in the previous section are ironed out:

(a) Nouns and verbs that in MSA variously have orthographic final -*ay*, final -*a?*, final -*a:?*, final -*a:*, and the pausal feminine noun ending -*a* are collapsed into the same orthographic -*ah* to reflect the Bahraini dialectal final short -/*a*/ pronunciation of all such forms, for example, *yirḍa* 'he is content', *yigra* 'he reads', *gaḍa* 'fate', *baladna* 'our country', *mara* 'woman'.

(b) The orthography of the feminine noun ending, which in MSA is a compromise letter pronounced -/*a*/ in pause and -/*at*/- in juncture is brought into line with dialectal morphophonology: thus *mara* 'woman' in always spelled with "pausal" -*ah* in line with its pronunciation, except if it is in a genitive construction with a following noun, in which case it is pronounced and spelled with final -*t*, for example *marat faqi:r* 'a poor man's wife'.

(c) Bahraini Arabic regularly deletes short high vowels in initial *Cv* syllables, optionally prefixing a prosthetic /*?i*/- (see 2.2.2.1), for example, *ybi:ʿ* <(*yibi:ʿ*) 'he sells', *?itḡayyart* (<*tiḡayyart*) 'I changed', *?irkubaw* (<*rikibaw*) 'they rode', *ḍlu:ʿ* (<*ḍulu:ʿ*) 'ribs'. This prosthesis is marked in the orthography by the prefixing of an *?alif* to such forms.

(d) The definite article, as in MSA, is phonologically /*l*/. However, unlike in MSA, postpausally it always has an /*?i*/- onset if the first syllable of the noun is closed, for example, *?ilčalba* 'the bitch'. This non-MSA vocalic onset is marked in the script by a *hamza* written below the *?alif* of the article. If, however, the noun begins with a short open syllable, its vowel is regularly deleted (see (c) above), and an epenthetic *schwa* is inserted after /*l*/, for example, *naxi:l* 'palm trees' but *lənxi:l* 'the palm trees'. Orthographically, this kind of syllabic form is marked by the omission of the *?alif* of the definite article, apparently in order to distinguish it from those that begin with the /*?i*/-onset.

(e) The three CLA/MSA indefinite case endings -*un*, -*an*, and -*in* live on in Bahraini dialect in the form of the single ending -*in* (for all cases) in a few fossilized phrases. Dialect orthography renders this either by a subscript double diagonal line (as this ending would be written if it occurred in fully vowelled MSA text) or, most interestingly, by an orthographic *n* as in *ḥaddin yinsa balad laṣda:f* 'No one can forget the country of shells'.

(f) As in MSA, the short vowels are not normally written. But this is less of a problem for the reader of written dialect for two reasons. First, as a consequence of historical phonological change, much of the syntactic load carried by the unmarked short vowels in OA/MSA has been redistributed in the shape of dialectal morphological distinctions, which are marked in the script. So for example the OA/MSA active versus passive distinction *yaktubu* 'he writes' *yuk-*

tabu 'it is written' that does not appear in unvoweled text is expressed dialectally by a morphological (and orthographic) distinction between *yiktib* 'he writes' and *yinkitib* 'it is written'. Second, the dialect, unlike MSA, is the speaker's native language, which means he or she is able to supply the correct vowels automatically without their being written in.

It should be clear from these points that the effect of the script modifications is to bring the orthography used for writing dialect into even closer correspondence with its phonology than is true in the MSA case. The details of dialectal orthographic systems differ from one dialect to another, but the general aims are the same: to spell the dialectal consonants in the way they are pronounced and reflect, insofar as that is possible in a script that lacks short vowel marks, the salient junctural, pausal, and phonotactic features of the dialect that would be obscured by the standard orthographic conventions.

Notes

1. See al-Nassir 1993 for a study of Sībawayhi's treatment of Arabic phonology. The first mention of Arabic word stress in European grammars of the language is in the early seventeenth century but appears likely to have reflected usage in the contemporary Lebanese Maronite dialect. See Fleisch 1968, 27–28.

2. Badawi 1973, 119–27 lists the main distinguishing characteristics of "religious" style as opposed to "secular" style MSA as evidenced by Egyptian state television programs. See 2.1.6.

3. Harrell 1960, 6–30 describes in detail how Egyptian radio MSA typically deviates from MSA phonological norms as described in standard European descriptions. Diem's 1974 study contains a useful collection of scripted media texts but also includes some unscripted interviews, and from a wider range of locations. Other kinds of program that use a more or less standard kind of Arabic but in this case with a *deliberately* nonstandard phonology are pan-Arab children's television programs such as the Arabic version of *Sesame Street* (*Iftaḥḥ Yā Simsim*).

4. Exceptionally, the /*l*/ phonemes in the word *?alla:h* 'God' are always emphasized except before an /*i*/ vowel, for example, in the phrase *bismilla:h* 'in the name of God'. There are a few other (foreign) words in MSA that normally have "dark" /*l*/, for example, *alma:niy* 'German'.

5. Ferguson 1978, 164.

6. Harrell 1960, 26–30.

7. See, for example, the comparative analysis of emphasis spread in Palestinian and Cairene in Younes 1993.

8. The artificial, compromise nature of MSA pronunciation conventions is well illustrated in this area of phonology. Historically, it seems that the old Arabic dialects either had *ḍ* or *ð̣* but not both—at least, no modern dialect unaffected by the notion of "correctness" has a *ḍ–ð̣* distinction. The *ḍ* pronunciation is associated with sedentary, urban dia-

lects, such as Cairene, and ð with Bedouin ones, such as Muslimi Baghdadi and the dialects of eastern Arabia. Thus Cairene has *ḍarab* 'he hit' *ḍuhr* 'noon', and eastern Arabia has *ðarab* and *ðuhr*. In formal MSA pronunciation, however, a distinction is always drawn between the two: either *ḍaraba*/ *ðuhr* in cases where the speaker's dialect preserved the CLA interdentals θ, ð, and ð̣, or *ḍaraba*/ *zuhr* where the speaker's dialect shifted the interdentals to *t, d,* and *ḍ* to correspond to the orthographic distinction between them. In the latter case, the orthographic θ, ð, and ð̣ of MSA are rendered by the *naḥwi* pronunciations *s, z,* and *ẓ*.

9. /*a*/ is by far the most common, having been shown in a statistical study of a sample of the Koran to occur three times as often as /*i*/ or /*u*/ (Cantineau 1960, 192).

10. In fact, this randomly selected snippet is taken from Syrian television's main evening news bulletin of 9 April 1989.

11. Harrell 1960, 11.

12. Tomiche 1964, 91.

13. Beeston 1970, 21.

14. Harrell 1960, 39.

15. Ibid., 41–42.

16. Ibid., 39.

17. See Parkinson 1991, 54 ff.

18. Adapted from Badawi 1973, 120–22 and Mitchell 1990, 154–55.

19. Abdel-Jawwad and Awwad 1987, 86.

20. For example, Ibrahim 1986.

21. For a study of cross-dialectal "stylistic" switching in Jordan, see Abdel-Jawwad 1986. In a study of 116 women from the northern Jordanian cities of Sult, Ajloun, and Kerak, Al-Wer 2002 shows that a switch is taking place from the local reflex /θ/ of CLA /θ/, with which it is formally identical, toward /t/, the "incorrect" but socially prestigious reflex of Amman and other large Levantine cities. This shift is most pronounced among the educated speakers, that is, those whose education would have exposed them to an awareness of the prestige of /θ/. This is a clear indication of the fact that the phonological change in the dialect, although it may be led by the educated, does not necessarily (or even often) follow MSA phonological norms. See Holes 1986a for similar findings in Bahrain.

22. The historical development of /*k*/ and /*q*/ in the Arabian peninsula is discussed in Johnstone 1963. Holes 1991 elaborates on Johnstone's work in the light of the dialectal data that has since become available. The examples given in this section by no means exhaust the combinatorial possibilities. Parts of Oman and northern Iraq, for example, have preserved historical /*q*/ as the modern dialectal reflex also. These dialects exemplify an older type of sedentary dialect, similar *grosso modo* to the city dialects in other aspects of their phonology.

23. Abdel-Jawwad 1986, 55–61. Sex-based phonological differences have also been described in Haeri 1996 (palatalization of dental stops by Cairene women but not men) and by Abu Haidar 1988 (differences in the degree of emphatization and pharygealization, among other features, between Baghdadi men and women), and by Al Wer for Jordan (see n. 21).

24. It was reported to me on a visit to Yarmouk University, Irbid, in 1989 that first-year women university students from (B) or (C) dialectal backgrounds rapidly shift to the "urban" /ʔ/ pronunciation during their first semester, at least in publicly observable speech contexts.

25. Suleiman 1993 disputes Abdel-Jawwad's analysis, claiming that some at least of the switching needs to be seen against the background of the recent turbulent political history of Jordan. In particular, he claims, salient features of the (B) dialect that none of the competing dialects share, such as the /*k*/ realization of OA/MSA /*q*/, are avoided by (B) dialect speakers precisely because of their locally "Palestinian" and hence (in the Jordanian context) negative associations.

26. Ingham 1982, 31–32.

27. Johnstone 1967, xxviii. Johnstone seems to have overlooked the possibility that a Kuwaiti (or any Arab for that matter) would be most unlikely to speak in his unalloyed colloquial to an Arabic-speaking but manifestly nonnative stranger on first acquaintance. This is a simple matter of linguistic courtesy (would he understand?) and appropriateness (dialect is for peers, not outsiders). For a Kuwaiti to be addressed by an outsider in this way is sufficiently surprising to explain what happened. When foreigners speak Arabic, they are, as it were, not supposed to say *simač*, but *samak*.

28. That is, insofar as MSA has suprasegmental rules that are genuinely independent of the dialects, which in some matters (e.g., positioning and degree of stress) is doubtful. See Janssens 1972 for a comprehensive treatment of stress and word structure in the dialects.

29. Tomiche 1964, 78 notes that, in a continuous speech sample of Cairene Arabic consisting of 9,979 syllables, 83 percent of the total was roughly equally divided between *Cv* and *CvC*, with the other three types *Cv:, Cv:C,* and *CvCC* accounting for the remaining 17 percent between them. Cantineau (1960, 198) claims that in MSA, 75 percent of the syllables are *Cv* and *CvC*, and 25 percent *Cv:*.

30. In a number of recent studies, Angoujard argues for specific hierarchical relationships between syllable type ("heavy/light"), sonority (or vowel height ("low/high")), and metrical structure (the location of "strong" and "weak" syllables within "binary feet" ([+–]/[–+])) as the underlying process in the determination of syllable structure, rhythm, and stress in dialectal Arabic. Interdialectal differences are explained in terms of this model, with metrical differences proposed as the source of the major present-day dialectal differentiation between the syllable structures typical of the Maghreb and the eastern Arab world (1990, 151) and differences in the restrictions on the interaction of syllable type and vowel sonority proposed to explain differences between some dialects within the eastern group (1990, 231–41).

31. Marçais 1956, 123; Holes 1990, 274.

32. Olmsted Gary and Gamal-Eldin 1982, 125; Cowell 1964, 18. In "Sunni" Bahraini, *kitbat* and *katabat*-type forms, stressed in the "urban" way, are now becoming common in educated speakers' speech because of the influence of MSA (see Holes 1986b, 179–80).

33. Marçais 1956, 124.

34. In a study of variation in the speeches of President Nasser (Holes 1993), heavy use of dialect features on the one hand and MSA ones on the other is shown to be associated with particular kinds of ideational content and discourse moves. Concrete and person-

alized descriptions and explanations, as well as "discourse-organizing" and explanatory passages are associated with dialect; abstract, depersonalized descriptions and content presented as authoritative or quasiscriptural truth are in MSA. Allowing for the differences between speech-making and ordinary conversation, in which these elements will be in a much different proportion, there is nonetheless a similar relationship between content and code in ordinary speech among those familiar with MSA norms.

35. Holes 1987, 50–55.
36. See n. 25.
37. Of course, this argument applies only in places where /q/ is not itself a dialectal variant. Where it is, some very similar arguments to those that explain a shift to /g/ and /ʔ/ may explain a switch to /q/. A case in point is Tunisia, where /q/ is the normal dialectal reflex in the city of Tunis, and a shift toward it appears to be in full swing, motivated not by awareness that it is the "correct" MSA form, but that it is emblematic of urban sophistication—see Gibson 2002.
38. All are taken from actual data: (1) and (2) from political speeches, (3) from the Leeds University Corpus of Educated Spoken Arabic, and (4) was provided to me by a Cairene native speaker as the way she would normally express this idea in casual speech.
39. The acceptability of hybrids such as *ma: byuqbalš* and the questionable status of forms such as those asterisked needs to be established with more certainty through the inspection of a large database of natural conversation, followed up by formal tests of the acceptability of nonoccurring forms on representative samples of native speakers. In this way the nature of the mechanism that allows some but disallows other hybrid forms might be revealed.
40. Labov 1972, 216–26.
41. Unsuccessful private attempts were made in the 1950s and 1960s (e.g., by the Unified Arabic Alphabet Association of New York and Beirut) to promote a noncursive version of the alphabet on the grounds that this would help solve the problems of illiteracy and simplify printing by reducing the number of separate shapes required in typesetting from 400 to 30.
42. This does not of course mean that they cannot understand what they read, but rather that they might be unsure of the values of certain phonological features of MSA texts, whose contribution to the overall meaning context renders redundant. This problem increases the less well-versed in MSA the reader is, to the point where understanding is hindered. Bensallama 1971 discusses the social and literary dimensions of this problem in Tunisia.
43. Beeston 1970, 26.
44. Rabin 1951, 116–17, 160.
45. Wright 1896–98, 1:11.
46. Much of the historical material on which Hopkins 1984 is based falls into this category, and Meiseles 1979 illustrates the genre in the modern world.
47. Rafīʿ 1971.

3

Verb Morphology
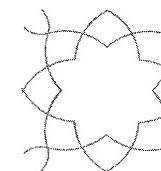
3.1 General principles

The principle of Arabic derivational morphology, which it shares with other Semitic languages, is that of the "root and pattern." The root of most words in the language is a semantic abstraction consisting of three consonants, symbolized here as $C_1C_2C_3$, from which actual words are derived by the superimposition of templatic patterns. Lexical sets are thereby formed that are both structurally and semantically related to the root. For example, the root *KTB* has the broad lexical value 'writing', and from this are derived words for 'book' (*KiTa:B*), 'desk' (*maKTaB*), 'library' (*maKTaBa*), 'correspondence' (*muKa:TaBa*), 'subscription' (*(ʔi)KtiTa:B*), and so on. The root meaning is expressed in its simplest form in the stems of the unaugmented verb, which consists of the root consonants alone and interdigitated vowel patterns.

There are rules and restrictions on the combinations of consonants that can occupy the three positions in the root, as follows:[1]

(a) Sonorants (/l/, /n/, /r/, /w/, /y/) may occur in any position with any other consonant preceding or following.
(b) C_1 and C_2 may not be homorganic.
(c) C_2 and C_3 may be identical, but if they are not, they may not be homorganic.
(d) There are very few roots where C_1 and C_3 are either identical or homorganic.

The patterns that are interdigitated with the roots are of two types, which we will term morphosemantic (3.2) and morphosyntactic (3.3).

The morphosemantic patterns consist of a small set of templates into which the consonants of the root are slotted and that modify the root meaning semantically in (to a large degree) predictable ways that will be made clear below. Morphologically, the templates augment the root by lengthening the middle radical, by inserting a long vowel between radicals, by adding consonantal affixes, and by combinations of these processes. The result, the "augmented" root, is still of course a semantic abstraction, not an actual word: like the unaugmented root, augmented roots become words through the further interdigitation of verbal and nominal morphosyntactic patterns.

Verbal morphosyntactic patterns specify the finite verb for categories such as aspect and voice, and in the noun (see chapter 4) cover a range of categories and functions—agent, patient, instrument, and plural, for example. The resulting combinations of interdigitated elements—root (with or without morphosemantic augmentation) plus morphosyntactic pattern—are then specified for other syntactic categories by affixation: person, number, gender, and mood in the verb (3.4), case and number (in some types of noun) in the noun (chapter 4).

This rather abstract sketch of the overall structure of Arabic noun and verb morphology is filled out below, first for MSA, and then (3.6) for dialectal Arabic.

3.2 Root and morphosemantic patterns: MSA

In MSA, nine "augmented" verbal morphosemantic patterns are in common use and can theoretically apply to any root (regardless of whether it has the most basic "Pattern I" verb of the type exemplified below). In practice, however, the maximum number of patterns found in any root is seven or eight, and most roots show fewer than this number. These augmented patterns, usually referred to in the orientalist tradition by Roman numerals, are as given below. C stands for radical consonant. Examples of patterns are printed in upper case to signify that they are abstractions, not actual words.

3.2.1 Pattern I: The root pattern *C1C2C3*

This, the unaugmented verbal root, is the basis of Arabic derivational morphology, both in the abstract sense and in the concrete one of frequency of occurrence. A recent analysis of the finite verb forms in a large corpus of literary, journalistic, documentary, and school texts showed 43,543 tokens representing 3,127 different verbs.[2] Of the 474 commonest verbs that occurred in the corpus twenty times or more, 224, or just under half, were Pattern I, with the other

nine common patterns (plus quadriliterals) accounting for the remaining 250 between them.

The root becomes a Pattern I verb through the interdigitation of short vowels between its consonants. In the suffix stem, the vowel patterns are, in descending frequency of occurrence in the lexicon: *a–a*, *a–i*, and *a–u*. Generally speaking, only one of these vowel patterns applies to any given root. The vowel patterns themselves are broadly associated with different categories of transitivity and dynamic versus stative meaning. The *a–a* pattern generally denotes an action, transitive or intransitive, performed by an agent, for example, *Fa'aL* 'do, make', *XaRaJ* 'go out', *DaXaL* 'go in', *ḌaRaB* 'hit'.

The *a–i* pattern also frequently denotes actions, transitive and intransitive, but ones in which, in Fleisch's words, the agent is *agent moyen*: not an agent pure and simple but one that affects itself in some way by the performance of its action.[3] Hence *XaSiR* 'lose', *WaJi'* 'feel pain', and *'aJiL* 'go quickly' all have the *a–i* pattern, as do many bodily functions, for example, *BaLi'* 'swallow', *šaRiB* 'drink', *Ṭa'iM* 'taste'; bodily attitudes and movements, for example, *TaBi'* 'follow', *RaKiB* 'mount', *XaṬiF* 'grab'. "Mental" verbs and those denoting contingent physical and emotional states, also have the *a–i* pattern, for example, *'aLiM* 'know', *FaHiM* 'understand', *SaMi'* 'hear', *NaSiY* 'forget', *RaGiB* 'desire', *XašiY* 'fear', *ĠaḌiB* 'feel angry', *FaRiḤ* 'feel happy'.

The *a–u* pattern is always intransitive and denotes the possession or acquisition of a quality that is permanent: *ḤaSuN* 'be/become good, beautiful', *QaBuḤ* 'be/become ugly', *SaĠuR* 'be/become small, young', *KaBuR* 'be/become large, great, old'.

3.2.2 Pattern II: Lengthen the middle radical = *C1C2C2C3*

Pattern II can modify the root meaning in two main ways: by making it more intensive/extensive or (in the vast majority of cases) by indicating that another agent caused it to happen, for example, (intensive/extensive) *JM'* 'collect', *JMM'* 'amass, pile up', *NBŠ* 'unearth, dig up' *NBBŠ* 'ransack'; (causative) *QDM* 'precede, go before' *QDDM* 'put forward', *'LM* 'know', *'LLM* 'teach' (i.e., cause to know). In a few roots, Pattern II is estimative or ascriptive, for example, *KBR* 'be great' *KBBR* 'praise' (i.e., declare to be great, cf. 'magnify' in the sense employed in the *Book of Common Prayer*: "My soul doth magnify the Lord"), *ṢDQ* 'be truthful', *ṢDDQ* 'believe' (i.e., ascribe or impute truth to).

In MSA, the causative meaning potential of Pattern II has been put to particularly heavy use in the coining of denominative verbs for scientific and other academic and modern purposes, in a way similar to the use of the English suffix '-ize', for example, from *'QM* 'sterile, barren' *'QQM* 'sterilize'; from *QṬR*

'dribble, drip', *QṬṬR* 'distil' (i.e., purify by causing to drip); from the passive of *ṬBʿ* 'be native, natural', *ṬBB* 'normalize'; from the noun *ha:miš* 'margin' *HMMŠ* 'marginalize'; from the noun *dawla* 'state' *DWWL* 'internationalize'. Some Pattern II roots that in older written Arabic had an intensive or extensive meaning (in any case relatively rare meanings for Pattern II) are now exclusively used in MSA in a causative sense, for example, *MWT* 'die' gives *MWWT*, which formerly meant 'die in large numbers' but in MSA is used only to mean 'cause to die', and *NWWM*, the Pattern II from *NWM* 'sleep' now means 'tranquilize, put to sleep' (i.e., with drugs) rather than 'sleep deeply'.[4]

3.2.3 Pattern III: Lengthen the short v between $C1$ and $C2$ = $C1v{:}C2C3$

The basic meaning of Pattern III is conative, that is, it denotes the making of an effort to achieve the Pattern I root meaning. So *QTL* 'kill' gives *Qv:TL* 'try to kill' or 'fight'. The involvement of a patient or of a participant in the action other than the subject is often implied by this pattern, so *KTB* 'write' *Kv:TB* 'correspond with', *ʿML* 'do' *ʿv:ML* 'treat', *QBL* 'move forward' *Qv:BL* 'meet', *MŠY* 'go, walk' *Mv:ŠY* 'keep abreast of, keep up with'.

3.2.4 Pattern IV: Prefix $ʔ$ = $ʔC1C2C3$

Pattern IV usually gives a causative or factitive meaning to the root: (causative) *FRḤ* 'happy' *ʔFRḤ* 'gladden'; (factitive) *NZL* 'go down' *ʔNZL* 'bring down'. In some roots there is an overlap with the causative meaning of Pattern II, for example, *XBR* 'know' gives *XBBR* and *ʔXBR* with the same meaning 'apprise, notify', whereas in others there is a difference, for example, *ʿLM* 'know' gives *ʿLLM* 'teach' but *ʔʿLM* 'inform'. There is also a subgroup of Pattern IV inchoative denominatives, for example, from *WRQ* 'leaf' *ʔWRQ* 'burst into leaf', *MṬR* 'rain' *ʔMṬR* 'rain' (cf. Fr. *faire du vent*, etc.), *SRʿ* 'speed' *ʔSRʿ* 'make haste'.

3.2.5 Pattern V: Prefix t to Pattern II = $tC1C2C2C3$

If Pattern II is most often transitive and causative, Pattern V most often expresses the effect of an action on its subject or is reflexive: thus Pattern II *ʿLLM* 'teach' gives Pattern V (effective) *tʿLLM* 'learn, become learned', and *JMMʿ* 'amass, pile up' gives (reflexive) *tJMMʿ* 'come or be brought together in masses'. Pattern V sometimes has an additional overtone of potentiality, so *KSSR* 'smash' gives *tKSSR* with the sense of 'be smashable' as well as 'be smashed'. As in the case of *ʿLM*, many roots show a regular set of stative (I)—causative (II)—effective/reflexive (V) semantic relationships. For example:

I	
stative	
ḤSN	'be good'
ḤQQ	'be true'
XFF	'be light'
MθL	'appear'

II	
causative	
ḤSSN	'make better'
ḤQQQ	'carry out'
XFFF	'lighten'
MθθL	'represent'

V	
effective/reflexive	
tḤSSN	'become better'
tḤQQQ	'materialize'
tXFFF	'unburden oneself'
tMθθL	'be represented'

3.2.6 Pattern VI: Prefix t to Pattern III = $tC1v{:}C2C3$

Both in form and semantic relationship, Pattern VI is to Pattern III as Pattern V is to II. Both often imply reflexivity. The difference between Patterns III and VI is simply that VI *necessarily* implies reciprocal action on the part of the participant. Examples: *Wv:FQ* 'agree to' (i.e., one party agrees to a proposition) gives *tWv:FQ* 'come to an agreement' (several parties agree by negotiation), *ʿv:WN* 'help' gives *tʿv:WN* 'cooperate', *Sv:BQ* 'compete with' gives *tSv:BQ* 'compete with one another'. The subjects of "reciprocal" Pattern VI verbs are necessarily plurals or collectives.

Pattern VI is also used to indicate stative relationships holding between the component parts of an entity as well as reciprocal action between different entities, for example, from *MSK* 'hold' *tMv:SK* 'hold together, cohere', from *KML* 'be complete' *tKv:ML* 'be complementary, integrated'. In some roots it indicates continuous or iterative action: *SQṬ* 'fall' *tSv:QṬ* 'fall continuously (e.g., missiles, rain)', *TBʿ* 'follow' *tTv:Bʿ* 'follow on after another', *RJʿ* 'go back' *tRv:Jʿ* 'recede' (i.e., to go back continuously). In a few roots there is a simulative sense to Pattern VI: *JHL* 'be ignorant' *tJv:HL* 'feign ignorance', *ðHR* 'appear' *tðv:HR* 'pretend'.

3.2.7 Pattern VII: Prefix *n* to Pattern I = *nC1C2C3*

Pattern VII is most commonly the passive correlate of Pattern I: *ḤLL* 'undo, untie' *nḤLL* 'be untied, dissolved', *QLB* 'overturn' *nQLB* 'be overturned, overthrown'. In some roots, Pattern VII has a signification similar to the reflexive–effective sense of Pattern V: it focuses on the effect of an action without the existence of a causative agent being directly implied: *KŠF* 'reveal' *nKŠF* 'come to light', *SḤB* 'pull' *nSḤB* 'withdraw, leave', *QṬʿ* 'cut' *nQṬʿ* 'cease, end', *XFḌ* 'lower' *nXFḌ* 'subside' (as well as the passive sense 'be lowered').

3.2.8 Pattern VIII: Insert *t* after *C1* = *C1tC2C3*

This pattern often has a reflexive sense similar to those of Patterns V and VII (indeed in some roots the senses of Patterns V and VIII are synonymous or nearly so): *JMʿ* 'collect' *JtMʿ* 'assemble, be gathered together' (*tJMMʿ* with the same sense), *NQL* 'move (something)' *NtQL* 'move oneself' (*tNQQL* with the same sense); *NḤR* 'slaughter' *NtḤR* 'kill oneself', *ḌRB* 'beat' *ḌtRB* 'become agitated' (i.e., beat and flap about oneself), *HMM* 'distress' *HtMM* 'concern oneself about', *BʿD* 'be distant' *BtʿD* 'distance oneself'. In some roots there is a benefactive sense discernible: *KSB* 'gain' *KtSB* 'earn' (i.e., gain for oneself).

3.2.9 Pattern IX: Lengthen *C3* = *C1C2C3C3*

This pattern is the rarest of those in normal use and is a closed class of inchoative verbs, applying only to roots that denote colors and physical defects, for example, *ḤMR* 'red' *ḤMRR* 'blush, turn red', *ḤWL* 'squint' *ḤWLL* 'be squint-eyed'.

3.2.10 Pattern X: Prefix *st* = *stC1C2C3*

Pattern X most often is the reflexive or reflexive-benefactive correlate of causative Pattern IV:

IV

ʔʿDD	'prepare'
ʔSLM	'deliver up'
ʔQWM	'set up straight'
ʔʿLM	'inform'
ʔʿML	'cause to work'

X

stʿDD	'prepare oneself'

stSLM	'surrender oneself'
stQWM	'stand oneself up straight'
stʿLM	'get information for oneself'
stʿML	'use' (i.e., cause something to work for one's benefit)

Two other types of Pattern X meaning are estimative: *ḤSN* 'be good' *stḤSN* 'approve', *XFF* 'be light in weight' *stXFF* 'scorn, disdain' (i.e., deem to be lightweight), *BʿD* 'be distant' *stBʿD* 'regard as unlikely', and eductive (i.e., eliciting or seeking): *ĠFR* 'pardon' *stĠFR* 'ask for pardon', *RXṢ* 'permit' *stRXṢ* 'ask for permission', *NJD* 'help' *stNJD* 'ask for help'.

3.2.11 Other patterns: X–XV

There are five other morphosemantic patterns, mostly "intensive" in the meaning extension they bring to the root, but these are rare even in CLA and are hardly used at all in MSA.

3.2.12 Quadriliteral roots

These are common. The root consonants pattern in the following ways:

(1) *C1C2C3C4*. Examples: *TRJM* 'translate', *SYṬR* 'control', *DḤRJ* 'roll', *HRWL* 'hurry', *BḤRN* 'Bahraini-ize', *BSML* 'say *bismilla:h* ('in the name of God').

(2) *C1C2C1C2*. Examples: *QʿQʿ* 'clatter', *ZLZL* 'shake, convulse', *ṢʿṢʿ* 'shine', *QṢQṢ* 'break into small pieces', *ĠMĠM* 'mumble'.

(3) *C1C2C1C3*. Examples: *BRBṬ* 'paddle, splash', *QHQR* 'retreat'.

(4) *C1C2C3C3*. Examples: *ZĠLL* 'dazzle'.

In MSA, Types (1) and (2) are far more common than (3) and (4) (although (3) is not uncommon in the dialectal Arabic of Syria–Lebanon and North Africa—see below). Most quadriliterals were derived historically by various processes of extension of triliteral roots: through a repetition of *C1*: *QHQR* 'retreat' < *QHR* 'defeat'; through dissimilation of a lengthened *C2*, where the intrusive consonant may be /r/, /l/, /n/, /y/ or /w/: *FRQʿ* 'crack' < *FQQʿ* 'crack'; through the addition of *C4*: *ŠMXR* 'be lofty' < *ŠMX* 'be high'. Type (2) verbs are reduplicatives and mostly derived from triliteral roots whose meanings they intensify: *JRJR* 'drag' < *JRR* 'pull', *QṢQṢ* 'break into small pieces' < *QṢṢ* 'break, cut'.

Quadriliterals have also been, and continue in the modern period to be, derived from foreign borrowings: *FLSF* 'philosophize', *TLFZ* 'televize', and *DBLJ* 'dub' (< Fr. *doublage*) and through secondary denominative derivation from

within Arabic: *NṢRN* 'Nasserize' (after the late president of Egypt), *BḤRN* 'Bahraini-ize'. A further, ancient use is in the abbreviation of religious phrases, as in the example *BSML* above.

Some quadriliteral roots show augmented patterns. The commonest pattern prefixes a *t*, which, like Pattern V of the triliteral root, makes the meaning reflexive: *ZLZL* 'shake' *tZLZL* 'shake (intrans.)', *DHRJ* 'roll' *tDHRJ* 'roll (intrans.)'. This pattern is also used in deriving secondary denominative verbs, for example, *tMRKZ* 'become concentrated' < *markaz* 'center' < *RKZ* 'firmly establish', *t2MRK* 'become American, americanized' < *ʔamri:ka:* 'America', *tBRṬN* 'become English, anglicized' < *bari:ṭa:nya:* 'Britain'.

3.3 Morphosyntactic patterns of the finite verb: MSA

The first distinction to be drawn is between patterns that are applied to the root (augmented or not) to generate finite verbs and those that generate nouns. The terminology "verb" and "noun" is however somewhat misleading, because certain common categories of words that are morphologically "nouns" can function syntactically like verbs, as we shall see in chapter 4.

There are two kinds of finite verb stem in Arabic: the prefix-stem and the suffix-stem. What we are designating the prefix-stem (hereafter p-stem) is specified for person, number, and gender by a set of discontinuous bound affixes. The p-stem is also marked for syntactic function (corresponding roughly to mood in European languages). It has a wide range of temporal/aspectual uses (see chapter 6), for which reason we prefer to avoid the potentially misleading label "imperfect" and simply refer to it by its morphological characteristics. For the same reasons, we refer to what is usually termed the "perfect" or "preterite" verb-stem by the neutral morphological label suffix-stem (s-stem).

The p-stem and s-stem each have two sets of internal vowel patterns for active and passive voice. In summarized tabular form these are (v = vowel):

3.3.1 Pattern I

		active		passive	
		v_1	v_2	v_1	v_2
p-stem:	-vCCvC-	a	a / i / u	a u	u	a
s-stem:	CvCvC-	a	a	i u	u	i

It can be seen from this that:

(a) v_1 is always /a/ in the active voice and /u/ in the passive

(b) s-stems in which the "theme" vowel, v_2, is /a/ (the commonest category, generally denoting actions that impinge on the world outside the agent performing them) can have p-stems whose theme vowel is /a/, /i/, or /u/. However, if C_2 or C_3 is a guttural (/x/, /ġ/, /ḥ/, /ʕ/, /ʔ/, /h/), the p-stem vowel is normally /a/. For example:

s-stem	p-stem	
FaʕaL-	*-aFʕaL-*	'to do, make'
ḌaRaB-	*-aḌRiB-*	'to hit'
KaTaB-	*-aKTuB-*	'to write'

(c) in the other two types of s-stem, which generally designate actions and states that impinge on the agent alone, the theme-vowel relationship with the p-stem is more predictable: with a couple of exceptions, the /i/ theme vowel in the s-stem corresponds to /a/ in the p-stem, and /u/ invariably corresponds to /u/. For example:

s-stem	p-stem	
FaRiḤ-	*-aFRaḤ-*	'to be happy'
LaBiS-	*-aLBaS-*	'to wear'
KaBuR-	*-aKBuR-*	'to grow old, great'
QaṢuR-	*-aQṢuR-*	'to become short, insufficient'

(d) in Pattern I verbs whose transitive meaning makes them susceptible to passivization,[5] v_1 is /u/ in both types of stem, and v_2, the theme vowel, is /i/ in the s-stem and /a/ in the p-stem:

s-stem	p-stem	
FuʕiL-	*-uFʕaL-*	'to be done'
ḌuRiB-	*-uḌRaB-*	'to be hit'
KuTiB-	*-uKTaB-*	'to be written'

It is noteworthy that the /i/ and /a/ theme vowels of the s- and p-stem passive patterns are the same as those of the active stems of verbs that designate actions and states that impinge on the agent alone. This formal parallel reflects a semantic similarity between the two, viz. in both cases the entity of which the verb is predicated is a patient: a patient only in the case of the passive stem such *ḌuRiB-* 'be hit', a patient as well as an agent in verbs such as *LaBiS-* 'wear', *FaHiM-* 'understand', *ŠaRiB-* 'drink', *FaRiḤ-* 'be happy'.

3.3.2 Augmented patterns

The vowel patterns of the augmented stems can be divided into four distinct groups:

(i)   VII, VIII, IX, X:[6]

	p-stem	s-stem
active	*a–a–i*	*a–a*
	↓	↓

$$\boxed{\begin{array}{l} -vnCvCvC- \\ -vCtvCvC- \\ -vCCvCvC- \\ -vstvCCvC- \end{array}} \qquad \boxed{\begin{array}{l} nCvCvC- \\ CtvCvC- \\ CCvCvC- \\ stvCCvC- \end{array}}$$

	↑	↑
passive	*u–a–a*	*u–i*

(ii) V, VI:

active	*a–a–a–a*	*a–a–a*
	↓	↓

$$\boxed{\begin{array}{l} -vtvCvCCvC- \\ -vtvCv:CvC- \end{array}} \qquad \boxed{\begin{array}{l} tvCvCCvC- \\ tvCv:CvC- \end{array}}$$

	↑	↑
passive	*u–a–a–a*	*u–u–i*

(iii) II, III:

active	*u–a–i*	*a–a*
	↓	↓

$$\boxed{\begin{array}{l} -vCvCCvC- \\ -vCv:CvC- \end{array}} \qquad \boxed{\begin{array}{l} CvCCvC- \\ Cv:CvC- \end{array}}$$

	↑	↑
passive	*u–a–a*	*u–i*

(iv) IV: [7]

active	*u–i*	*a–a*
	↓	↓

$$\boxed{-vCCvC-} \qquad \boxed{2vCCvC-}$$

	↑	↑
passive	*u–a*	*u–i*
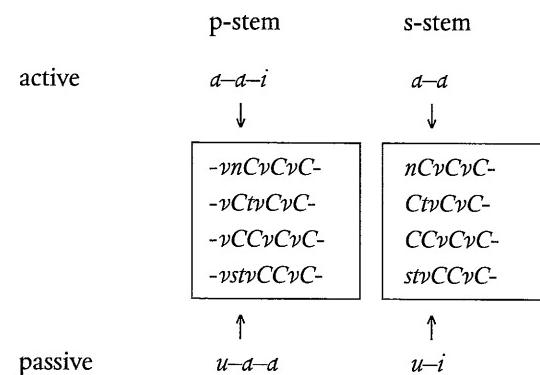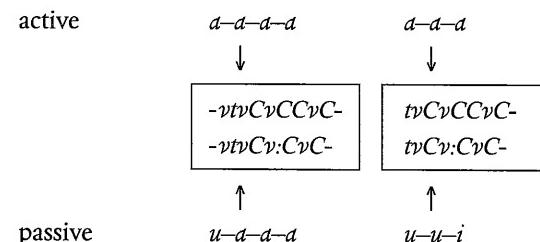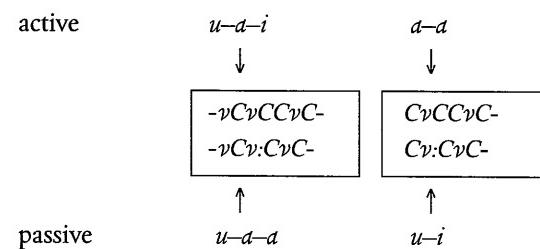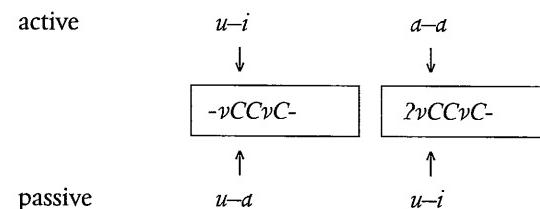
There are a number of structural parallels here with the Pattern I vocalization system that suggest that the augmented pattern system may be an extension of it:

(a) The active s-stem *a–a/ a–a–a* is the same as the main Pattern I active stem *a–a*.

(b) The passive s-stem vocalization *u–i/ u–u–i* echoes Pattern I *u–i*.

(c) The passive p-stem vocalization *u–a/ u–a–a* parallels the Pattern I vocalization *u–a*.

(d) The active p-stem shows a variety of vocalization patterns, as does that of Pattern I.

3.4 Verbal affixes

Finite verb stems, augmented or unaugmented, are specified for person, gender, and number (in which singular, dual, and plural (= more than two) are separate categories) by two sets of affixes, as in table 3.1.

The paradigm for the p-stem (minus the elements in brackets) is morphologically the base set of forms into which the actual p-stem itself is slotted in place of " . . . ." The commonest uses of it in MSA are after negative particles:

Table 3.1   CLA/MSA Verbal Affixes for Person, Number, Gender

Person/Gender	P-STEM		
	Number		
	Sg.	Pl.	Dual
1	*2- . . .*	*n- . . .*	(none)
2m.	*t- . . .*	*t- . . . -u:(na)*	*t- . . . -a:(ni)*
2f.	*t- . . . i:(na)*	*t- . . . -na*	*t- . . . -a:(ni)*
3m.	*y- . . .*	*y- . . . -u:(na)*	*y- . . . -a:(ni)*
3f.	*t- . . .*	*y- . . . -na*	*t- . . . -a:(ni)*

Person/Gender	S-STEM		
	Number		
	Sg.	Pl.	Dual
1	*. . . -tu*	*. . . -na:*	(none)
2m.	*. . . -ta*	*. . . -tum*	*. . . -tuma:*
2f.	*. . . -ti*	*. . . -tunna*	*. . . -tuma:*
3m.	*. . . -a*	*. . . -u:*	*. . . -a:*
3f.	*. . . -at*	*. . . -na*	*. . . -ata:*

after *lam* for completed actions and states in the negative and after *la:* for negative commands, for example, from Pattern I verbs: *-aFRaH-* 'be happy' *lam ʔafraḥ* 'I wasn't happy'; *-aKTuB-* 'write' *lam yaktubu:* 'they didn't write', *la: taktub* 'don't write!'; *-uDRaB-* 'be hit' (with passive vowel pattern) *lam tuḍrabi:* 'you (f.) weren't hit'; from Pattern V active 'speak' *-atKaLLaM- lam tatakallama:* 'you two didn't speak'; from Pattern X passive 'use' *-ustaʿMaL- lam yustaʿmal* 'it wasn't used'. This base set of p-forms, minus the personal preformative (= prefix plus vowel), also forms the positive command form.[8]

There are two other sets of p-stem forms that we term the a-set and the u-set. The u-set differs from the base set by adding a *-na* or *-ni* suffix (bracketed above) to forms that have a long vowel suffix. The a-set adds a final *-a* to all forms that are suffixless in the base set, whereas the u-set adds a final *-u*. The main use of the a-set is in the description of actions and states that may or may not actually have come to pass but that, within the time frame in which they are located, are/were in the realm of the possible, the intended, wanted, or striven after.[9] The a-set verb occurs mainly in subordinate clauses and is not dissimilar to the subjunctive in French. It is always the automatic result of the use of certain governing particles and is therefore less "independent" than the u-set verb. The u-set verb usually occurs as the main verb in declarative statements of timeless validity or in describing actions and states that are/were/will be ongoing (habitual, durative, repetitive) rather than ones that have been completed once and for all. If such actions involve future intention, the particle *sa-* or *sawfa* is often prefixed to the u-set, for example, *sayaktubu* 'he will write'. Negative future intention (or prediction) is expressed by *lan* + a-set, for example, *lan nastaslima* 'we shall/will not surrender'.

The s-stem, active or passive, is slotted directly into the s-stem set of suffixes: Pattern I active *katabna:* 'we wrote'; Pattern II passive *ʿullima* 'he was taught'; Pattern VII *nqalabu:* 'they were overthrown', Pattern VIII active *štaʿaltu* 'I worked'. S-stem verbs are used to cover a wide range of meanings (chapter 6), but the main difference with the p-stem is that s-stem verbs depict the action or state they describe (whether in the past or present, or projected into future time) as completed or ended.

3.5 Morphophonological adjustments

Historically, certain combinations of consonants and vowels that resulted from the processes of interdigitation and affixation described above became subject to adjustment because they violated phonotactic rules. Impermissible strings arose in two ways:

(a) where one of the semivowels /w/ or /y/ occupied one of the C positions in any word, whether derived from Pattern I or from certain of the augmented patterns;
(b) where $C_2 = C_3$.

However, statement (a) needs to be qualified, since synchronically the susceptibility of strings containing semivowels to the adjustment rules is not the same in all roots, not the same between different word classes within the same root, and not the same as compared between strings from different subsets of the same verb stem. The reasons for much of this are historical: some varieties of ancient Arabic had "strong" (fully consonantal) semivowels and some had not, and present-day MSA contains morphophonological and lexical vestiges of this historical difference. Be that as it may, some of the phonological processes that affect the semivowels /w/ and /y/ have in various ways become morphologically conditioned in MSA, as we shall now show with some examples.

3.5.1 Semivowels

If /w/ or /y/ occurs singly, and especially intervocalically, in an underlying string, the string is adjusted in various ways. If they are lengthened or preceded by a long vowel (e.g., as C_2 in Patterns II, III, V, VI), /w/ and /y/ behave as "strong" consonants. /w/ and /y/ as C_1 are also "strong" in all patterns except I and VIII, in which /w/ is "weak."

3.5.1.1 C_1 = semivowel ($)

In the Pattern I s-stem, such verbs behave like normal strong verbs, for example, from root *WRθ* 'inherit' *wariθtu* 'I inherited', from root *YQN* 'know for sure' *yaqinat* 'she knew for sure'. But in the p-stem (base, a-, and u-sets) active voice and in the command form, some verbs in which /w/ is C_1 drop /w/. For example:

from *WSL* : *yaṣilu* 'he arrives' not *yawṣilu; ṣil* 'arrive (m.)!' not *ʔiwṣil*
WDʿ: *yaḍaʿu* 'he puts' not *yawḍaʿu; ḍaʿ* 'put (m.) !' not *ʔiwḍaʿ*

but:

from *WJʿ*: *yawjaʿu* 'he is in pain'
WJZ : *yawjazu* 'it is short'

It is noticeable that it is mainly the verbs of action, not those denoting contingent states or durable qualities (3.2.1), that have the shortened p-stem. Fleisch[10]

has speculated that the historical reason for this is that the imperative, which in most languages including Arabic is usually the most morphologically abbreviated verb form and which also by its nature is used exclusively with action verbs, is the source of the morphological difference we see in MSA: historically, from an original *ḍaʿ* 'put!' the p-stem forms *yaḍaʿu*, etc., developed by analogy with those of what we have come to regard as the "strong" verb and later, by further analogy, the initial /w/ was added to the s-stem forms to bring these verbs into line with the triradical system of morphology. But verbs with initial /w/ denoting states and qualities, in which the imperative is for obvious reasons a "theoretical" form only, developed in the same way as other verbs of the same morphosemantic category (the verbs that have *a–i* and *a–u* voweling in the s-stem). Alternatively, it can plausibly be argued that the CLA/MSA forms are the mixed result of amalgamating ancient dialectal forms that treated verbs with initial /w/ differently.[11]

Whatever the merits of these arguments, synchronically the /w/ initial verbs in MSA Pattern I verbs divide into two categories: those with the p-stem /i/ theme vowel that do not have a *C1* in the p-stem active (mainly transitive), and those that have the p-stem /a/ or /u/ theme vowel that have *C1* as a "strong" /w/ in the p-stem (mainly intransitive, expressing contingent states or durable qualities). In the passive, the missing /w/ returns, for example, from *WJD* we get *yajidu* 'he finds' but *yuwjadu* 'he/it is found' (sc. 'there is'). Such passive forms are explicable historically on the basis of an analogy with the morphology of the strong verb; synchronically, their structure provides a neat parallel with that of the active form of the stative and quality verbs with which they share the semantic feature of agentlessness.

In Pattern VIII, /w/ is assimilated to the inserted /t/:[12] thus from root *WṢL*, Pattern VIII s-stem active *ttaṢaL-*, p-stem *-attaṢiL-* 'contact, communicate', not *ʿWtaṢaL-*, *ʿ-aWtaṢiL-*. Although (anciently) phonological in origin,[13] this change is fully morphologized in MSA, affecting only historical /w/ as *C1* when in contact with inserted "morphological" /t/. In other morphological contexts, contiguous /w/ and /t/ do not behave in this way, for example, *ṣawt* 'voice', *ʾawtaːd* 'tent pegs'.

3.5.1.2 C3 or C2 = $

If v_2 in v_1\$$v(:)_2$ occurs in open syllable, the following adjustments generally occur in both verbs and nouns:

in *v\$v*:

a\$a → /aː/	i\$a → /iya/	u\$a → /uwa/
a\$i → /aː/	i\$i → /iː/	u\$i → /iː/
a\$u → /aː/	i\$u → /iː/	u\$u → /uː/

That is:

Rule (1): low v_1 assimilates any v_2.
Rule (2): low v_2 is stable.
Rule (3): high front v assimilates high back v in all positions.
Rule (4): high back v_2 assimilates high back v_1.

in *v\$v*:

a\$aː → /a\$aː/	i\$aː → /iyaː/	u\$aː → /uwaː/
a\$i → /ay/	i\$i → /iː/	u\$i → /iː/
a\$u → /aw/	i\$u → /uː/	u\$u → /uː/

That is:

Rule (5): low v_1 reduces any high $v:_2$.
Rule (6): low $v:_2$ is stable.
Rule (7): high $v:_2$ assimilates high v_1.

From the patterning of these assimilatory and reductional processes, the relative strength of the vowels emerge and seems to correspond to degree of sonority.[14] The "strongest" vowel in terms of its stability, assimilatory and reductive power, and sonority is /a/, followed by /i/, with /u/ the "weakest." The same rank order applies to the semivowels that correspond to these vowels: in those strings where final /a/ or /aː/ exercises a stabilizing influence, "weak" /w/ in *ʿiwa* and *ʿiwaː* is replaced by "stronger" /y/. Thus, historically, the root consonants *RḌW* 'be content' have given Pattern I s-stem active *raḍiya* 'he was content', not *ʿraḍiwa*; synchronically, root *ĠZW* 'raid' gives the Pattern I s-stem passive dual *ġuziyaː* 'the two were attacked', not *ʿġuziwaː*.

The rules laid out above admit of a few exceptions in the verb: -/awi/- occurs in *ʿaWiR-* 'be one-eyed', *ʿaWiZa* 'be needy', and a few others, which are most likely explained as late denominative derivations (c.f. *ʾaʿwar* 'one-eyed', *ʿawz*, *ʿawaz* 'destitution, need'). Minimal pairs (in MSA admittedly not that many) can also be adduced in which -/awa/-, -/iwa/- and -/iwaː/- remain stable in nominal forms in an environment in which they are invariably reduced in verbal ones. For example, from the root *XWN* 'behave traitorously' we get, by regular rule, *xaːnat* < *ʿxawanat* 'she was traitorous' but the plural noun *xawana* 'traitors'; from *JWR* 'oppress' we get *jaːrat* < *ʿjawarat* 'she oppressed' but the plural noun *jawara* 'despots'. But more important than in these somewhat marginal survivals, the semivowel deletion and vowel assimilation rules are ignored where, if they were applied, important morphological differences would collapse. This suspension of otherwise general phonological rules in particular areas of morphology where if applied they would eliminate distinctions is a regular "design feature" of MSA. An example within verb morphology is given directly below, and in 3.5.2 we see how strings of identical root consonants separated

only by a short vowel are treated differently depending on whether they are nouns or verbs.

An example of the maintenance of morphological distinction through the "selective" application of phonological rules is found in the p-stem of verbs that have $C_3 = \$$. In the strong verb, the following is the standard set of base set, a-set, and u-set, for example, (from p-stem *-aKTuB-* 'write', 2msg./3fsg.):

base	a-set	u-set
taktub	*taktuba*	*taktubu*

In the a- and u-sets of verbs where $C_3 = \$$, forms arise that undergo the rules stated above regularly, for example (from *-aRMiY-* 'throw', *-aBQaY-* 'remain', *-aD'uW-* 'call'):

a-set		u-set	
tarmiya	(Rule 2)	*tarmi:* <*tarmiyu*	(Rule 3)
tabqa: <*tabqaya*	(Rule 1)	*tabqa:* <*tabqayu*	(Rule 1)
tad'uwa	(Rule 2)	*tad'u:* <*tad'uwu*	(Rule 4)

These phonological changes have relatively little effect on morphology: the p-stem theme vowel /a/ verbs lose distinctions (exemplified by *tabqa:*) in some persons of the a- and u-set only; otherwise, distinctions are preserved. But in the base set, in which in the "strong" verb paradigm there is no final vowel in persons of the verb that do not have a gender or number suffix, syllable final *-iy*, *-ay*, and *-uw* would arise in many persons of verbs with $C_3 = \$$. In fact, in base set forms of this type, $\$$ in C_3 position are deleted, leaving a final short vowel: *lam yarḍa* 'he wasn't content', *lam narmi* 'we didn't throw'. Although it is true that this final $\$$ deletion could be seen as a part of the same phonological processes that affect the a- and u-sets—namely, the deletion of "weak" $\$$—this seems inadequate as an explanation of why it occurs. For one thing, there is no general phonological prohibition in MSA on the occurrence of word final *v$\$$, as a-set forms such as *tarḍay* (itself a reduction of *tarḍayi:*) 'you (f.) may be content' show. The explanation for the deletion of final $\$$ in the base set seems to be as much the need to reduce homophony and ambiguity in the morphology between the phonologically reduced u-set forms such as *tarmi:* and other persons of the verb in the base set, which would collapse if the base set were not so reduced. Were it not for the deletion of final *-y* in a form such as *tarmiy* in the base set, the base set 2fsg., 2msg., 3fsg., the a-set 2fsg., and the u-set 2msg. and 3fsg. would all be phonetically the same—*tarmi:/*tarmiy*. The $\$$ deletion rule in the base set would therefore seem to be more morphosyntactically than phonologically motivated.

In the a- and u-sets of the p-stem verbs that have $\$$ in C_2 position, a set of semivowel assimilation rules apply in open syllables that work in the same way as the vowel assimilation rules described above:

$$\$a \rightarrow /a{:}/$$
$$\$i \rightarrow /i{:}/$$
$$yu \rightarrow /i{:}/$$
$$wu \rightarrow /u{:}/$$

Thus, p-stem u-set:

-aHYaB-	'dread'	Pattern I active: *yaha:bu* <*yahyabu* 'he dreads'
-aXWaF-	'fear'	Pattern I active: *yaxa:fu* <*yaxwafu* 'he fears'
-aSYiR-	'go'	Pattern I active: *yasi:ru* <*yasyiru* 'he goes'
-aQWuL-	'say'	Pattern I active: *yaqu:lu* <*yaqwulu* 'he says'
-aṬWuL-	'be long'	Pattern I active: *yaṭu:lu* <*yaṭwulu* 'it is long'

In the base set generally, and in the a-/u-stem forms that have a consonant-initial suffix, the long vowel is shortened because $Cv{:}C$ is generally an impermissible syllable in MSA. So we get in the base set *yahab*, *yaxaf*, *yasir*, etc., and in all sets the 2/3fpl. forms *tahabna/yahabna*, *taxafna/yaxafna*, *tasirna/yasirna*, etc.

Noteworthy is the fact that the phonological rule that reduces *yaxwafu* to *yaxa:fu* is lexically and morphologically sensitive: a few "older" roots have "strong" /w/ (*ya'wazu* 'he is unable') and certain nominal patterns that have analogous *-$a-* strings are not reduced, for example, *miqwad* 'steering wheel', not *miqa:d*, and *miqwal* 'gramophone' not *miqa:l*.

Noteworthy also is the fact that it is the quality of the underlying vowel in the p-stem that is adopted as that of the short "theme" vowel of the s-stem, contrary to what would be expected if the rule that applied were not *morphologically* determined. For example, from the root *QWL* 'say' we get the s-stem *Qa:L-* from underlying *QaWaL-* by Rule (1) above. But if a consonant-initial enclitic is added to this stem, the long vowel has to be shortened to conform to syllable structure rules, which, as noted immediately above, do not allow $Cv{:}C$ syllables. However, what we get is not *qaltu* but *qultu* 'I said', the /u/ reflecting the /u/ of the underlying p-stem form *yaqu:lu* <*yaqwulu* 'he says', and not the shortening of the stem /a:/ to /a/, which is what might have been expected if all that was happening was vowel shortening. Similarly, from the root *BY* 'sell' we get from the underlying *baya'tu* not *ba'tu* but *bi'tu* 'I sold' because the theme-vowel of the p-stem is /i/, e.g., *yabi:'u* <*yabyi'u* 'he sells'. That the short s-stem vowel is not merely the vocalic correlate of the semi vowel of the root is shown by the fact that from *XWF* 'be afraid' we get *xiftu* 'I was afraid' <*xawiftu* and not *xuftu*.

Whatever its historical origin, the effect and likely explanation for the "pattern-matching" of underlying p-stem/short s-stem theme vowels in verbs that have $ as C_2, as it was in the case of the apocopation of final -*ay*, -*iy*, and -*uw* in the base set forms for verbs with $ = C_3, is the preservation of surface morphological distinctions, in this case between different types of verbs, all of which happen to have a semivowel as C_2.

3.5.2 C2 = C3

Where C_2 and C_3 are the same and occur separated from each other only by a short vowel, the resulting string may undergo adjustment. In the verb, C_2vC_3 is stable if it forms a closed syllable: so from *MLL* 'to be bored' we get the typical forms:

s-stem: *maliltu* 'I got bored'
p-stem base set: *lam yamlal* 'he didn't get bored'

In open syllables, as a result of the suffixing of a vowel-initial suffix, the v separating C_2 and C_3 is deleted, for example, in the s-stem:

ḥajja < *ḥajaj-a* 'he went on pilgrimage'
malla < *malil-a* 'he became bored'
ṣadda < *ṣadad-a* 'he turned (someone) away'

But the same underlying string is treated differently in some types of noun. Noun plurals, verbal nouns, and other nouns of the general form $C_1vC_2vC_3$ remain in this form where two short syllables C_2vC_3v (shown in bold) occur as a result of the suffixing of case endings. For example:

ḥijaj-an	'pilgrimages' (acc.)
malal-an	'boredom' (acc.)
ṣadad-a	'in respect of' (acc.)

It seems from this that the short vowel deletion rule either operates or does not just in order to maintain an important morphological distinction in forms that are phonologically superficially similar.

But in the p-stem, u-set (and a-set) we get, via a rule of short vowel deletion and the epenthesis of a short vowel identical to the deleted one to break up the resulting three-consonant cluster:

yaḥujju < *yaḥjuju* 'he goes on pilgrimage'
yamallu < *yamlalu* 'he becomes bored'
yaṣuddu < *yaṣdudu* 'he turns away'

In nouns that have C_2vC_3v, as in these verbs, the same rules apply: for example, *maḍarra* 'harm', not *maḍrara*; *ʔaḥaqqu* 'more entitled', not *ʔaḥqaqu*; *ʔadilla* 'proofs', not *ʔadlila*; *ʔaṭibba:ʔ* 'doctors', not *ʔaṭbiba:ʔ*; *ʿa:mma* 'general public', not *ʿa:mima*. The explanation for these "adjusted" forms may lie in the greater degree of surface morphological difference that exists between p-stem verbs and nouns (that is, the adjustment rule can be allowed to apply without morphological "risk"), compared with that between s-stem verbs and (some categories of) nouns that are morphologically more similar to each other. Indeed, even in the base set of the p-stem verb and the command form, *yamlal*-type forms are being generally replaced in MSA by "adjusted" a-set *yamalla*-type forms.

3.6 Dialectal verb morphology

3.6.1 Introduction

The point of this section is not merely to list the ways in which the verb morphology of today's spoken Arabic differs from that of CLA/MSA.[15] Our concern is rather to highlight the major typological features that set the dialects as a whole, or particular groups of dialects, apart from MSA and to explain broadly, although without delving into too much historical detail (which in any case is frequently lacking), how and why the present situation came about. In general, as far as is consistent with clear presentation, detailed discussion and exemplification of syntactic function are left to chapters 6 and 7.

It will be obvious from the excursus on phonological variation and hybridization in the last chapter (2.3.1, 2.3.2.) that it would be wrong to describe the language used in many or even most of ordinary people's everyday social interactions as "pure dialect." Everyone's speech nowadays is more or less affected by other varieties of Arabic—principally MSA, but often also the dialects of other regions of the same country or other countries entirely. However, for the sake of presentational clarity, forms of language are presented here that are representative of the unadulterated lowest common denominator vernacular ("Cairene," "Bahraini," etc.) even though, at the level of the individual idiolect, such a form of language is a linguist's fiction.

The morphological structure of the Arabic dialects as a whole is simpler than that of CLA/MSA in the sense that there are fewer separate categories. In the

verb, some morphosemantic patterns have been lost; the irregularities that arise from the appearance of semivowels and identical consonants as root consonants have been ironed out through analogical reduction, and distinctions of mood and voice are no longer carried by internal vowel patterns but have become incorporated into the consonantal skeleton of the stem. In inflectional morphology, a number of categories have disappeared completely. These are generalizations that apply in a "more or less" kind of way: more in the case of the urban vernaculars, less (and differently) in the case of the Bedouin ones, and differently again in the Maghreb as compared with the eastern Arab world.

The losses and changes can be attributed partly to morphological redundancy in the linguistic ancestors of the modern dialects, but also in part reflect the long-term drift from a "synthetic" morphological structure to a more "analytic" one that occurred in the conditions of mass and haphazard learning of Arabic as a foreign language by polyglot urban populations in the centuries after the Islamic conquests (see chapter 1). We know from the early Arab philologists that, before the reification of a mixed tribal poetic koinè into the "Classical Arabic" that came eventually to serve as a written standard, there were tribal and geographical dialect differences. These ancient tribal dialects, although different from each other in detail, were certainly similar to each other typologically. The modern dialects represent the end point of developments from these slightly different pre-Classical starting points, although we still have only an extremely sketchy idea of what these starting points were and how and when the stages in the linguistic evolution of the dialects occurred that brought them to their present stage of development. The parallel but separate development of MSA from CLA as a written literary standard over the same period of time, by contrast, is relatively well documented.

In considering the structure of modern dialectal Arabic, one should bear two points in mind. The first is that the present-day prescriptive distinction between "Classical" (in Arabic *fuṣḥa:*) and "colloquial/dialectal" is a somewhat arbitrary one, historically speaking, and reflects age-old politicoreligious and aesthetic considerations. Many contemporary features condemned as "dialectal" are in fact just as "Classical," if we define that word as meaning "having an ancient textual historical precedent," as the forms that actually entered the Classical canon.[16] The second point follows logically from this: one should avoid the automatic "derivation" of contemporary colloquial features from Classical ones, as if there were always a direct line of development between the two. Although it is true that many modern colloquial features *do* represent, quite literally, a continuation of features that happen also to have been included in the philologists' approved canon, others are survivals of Old Arabic forms of expression that

were coeval with "Classical" ones but not in any sense "derived" from them, whereas yet others are the accretions of gradual internal developments that have occurred over many centuries. It would be quite wrong, although it is still often done, to represent the dialects as mere corruptions of some imaginary linguistic gold standard that existed in the eighth century.

In the modern period, as we briefly noted at the end of chapter 1, "Classical Arabic," in the guise of its modern descendant MSA, and the colloquials have once more begun to converge in form and usage by dint of MSA assuming some of the characteristics of a living, rather than an institutionally imposed, standard language. This too poses descriptive problems: even if we exclude "hybridized" forms of Arabic from the "unadulterated dialect" on the grounds that they are obviously the result of dialect–MSA contact, we are still faced with the problem of deciding whether some usages are "really dialectal" (in the sense of having been part of ordinary nonliterate speakers' repertoire of forms before the advent of modern schooling and the media) or whether they have been recently "borrowed" from MSA. Borrowing from another dialect is usually much easier to spot.

3.6.2 Losses and gains in inflectional categories and morphosyntactic patterns

3.6.2.1 Urban dialects

If we take CLA morphology as the point of comparison, the major changes in the morphosyntax and inflection of the verb in the eastern urban dialects (e.g., Cairo, Damascus, and Jerusalem) can be summarized as follows:

(a) Number and gender
   (i) Gender distinctions in the second and third person have disappeared in the plural—the masculine forms have been generalized.
   (ii) Dual number has been completely lost. The masculine (= now common) plural or the feminine singular is used instead.
   (iii) The suffixes in the s-stem have been simplified: 1sg. and 2msg. have the same suffix -*t*, 2nd (common) pl. has –*tu*, and the 3sg. has no suffix.
(b) Mood and voice
   (i) The u-set and a-set mood suffixes have been lost. The dialects have a paradigm similar to the CLA base set but with aspectual/modal prefix particles. In the paradigms below, one of these prefixes, which generally has an indicative/declarative function (= CLA u-set) is placed in brackets before the base set as an example.

Table 3.2   Urban Levantine Dialects: Verbal Affixes for Person, Number, Gender

| | P-STEM | | S-STEM | |
| | Number | | Number | |
Person/Gender	Sg.	Pl.	Sg.	Pl.
1	*(b)a-* . . .	*(b)ni-* . . .	. . . *-t*	. . . *-na*
2m.	*(b)ti-* . . .	*(b)ti-* . . . *-u*	. . . *-t*	. . . *-tu*
2f.	*(b)ti-* . . . *-i*	*(b)ti-* . . . *-u*	. . . *-ti*	. . . *-tu*
3m.	*(b)yi-* . . .	*(b)yi-* . . . *-u*	. . . *-ϕ*	. . . *-u*
3f.	*(b)ti-* . . .	*(b)yi-* . . . *-u*	. . . *-it*	. . . *-u*

(ii) The discontinuous vowel patterns that signify passive voice in CLA/MSA have been lost. Passive voice is indicated by a prefix that is attached to the stem of the verb, sometimes *ʔin-* and sometimes *ʔit-*, depending on the verb and the dialect.

(iii) The form of the prefix has been generalized in all persons except 1sg. to *ti- /yi- /ni-* . . . , where CLA/MSA have either /*a*/ or /*u*/ as the prefix vowel depending on the stem pattern and active/passive distinctions.

The simplified affixational paradigm for the urban dialects of the Levant, ignoring local phonological differences, forms table 3.2. (compare with table 3.1. for CLA/MSA).

3.6.2.1.1 Loss of marked categories

"Marked" categories, in the sense of low frequency and usefulness, have in general tended to disappear in the urban dialects: hence, here the loss of the feminine plural form, normally used in CLA/MSA only to refer to female human plural groups of more than two members, and feminine gender nonhuman plurals of between two and ten members.[17] The dual, another highly specified category, continues in the dialects as a fossil in the noun only. In the pronoun, adjective, and verb it has been lost and is replaced with the plural form. Dual nouns are treated, from the point of view of concord, as if they were plurals.[18] Compare the typical forms:

CLA/MSA:
ʔalbana:tu daxal-na
'The girls entered-3fpl'
ʔarrija:lu daxal-u:
'The men entered-3mpl'

ʔalbint-a:ni daxal-ata:
'The girl-dual entered-3fdual'
ʔarrajula:ni daxal-a:
'The man-dual entered-3mdual'

modern Cairene:
ilbana:t
'The girls
irrigga:la
'The men
ilbinte:n   *daxal-u*
'The girl-dual   entered-3compl'
irragle:n
'The man-dual

3.6.2.1.2 Paradigm symmetry and simplicity

In the remaining forms, the dialect paradigm has shed elements of the affixes of both the p-stem and s-stem. In the p-stem, the redundant *-na* element of the *-i:na* and *-u:na* suffixes has been lost, and the resulting redundantly long vowels shortened,[19] bringing all suffixed forms into line. In the s-stem, final short vowels (except in 2fsg. *-ti*) have disappeared as a result of the generalization of prepause forms (= loss of final short vowels and shortening of long ones). No ambiguity necessarily results because the interactional speech context (obviously a factor affecting the linguistic economy of exclusively spoken varieties of language) generally makes it clear whether an action is being predicated of "I" or "you." Moreover, the disjunctive pronouns *ana* 'I', *inta* 'you(m.)', *inti* 'you(f.)' are also more often used as overt verbal subjects in the dialects than are their counterparts in CLA/MSA, which (because of the higher degree of morphological differentiation in the CLA/MSA verb) tended to be used with full verbs only in contrastive or emphatic contexts. The overall effect is to create greater simplicity and paradigm symmetry in the dialects without creating new ambiguities.

The same type of system as described here for the eastern Arab urban dialects is also found in the Maghreb, although here the process of analogical leveling has gone even further, reaching its most advanced stage in the big cities.[20] Everywhere in the region the first person number distinction has been made isomorphic with that in the second and third person, so that *n-* . . . has become 1sg. and *n* . . . *-u* 1pl. In the Moroccan urban dialects, the gender distinction has been lost in the 2sg. as well as in the 2pl. and 3pl., although in different ways: in the s-stem via the generalization of the *-ti* f. suffix, and in the p-stem via the gen-

Table 3.3   Tunis Arabic: Verbal Affixes for Person, Number, Gender

| | P-STEM | | S-STEM | |
| | Number | | Number | |
Person/Gender	Sg.	Pl.	Sg.	Pl.
1	n- . . .	n- . . . -u	. . . -t	. . . -na
2	t- . . .	t- . . . -u	. . . -t	. . . -tu
3m.	y- . . .	y- . . . -u	. . . -φ	. . . -u
3f.	t- . . .	y- . . . -u	. . . -it	. . . -u

eralization of the suffixless masculine form. In Tunis, the second person gender distinction has been lost also, although here through the generalization of the masculine forms in both paradigms (table 3.3.).

3.6.2.1.3 Loss of final vowels = loss of mood distinctions = compensatory development of modal/aspectual particles

The loss of the vowel suffixes indicating mood in the p-stem was a further consequence of the historical loss of all word-final short vowels and the generalization of "pause forms" to all positions. This has meant that a set of forms similar to the base set of CLA/MSA has become the sole p-stem paradigm in the dialects. Mood distinctions carried by vowels (or absence of them) have been replaced by distinctions carried by particles (or absence of them) derived from decayed functional or lexical elements. This type of "analytical" development has been a very widespread and productive one in the morphology in all urban dialects. In this case, a *b(i)*- element is prefixed to the base set p-stem to cover roughly the same range of meanings covered by the CLA/MSA u-set: statements of fact that may be timeless in their validity, ongoing at the time of speech, and habitual, durative, repetitive actions or states. Instead of the particle *sa*- + u-set, the dialects prefix *ḥa*- to the base set to refer to future actions or intentions. There is no prefix in cases where CLA/MSA has the a-set or base set, for example, verbs in subordinate noun clauses that describe actions and states in the realm of the possible, intended, wanted, or striven after, as well as jussive (command) forms, affirmative and negative (see 6.2). Table 3.4 (taking Cairene as a typical representative of an eastern urban-type dialect) provides a summary.

3.6.2.1.4 Internal vocalic changes = loss of internal active/passive distinction = compensatory extension of stem prefix system

Changes in the quality and distribution of short vowels conditioned by contiguous consonants (3.6.2) also caused the blurring of the distinction between the active and passive internal vowel patterns and the eventual loss of the latter. Here the dialectal solution was to extend the use of the existing alternative

Table 3.4   P-Stem Mood and Aspect Markers: CLA/MSA and Cairene

Mood/Aspect	CLA/MSA	CAIRENE
Present (factual)	u-set	*b(i)* + base set
Present (nonfactual)	a-set	base set
Future/Intentional	(*sa*) + u-set	*ḥa* + base set
Jussive	base set	base set

means of passivization—the prefixing of a consonant to the Pattern I stem. Depending on the verb and the dialect, an *ʔin*- prefix (i.e., Pattern VII) or an *ʔit*-prefix (cf. Pattern V) is used in today's urban dialects. Contrast the typical forms:

	CLA/MSA	DIALECT
s-stem:		
'write'	*KaTaB*-	*KaTaB*-
'be written'	*KuTiB*-	*ʔitKaTaB*- / *ʔinKaTaB*-
p-stem:		
'write'	-*aKTuB*-	-*KTiB*-
'be written'	-*uKTaB*-	-*tKiTiB*- / -*nKiTiB*-

3.6.2.1.5 Summary

Changes in the inflectional system were thus occasioned by two factors: the loss of low-frequency, marked categories and the reduction of morphological redundancy in the rest of the paradigm, resulting in greater simplicity and symmetry. In morphosyntax, the functional load of vowel patterns has been reduced and that of particle and affixational elements increased to compensate for the loss.

3.6.2.2 Bedouin dialects

We see a somewhat different pattern of development in the dialects of recently sedentarized Bedouin (there are no true nomads left) living in areas remote from the large population centers of the Mediterranean basin and its hinterland. Here, many features of the ancient CLA verb system are preserved. The populations that speak these dialects did not mix and intermarry with non-Arabs and were not subjected to outside influences, which explains the conservatism of the dialects they speak. For example, in the dialect of the Najdi town of Ha'il, 400 miles north of Riyadh in central Arabia, we find the following:[21]

- the p-stem -*iːn* and plural -*uːn* suffixes survive,
- gender distinctions in the second and third person plural are maintained,

Table 3.5 Ha'il Arabic: P-Stem Affixes for Person, Number, Gender

Person/Gender	Sg.	Pl.
	Number	
I	a- . . .	n- . . .
2m.	t- . . .	t- . . . -u:n
2f.	t- . . . -i:n	t- . . . -in
3m.	y- . . .	y- . . . -u:n
3f.	t- . . .	y- . . . -in

Table 3.6 Bahraini Arabic: P-Stem Affixes for Person, Number, Gender

Person/Gender	Sg.	Pl.
	Number	
I	a- . . .	ni- . . .
2m.	ti- . . .	ti- . . . -u:n
2f.	ti- . . . -i:n	ti- . . . -u:n
3m.	yi- . . .	yi- . . . -u:n
3f.	ti- . . .	yi- . . . -u:n

- the internal passive survives alongside the reductional change to an *ʔin-* prefix, and
- a system of mood particles has not developed.

Compare now the p-stem affixational paradigms for CLA/MSA (table 3.1) and for the urban dialects of the Levant (table 3.2) and Tunis (table 3.3) with that for Ha'il in table 3.5.

In some respects—complete loss of dual number, loss of the 1sg./2msg. distinction in the s-stem, loss of final short vowels, and the shortening of final long ones—developments in Ha'il have paralleled those of the urban dialects. But it is well to emphasize the word "paralleled": as we have noted, the urban and Bedouin dialects have had different social histories, developing independently of each other but from what must have been similar starting points many centuries ago. Their present-day similarities vis-à-vis CLA seem to be the result of a natural drift from a high to a reduced degree of category differentiation in inflectional morphology, triggered initially by phonological change, which has taken them separately in the same general direction. Where they differ is in the speed, scope, and detail of the change.

The Bedouin-descended dialects of the seaports of southern Iraq and the Gulf littoral are an interesting typological compromise between the reductional tendencies of the Levant cities and the conservatism of inner Arabia, which again reflect the sociolinguistic history of their speakers. A "conservative" Bedouin dialect, probably similar to the type spoken in Ha'il, was exported from central Arabia about 200 years ago and has been subjected in the interim to the long-term leveling influences of interdialectal mixing and learning by large numbers of immigrant non-Arabs (Persians and Indians in the main). In the town dialects of Kuwait and Bahrain, for example, we find:

- the *-i:n* and *-u:n* suffixes survive (as in the source central Arabian dialects),

- the 2/3pl. gender distinctions have been lost (like the Levant cities, unlike central Arabia),
- the internal passive has been lost except in a few fossilized expressions, and affixation is the normal means of passivization (like the Levant cities, unlike central Arabia), and
- a partial system of mood particles has developed (for the expression of future intention only), but the system is different and less fully morphologized than in the Levant, with the lexical origin of the mood elements more transparent.

The p-stem affixational paradigm for Bahraini (table 3.6.) is a typological "halfway house" between the Levantine paradigm and the CLA-like Ha'il paradigm.[22]

3.6.3 Morphosemantic developments in the dialect verb

We are concerned here with developments in the root-pattern system. Broadly speaking, similar types of reductional change occurred in the urban dialects as occurred in affixational and morphosyntactic elements, with the Bedouin dialects somewhat less (and differently) prone to innovation. The major changes can be summarized as follows:

(a) restructuring of the formal correspondence of interdigitated vowel morphemes to meaning types in Pattern I verbs,

(b) reduction in the number of Pattern I stem types, and

(c) loss of some augmented patterns, extension of others.

We will deal with these in turn.

3.6.3.1 Restructuring of Pattern I

Recall that CLA has three s-stem types (theme vowels /a/, /i/, /u/) that correspond as follows to p-stem theme vowels:

s-stem   *a*   *i*   *u*

p-stem   *a i u*   *a*   *u*
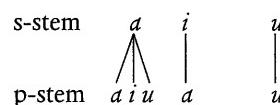
Verbs with s-stem /*a*/ are generally verbs of action, mainly but not entirely transitive; /*i*/-stems mainly denote actions that are grammatically transitive but whose consequences are experienced by the agent, which we term "medial" or "agent moyen," for example, 'understand', 'know/come to know', 'wear/put on (clothes)', 'get in (a car, etc.)', 'hear', and 'drink' and verbs of being and becoming that express contingent states, for example, 'be/become afraid', 'be/get angry', 'be/become happy'; /*u*/-stems denote the possession or the coming into possession of more permanent qualities ('be/become ugly', 'be/become old'). The correspondences between s-stem and p-stem vowels are partially predictable: (i) s-stem /*a*/ most often has p-stem /*u*/, whereas p-stem /*a*/ is usually found in verbs where C_2 or C_3 is a postvelar (/*x*/, /*ġ*/, /*ḥ*/, /*'*/, /*ʔ*/, /*h*/); (ii) s-stem /*i*/ in the vast majority of cases has p-stem /*a*/, although there are some verbs with p-stem /*i*/; (iii) s-stem /*u*/ almost invariably has p-stem /*u*/.

In dialectal phonology, there has been a general tendency for frontness/backness in the high short vowels to become nondistinctive and for vowel quality in general to become more dependent on consonantal environment, both developments entailing the blurring of morphological distinctions that depended on them. Coincidentally, as elsewhere in the morphology, realignments have occurred that are purely morphosemantic in character, and categories that are marginal to the functioning of the system as a whole have tended to lose their distinctive marking, just as in affixational morphology. Both of these tendencies have dislocated and simplified the CLA system of verb classification, as well as the s-stem/p-stem vowel correspondences, without, however, totally destroying them. Cairene Arabic provides a typical "urban" case in point.

Like CLA, Cairene has a three-way division of s-stems, but, unlike it, v_1 always harmonizes with the theme-vowel, v_2. For example:

Cairene	*KaTaB-*	*FiRiḤ-*	*RuXuṢ-*
cf.			
CLA	*KaTaB-*	*FaRiḤ-*	*RaXuṢ-*
	'write'	'be happy'	'become cheap'

A majority of the verbs that have s-stem /*a*/ theme vowel in CLA have it also in Cairene, and the correspondence is even closer between CLA /*i*/ theme vowel and Cairene *i–i* verbs. However, two major innovations have occurred. The first is that a considerable number of verbs that had CLA /*a*/ theme vowel but that

were intransitive or stative in meaning have been transferred to the Cairene *i–i* class, for example, *RiGiʕ-* 'return', *WiLiD-* 'give birth', *XiLiṢ-* 'come to an end', *ĠiTiṢ-* 'dive, become submerged', *FiSiD-* 'become corrupt', *ṬiLiʕ-* 'go out/up', *FiLiT-* 'work loose, escape', *SiKiT-* 'become silent', *HiLiK-* 'die, perish', *HiRiB-* 'flee', *BiRiD-* 'cool off', *WiʔiF-* 'stand, stop', etc. Conversely, a number of verbs that were transitive but had a CLA /*i*/ theme vowel have been shifted to the Cairene *a–a* class, for example, *ʕaMaL-* 'make, do', *TaBaʕ-* 'follow', *XaṬaF-* 'grab', *ḤaFaẒ-* 'keep', *ḤaSaB-* 'deem, regard as'. Cairene thus seems to have adjusted (or to be adjusting?) the basis of the correspondence between vowel pattern and verb type so that transitivity is more consistently marked morphologically.[23] There are some exceptions in verbs that could have been expected to move into the Cairene *i–i* class but that have not, for example, *ŠaRaD-* 'flee' (compare *HiRiB-*, same meaning, that did), *SaʕaL-* 'cough', *RaʔaD-* 'lie (flat)', but the trend is clear nonetheless.

The second difference is that the verbs in class *u–u*, ultimately descended from the CLA /*u*/-stems (the most specialized from the point of view of semantic class, originally denoting permanent as opposed to contingent qualities and states of being) have without exception free variants in *i–i*, for example, *RuXuṢ-*/*RiXiṢ-* 'become cheap' *ṢuĠuR-*/*ṢiĠiR-* 'become small', *KuBuR-*/*KiBiR-* 'become big/old', *ʔuṢuR-*/*ʔiṢiR-* 'become short', *ʔuRuB-ʔiRiB-* 'be/come near', *KuTuR-*/*KiTiR-* 'become numerous'.[24] Some verbs that were "transferred" from the CLA /*a*/ to the Cairene *i–i* type have also developed free variants in *u–u*, for example, *SiʔiṬ-*/*SuʔuṬ-* 'fall', *ĠiTiṢ-*/*ĠuṬuṢ-* 'dive, be submerged', *ṢiBiR-*/*ṢuBuR-* 'be patient', *XiLiṢ-*/*XuLuṢ-* 'come to an end'. The effect of these developments is that Cairene no longer has a fully differentiated three-way division in the s-stem Pattern I verb but rather a binary low-vowel/high-vowel contrast (with vowel harmony) in which the dialectal analogs of CLA /*u*/-theme verbs have merged morphologically with the analogs of the /*i*/-theme. Although, as noted, there are some exceptions, the system of theme-vowel marking basically mirrors transitivity and (hence) differences in complementation. Outside this general scheme is a group of verbs, termed here "medial," that is also a subcategory of the CLA /*i*/-stems (see above). These verbs are in some cases transitive and in other cases not, but their common distinguishing feature is that they denote bodily or cognitive actions/functions, the major effect of which is felt by the agent.

These divisions are illustrated below with sample verbs in each category:

TRANSITIVE	INTRANSITIVE	MEDIAL
ʕaMaL- 'make'	*RiGiʕ-* 'return'	*ŠiRiB-* 'drink'
KaSaR- 'break'	*Wiʔiʕ-* 'fall'	*ʕiRiF-* 'know'

DaRaS- 'study'	*ṬiLi?-* 'go up'	*'iRi?-* 'sweat'
XaDaM- 'serve'	*WiLiD-* 'give birth'	*Li'iB-* 'play'
FaTaḤ- 'open'	*ĠiRi?-* 'drown'	*SiMi'-* 'hear'
ḤaMaL- 'carry'	*KiTiR-* 'become	*FiHiM-* 'understand'
ḌaRaB- 'hit'	(*KuTuR-*) many'	*RiKiB-* 'get on, mount'

Other developments in the dialect complement and support the general the-sis proposed here of a more clearly marked transitive–intransitive opposition. In some roots a new *a–a/i–i* contrast has evolved, for example, *RaKaZ-* 'set up (something) securely', *RiKiZ-* 'come to rest safely', *SaKaN-* 'live in (a house)', *SiKiN-* 'become occupied', *FaṬaS-* 'flatten', *FiṬiS-* 'die', *ĠaLaB-* 'overcome', *ĠiLiB-* 'suffer hardship', *HaLaK-* 'exhaust', *HiLiK-* 'be exhausted'. In some of these pairs (e.g., *RaKaZ-/RiKiZ-*), the *i–i* stem does not hark back to a corresponding CLA /*i*/-stem verb but seems rather to be a fossilization of a CLA internal (*CₚuC₂iC₃-*) passive, whereas in other cases (*HaLaK-/HiLiK*) a CLA Pattern I/Pattern IV sta-tive–causative contrast has been transmuted to an *i–i/a–a* dialect contrast.

Vowel deletion rules introduce a further dimension of morphological con-trast between the two stem types. In those forms of the s-stem whose suffix is vowel initial, a *CvCvCv(C)* string would result. If v_2 in such a string is a high vowel, it is deleted, so *katabit* 'she wrote', *'amalit* 'she made', *ḍarabit* 'she hit', but *rig'it* 'she returned', *širbit* 'she drank', *fihmit* 'she understood'.

In the p-stem, there are three patterns, *i–a, i–i*, and *i–u* (var. *u–u*), for example, -*XṬaF-* 'grab', -*KTiB-* 'write', -*SKiT-* or -*SKuT-* 'remain silent'. The stem-vowel pattern here is partly dependent on verb class and partly on consonantal envi-ronment, as to some degree it is in CLA, although the details of the corre-spondence are different. An examination of some 116 verbs chosen randomly from Hinds and Badawi's dictionary revealed the following:

(a) If a verb has *i–i* or *u–u* in the s-stem, it almost invariably has theme vowel /*a*/ in the p-stem (52 out of 56 verbs examined), whatever the consonants before and after it and whatever p-stem vowel it may have had in CLA, for example, -*BXaL-* (CLA /*a*/) 'be miserly', -*RXaS-* (CLA /*u*/) 'become cheap', -*XLaS-* (CLA /*u*/) 'come to an end', -*BRaD-* (CLA /*u*/) 'cool off' -*RGa'-* (CLA /*i*/) 'return', -*'RaF-* (CLA /*i*/) 'know'.

(b) In the *a–a* s-stems, the p-stem vowel can be /*a*/, /*i*/, or /*u*/ but is largely predictable phonologically: if C₂ or C₃ is postvelar (i.e., /*x*/, /*ġ*/, /*ḥ*/, /*'*/, /*?*/, /*h*/) or emphatized, the p-stem vowel is /*a*/ (29 out of 30 verbs examined), for example, -*GRaḤ-* 'wound', -*B'aT-* 'send', -*ḤDaR-* 'attend', -*ḤFaẒ-* 'keep'; if C₂ or C₃ is neither postvelar nor emphatic, the vowel is /*i*/ or /*u*/, or in a few cases in free variation (27 verbs out of 30 verbs examined), for example,

-*KTiB-* 'write', -*XDiM-* 'serve', -*DRiS-* 'study', -*TRuK-* 'leave', -*'MiL-* 'do, make', -*KNiS-/-KNuS-* 'sweep'. To some degree (in the effect of the postvelars), the consonantal conditioning of the p-stem vowel in Cairene echoes that of CLA.

The result of the reorganization of the Pattern I verb stem voweling system in Cairene thus seems to have been a more consistent and (from the user's point of view) predictable marking in the morphology of transitivity types. Phono-logical conditioning also played a part, although a subordinate one within this reorganized system based on transitivity marking. Compare the CLA system at the beginning of this section with the modern Cairene system of s-stem/p-stem vowel correspondences:

s-stem:		*a*		*i/u*
		(mainly trans.)		(mainly intrans.)
		/	\		
p-stem	*a*		*i/u*	*a*
	(+ emphatic/		(– emphatic/	(all
	postvelar)		postvelar)	environments)
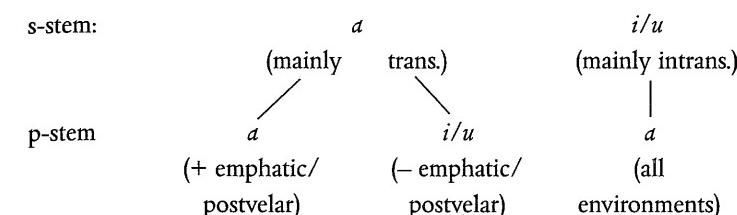
Let us now briefly compare the Cairene development, typologically similar to what happened in other major urban areas of the eastern Mediterranean, with what occurred in the Bedouin dialects of northern and eastern Arabia. Here the CLA correspondence of s-stem vowel to verb type has also been realigned, although here principally by phonological change. In these dialects, general restrictions have developed on the occurrence of /*a*/ in certain consonantal en-vironments, and there has been a partial merger of /*i*/ and /*u*/ such that /*i*/ oc-curs to the exclusion of /*u*/ in all except environments where a labial consonant is present. The basic rule is that any short vowel in a nonfinal open syllable is /*a*/ if any contiguous consonant is a postvelar or if the following consonant is a liquid (/*l*/, /*r*/) or /*n*/; otherwise, it is /*i*/ or /*u*/,[25] the latter generally if a labial (especially /*b*/ or /*w*/) is in contiguity. So, for example, we get the fol-lowing typical Pattern I s-stems in the Bahraini 'Arab (BA) dialect:[26]

$v_1 = /a/$

BAHRAINI SUNNI (BA)		compare: CAIRENE	CLA
XaDaM-	'serve'	*XaDaM-*	*XaDaM-*
'aRaF-	'know'	*'iRiF-*	*'aRaF-*
FaHaM-	'understand'	*FiHiM-*	*FaHiM-*
La'aB-	'play'	*Li'iB-*	*La'iB-*
ŠaRaB-	'drink'	*ŠiRiB-*	*ŠaRiB-*
ṬaLa'-	'go up'	*ṬiLi'-*	*ṬaLa'-*

v_1 = /i/ or /u/:

KiTaB-	'write'	KaTaB-	KaTaB-
KiSaR-	'break'	KaSaR-	KaSaR-
NiGaL-	'transport'	Na?aL-	NaQaL-
GiDaR-	'be able'	?iDiR-	QaDiR-
SiMa'-	'hear'	SiMi'-	SaMi'-
KiBaR-/KuBaR-	'grow old'	KiBiR-/KuBuR-	KaBuR-
WuGaF-	'stand, stop'	Wi?iF-	WaQaF-
WuṢaL-	'arrive'	WiṢiL-	WaṢaL-

It is clear from these examples that the effect of the phonological change in Sunni BA has been to destroy the relationship between s-stem theme vowel and verb type, and to reduce all Pattern I strong verbs to a single type in which voweling differences in v_1 position are phonologically determined. The categorical occurrence of /a/ as v_2, however, seems to be a matter of morphology (in the Shi'i dialects, /i/ can and does occur as v_2), a CvCaC- shape having been generalized for all Pattern I s-stems but not for the nouns that had a CvCvC shape in CLA. Moreover, like many Bedouin-descended eastern Arabian dialects, BA also has a rule that resyllabifies all CvCvCv(C) strings to CCvCv(C) (see 2.2.2.1 above). Such strings occur in the Pattern I strong s-stem in the 3fsg./3pl. as a result of suffixation, resulting in the deletion of v_1 of the stem. A consequence of this deletion is that v_2 of the stem is now in open syllable and is also subject to the conditioning rule noted above. So, unlike Cairene, which has a high vowel deletion rule whose effect is to consistently distinguish a–a from i–i/u–u stems in suffixed forms (e.g., katabit 'she wrote' versus siktit 'she became silent' /?urbit 'she drew near'), BA treats all such forms in the same way, via resyllabication and vowel conditioning. The morphological effect is again to collapse distinctions between verb types. For example,

	RESYLLABICATION	VOWEL CONDITIONING		
kitab + at →	(ktabat) →	ktibat	'she wrote'
sikat + at →	(skatat) →	skitat	'she became silent'
garab + at →	(grabat) →	grubat	'she drew near'
kisar + at →	(ksarat) →	ksarat	'she broke'
sima' + at →	(sma'at) →	sma'at	'she heard'

The same overarching vowel-conditioning rule involving the postvelars applies in the p-stem. Here, if C_2 or C_3 is a postvelar, the theme vowel is /a/ whatever the original vowel may have been in CLA, and v_1 dissimilates to /i/. In nonpostvelar environments, the theme vowel is generally (as in CLA) /i/ or /u/ except in the "medial" group, where (again as in CLA) it is /a/. The prefix v again

automatically dissimilates. There are two sets of apparent exceptions to this general scheme. The first is that in verbs in which an emphatized consonant (or one secondarily emphatized through emphasis spread) occurs in C_2 or C_3 position, the vowel is backed and lowered to /a/, whatever its quality may have been in CLA. This, however, is only an apparent exception, because emphatization involves retraction of the tongue root and constriction of the pharynx, a feature it shares with the postvelars, which, as we have seen, regularly condition the quality of contiguous short vowels in the same way. The second exception, interestingly, is in verbs that in CLA denoted permanent states or qualities: even if they do not have a postvelar in C_2 or C_3 position, they have merged, exactly as they have done in Cairo, with the "medial" group of verbs denoting bodily actions/functions and contingent qualities and states and have /a/ p-stem vowel. Examples (CLA p-stem vowel in brackets for comparison):

C_2 OR C_3 = POST-VELAR	OTHER VERBS
-iFHaM- (a) 'understand'	-a'RiF- (i) 'know'
-iBLaĠ- (u) 'reach'	-aHLiK- (i) 'die'
-iṬLa'- (u) 'go up'	-aĠSiL- (i) 'wash'
-iBṬaH- (a) 'knock down'	-aRSiM- (u) 'draw'
-iṬBaX- (u/a) 'cook'	-aBRiZ- (u) 'be ready'

C_2 OR C_3 = EMPHATIZED	"MEDIAL VERBS"
-RFaD- (u) 'refuse'	-iŠRaB- (a) 'drink'
-HDaR- (u) 'attend'	-iLBaS- (a) 'dress'
-DRaB- (i) 'hit'	-iGDaR- (a) 'be able'
	-iRKaB- (a) 'mount'

	"PERMANENT STATE"
	-iṢĠaR- (u) 'be small'
	-iKBaR- (u) 'be old'
	-iČθaR- (u) 'be many'

Thus in BA the primary determinant of the p-stem vowel is phonological environment: where postvelar articulation or coarticulation (= the emphatics) is present, the vowel is low. Otherwise, with the exception of the merged "medial/permanent state" group, the p-stem vowel is high.

In conclusion, if we take the CLA system as our starting point, the basic difference between the Cairene and the BA systems for Pattern I verb classification is that in Cairene, morphological marking according to semantic type predominates over the power of phonological environment as a simplifying/organiz-

ing principle, whereas in BA it is the other way around. As noted, there are cat-egories of verb that in each case are exceptional, but in each case the result is nonetheless (from the user's/learner's standpoint) a reduction in complexity and an increase in predictability.

3.6.3.2 Reduction in the number of Pattern I stem types

In CLA, the various vowel reduction and assimilation processes described in 3.5.1–2 resulted in four basic s-stem types in Pattern I verbs (henceforth desig-nated by the Arabist nomenclature), which we illustrate with examples involv-ing the stem + vowel-initial suffix and stem + consonant-initial suffix:

	STEM	STEM + v	STEM + Cv
		3msg	1sg
"strong"	KaTaB-	kataba	katabtu
"hollow" (C_2 = /w/ or /y/)	QaWaL-	qa:la	qultu
"weak" (C_3 = /w/ or /y/)	RaMaY-	rama:	ramaytu
"doubled" ($C_2 = C_3$)	ŠaDaD-	šadda	šadadtu

Disregarding phonological differences of detail, the corresponding forms in all the modern dialects, urban and Bedouin, North African and Middle Eastern, are of the following general shape (Cairene and BA forms are used as an illustration):

CAIRENE		BAHRAINI SUNNI	
katab	katabt	kitab	kitabt
ʔa:l	ʔult	ga:l	gilt
rama	rame:t	rima	rime:t
šadd	šadde:t	šadd	šadde:t

Thus all dialects have ended up simplifying the system in the same way, col-lapsing the distinction between the doubled and weak verbs in suffixed forms. The same merger has occurred in weak and doubled verbs in the augmented pat-terns, for example, Cairene ʔirtaḍe:t (Pattern VIII, root RḌY), štadde:t (Pattern VIII, root ŠDD), Bahraini iltage:t (Pattern VIII, root LGY), staḥabbe:t (Pattern X, root ḤBB). However, this system was not always as dialectally universal as it is today. In medieval Christian Arabic in south Palestine a different simplification of the CLA system occurred[27]—some doubled verbs at least were treated like hollow verbs. So from MSS 'touch', we got masna: 'we touched' (cf. hollow forms such as qulna: 'we spoke' from QWL) rather than the present-day general masse:na. In present-day southern Iraq, Bahrain, and northern Qatar, some speakers have generalized the weak paradigm in suffixed forms to verbs of *all* stem types in all patterns, by simply adding -e:C to the stem, for example, ktibe:t 'I wrote', ga:le:t

'I said', tzawwaje:t 'I married' as well as the expected rime:t, šadde:t.[28] This mi-nority variety of extreme reductionism is now dying because of contact with (more prestigious) dialects that have merged suffixed forms only in the weak and doubled stem types. Dialect contact is also almost certainly the explanation for the loss in modern Palestinian dialects of the doubled verb masna: -type forms cited by Blau for the thirteenth century.

The morphological principle involved in the weak/doubled merger that we now see everywhere in the dialects is analogical leveling, probably originating in the coincidental surface similarities between the Pattern II weak verb and the Pattern I doubled verb in CLA. A typical example of the former is, in stem form, xalla 'he let, allowed' (root XLW). In suffixed forms of the CLA p-stem, these two verb types have, because of the vowel reduction and consonant assimila-tion rules, isomorphic sets of forms. For example:

	PATTERN I DOUBLED	PATTERN II WEAK
p-stem:	tašiddi:na 'you(fsg) pull'	tuxalli:na 'you(fsg) let'
	tašiddu:na 'you(mpl) pull'	tuxallu:na 'you(mpl) let'
	yašiddu:na 'they(mpl) pull'	yuxallu:na 'they(mpl) let'
commands:	šiddu: 'pull(mpl)!'	xallu: 'let(mpl)!'
	šiddi: 'pull(fsg)!'	xalli: 'let(fsg)!'

What seems to have happened subsequently in the dialects is a reanalysis of suf-fixed doubled stem forms as "weak stems" on the basis of the kind of surface similarities exemplified in the CLA p-stem here, which increased as the dialects evolved: loss of the -na suffixes through generalization of the base set and the replacement of the feminine plural forms by masculine ones (discussed above) would have increased the similarity between them even further. The final stage would have involved extension of this structural isomorphism to the suffixed forms of the s-stem doubled verbs.

A similar process of paradigm convergence has resulted in the general loss in the dialects of weak verbs in which C_3 = /w/, which have merged with weak verbs in which C_3 = /y/. Again, because of the effect of the vowel reduction and assimilation rules, many forms in these verb types already resembled each other morphologically in CLA. For example:

ROOT	DʿW 'call'	MŠY 'walk, go'
p-stem:		
2mpl	tadʿu:na <*tadʿuwu:na	tamšu:na <*tamšiyu:na
3mpl	yadʿu:na <*yadʿuwu:na	yamšu:na <*yamšiyu:na
2fsg	tadʿi:na <*tadʿuwi:na	tamši:na <*tamšiyi:na

s-stem:

3msg	*da'a:* <**da'awa*	*maša:* <**mašaya*
3fsg	*da'at* <**da'awat*	*mašat* <**mašayat*
3mpl	*da'aw* <**da'awu:*	*mašaw* <**mašayu:*

After the generalization in the dialects of the base set p-stem and the generalization of the masculine plural to the feminine plural, the result, as in the case of the doubled/weak verb s-stem merger, was that the paradigms came to resemble each other even more closely. The next stage would have been an analogical leveling of the parts of the paradigms that remained different. The reason this affected verbs with $C_3 = /w/$ was because there were far fewer of them than verbs in which $C_3 = /y/$, and they were more vulnerable to loss in any merger of categories.

Category change, although not reduction, also occurred through simple phonological loss. This has happened to verbs that in CLA were strong and had C_I or $C_3 = /?/$. In all dialects, historical /?/ has tended to disappear completely in final position and in other positions to be replaced (depending on its morphological function) by vowel length or /w/ or /y/. This has caused reclassification of some types of verb stem.

3.6.3.2.1 Verbs with $C_3 = /?/$

After the general loss of final short vowels, /?/ in final position in stems in which $C_3 = /?/$ became unstable and was eventually lost, which resulted in such stems becoming indistinguishable from those of weak verbs. Compare:

CLA	CAIRENE	GULF	
yaqra?u	*yi?ra*	*yigra*	'he reads'
qara?a	*?ara*	*gara*	'he read'
qara?tu	*?are:t*	*gare:t*	'I read'

In such forms, all CLA *CC?* roots have effectively been reanalyzed in the dialects as *CCY* with p-stem vowel /a/ and have evolved exactly like verbs with historical /y/ as C_3 such as *RDY* 'be content'. Such new roots are of course available for morphosemantic extension using the dialectal augmented patterns (see 3.6.3.3), for example, Cairene *?it?ara* 'to be read', Gulf *garra* 'cause to read'.

3.6.3.2.2 Verbs with $C_I = /?/$

Here, there have been different treatments. Some dialects replaced /?/ in p-stem forms by lengthening the /a/ prefix vowel, viz. *ya?kul* → *ya:kul* or *ya:kil*. The resulting p-stem *-KvL-* then formed the basis for a dissimilated s-stem *KaL-*.

So in modern Cairene and many other present-day dialects, *kal* 'he ate' and *ya:kul* 'he eats' contrast with CLA/MSA *?akala* and *ya?kulu*. In other parts of verb morphology, the phonological circumstances causing the loss of /?/ were different and entailed a different morphological result. In Pattern II verbs, forms such as *yu?akkil* 'he feeds' and *yu?addi:* 'he sends' became *yuwakkil* and *yuwaddi*. The changes were then generalized to the s-stem to produce *wakkal* and *wadda* with the notional roots *WKL* and *WDY*. Other dialects (e.g., Omani, some Bahraini) replaced /?/ with /w/ in the p-stem, viz. *ya?kul* → (*yawkul*) → *yo:kul*, resulting in a merger of the resulting forms with those in which $C_I = /w/$, for example, *yo:ṣal* 'arrive' (see n. 11).

3.6.3.3 Loss of some augmented patterns, extension of others

It will be obvious from the exposé of the CLA/MSA system of stem augmentation that there is a certain amount of redundancy in the morphological encoding of grammatical categories such as causative, passive, reflexive, and so on. In general the dialects shed a good deal of this duplication of function in the course of their development and have also tended to narrow the range of types of meaning associated with a particular type of stem augmentation. The end result is that the augmented patterns that were retained from the CLA system, and in some cases massively expanded in scope, are more predictable in the way they modify the root meaning. As in other parts of verb morphology, however, the extent to which the CLA system is still reflected in a particular dialect depends on its history: the urban dialects and those nonurban dialects (such as those of the Gulf states) that have had a history of foreign-language and other dialectal contact show a greater degree of innovation than the Bedouin ones, which are closer to the CLA system. In particular, large numbers of foreign borrowings relating to a sedentary, literate culture have been assimilated into Pattern II and its reflexive/mediopassive correlate, Pattern V.

3.6.3.3.1 Passives, mediopassives and reflexives

One of the consequences of changes in the vowel systems of the dialects was, as we have noted, the loss of the "internal" passive. Only in (some) Arabian dialects does this means of passivization still seem to be more or less functional, and even here there has been considerable innovation and specialization of pattern function, as we shall see below. Everywhere outside Arabia (and now to a great extent there too), the passive is expressed by affixation, most commonly by the prefixing of *?it-* or *?in-* to any Pattern I stem, for example, from *ŠḤN* 'load up' Cairene *?itŠḤN*, Bahraini *(?i)nŠḤN* 'be loaded up'. The *t* stem-affix is

of course an important and flexible element in the augmented pattern morphology of CLA: through prefixation to Pattern II and III verbs, and infixation in Pattern I, it adds a reflexive or mediopassive dimension to the meaning. It is not therefore surprising that its use should have been extended also to the "true" passive in some dialects such as Cairene. In Cairene, the *t-* prefix of Patterns V and VI has in fact become identical to *ʔit-* of the innovated passive so that there has developed in some roots a regular set of morphosemantic relationships such that:

	UNPREFIXED		PREFIX *ʔit-*
(Pattern I)	transitive	→	passive
	↓		
(Pattern II)	causative	→	reflexive (Pattern V)
Example:	ʿRD	→	ʔitʿRD
	'show'		'be put on show'
	↓		
	ʿRRD	→	ʔitʿRRD
	'expose'		'be vulnerable'

or, in the case of roots in which Pattern II carries an intensive rather than causative meaning:

ʔTL	→	ʔitʔTL
'kill'		'be killed'
↓		
ʔTTL	→	ʔitʔTTL
'kill in numbers'		'be killed in numbers'

The *n-* prefix to Pattern I, forming Pattern VII, also formed part of the CLA system, and its use often overlapped with that of the internal passive. It was then to be expected that with the loss of the internal passive for phonological reasons, this *n-* prefix would assume a heavier functional load. This seems to be what has happened. The dialects have simply "recycled" this affix, massively extending its use in order to compensate for the loss of the internal passive. It can be attached to virtually any Pattern I verb in those dialects in which it, rather than prefix *ʔit-*, has become the marker of "true" passivization. Here are some examples of *n-* prefix passive verbs taken from Bahraini Arabic, which frequently, in the p-stem, carry a sense of "passive potential" as well as of "passive actual":

SMʿ 'hear'	nSMʿ 'be heard/audible'
ŠRB 'drink'	nŠRB 'be drunk/potable'
ŠYL 'remove'	nŠYL 'be removed/removable'

TRS 'fill'	nTRS 'be filled/fillable'
FŠG 'smash'	nFŠG 'be smashed/smashable'
DRY 'know'	nDRY 'be known/knowable'

The dialects of Iraq and Arabia in general tend to reserve the *n-* for "true" passives, and the prefix *t-* is used for reflexive or mediopassive correlates of Pattern II. In a few verbs, infixed *-t-* is applied to Pattern I (forming an analog to CLA/MSA Pattern VIII) but seems to be a "true" passive, interchangeable with the *n-* prefix. The semantic differences in *n-* and *t-* prefixed patterns is illustrated by the root *BṬṬ* 'break, burst': *nBṬṬ* 'be broken, smashed' is used to describe the result of an action external to the subject of which it is predicated, as in the description of a motorcyclist's head striking (or being struck by) the curb as a result of a traffic accident and 'bursting open'; *tBṬṬ*, on the other hand, is a reflexive, used for example to describe the bud of a flower 'bursting open'. These examples from Bahraini Arabic show semantic contrasts of prefixed versus infixed *-t-*:

t- prefixed to Pattern II/quadriliteral stems:

BṬṬṬ 'break open'	tBṬṬṬ 'burst open'
GDDD 'gnaw'	tGDDD 'be eaten away'
BNND 'close (trans.)'	tBNND 'be closed, close (by itself)'
BṬṬL 'open'	tBṬṬL 'be opened, open itself'
XSSS 'ruin'	tXSSS 'go off, spoil'
GHWY 'give coffee'	tGHWY 'take/be given coffee'

-t- infixed to Pattern I, interchangeable with prefix *n-*:

BWG 'steal'	BtWG (or nBWG) 'be stolen'
FČČ 'open, release'	FtČČ (or nFČČ) 'be released'
FṢD 'bleed (trans.)'	FtṢD (or nFṢD) 'be bled'

The passive prefixes in the dialects are applied in other ways that have no analogs in CLA. Intransitive Pattern I verbs may be prefix-passivized to give an impersonal meaning, for example (see n. 5), as the following examples make clear:[29]

Cairo, from *ʔʿD* 'sit', *ʔitʔʿD* 'be sat next to':
di lbint illi yitʔiʿid maʿaːha
'This is the girl to be seated with'
(lit. 'This (is) the-girl which it-is-sat with-her')

Baghdad, from *NWM* 'sleep', *nNWM* 'be slept':
halbeːt maː yinnaːm biːh

'This house can't be slept in'
(lit. 'This the-house not it-is-slept in-it')

Bahrain, from *ŠYX* ('alā) 'to be too much (for)', *nŠYX* 'be overcome':
ma: yinša:x 'ale:h ilmaraḍ
'The disease can't be overcome'
(lit. 'Not it-is-triumphed over it, the disease')

Even more strikingly divergent from CLA is the flexible prefixing of the *n*-element to augmented patterns, although this is unusual. For example:

'Ibrā', Oman, prefixed to Pattern II *'WWR* 'hurt', *n'WWR* 'be hurt':
ʔin'awwar 'ala lḥadi:d
'He hurt himself on the iron (tent pole)'

Ha'il, central Saudi Arabia, prefixed to Pattern VI *tṬa:RD* 'pursue each other',
ntṬa:RD 'be pursued, each by the other':
yintta:radu:n
'They can be pursued by each other'

3.6.3.3.2 Causatives and factitives

We saw in the description of the CLA/MSA system that causative and factitive meanings could be expressed by both Patterns II and IV. Sometimes a difference in connotation exists between the two, especially if they are applied in the same root, for example, from *NZL* 'go down' MSA Pattern II *NZZL* is associated with 'bringing down' prices, whereas CLA/MSA Pattern IV *ʔNZL* often collocates with divine agency, as in the 'sending down' from heaven of holy writ; but sometimes very little difference can be discerned between the two patterns, as in *XBR* 'know', which gives *XBBR* and *ʔXBR* both with the same meaning 'inform'. The dialects taken as a whole have virtually eliminated Pattern IV from their repertoire of devices for meaning augmentation. In some roots, where the CLA causative sense had become opaque or was a noncausative denominative, the stem was simply absorbed into dialectal Pattern I, for example, CLA *ʔḤBB* 'like', Cairene *ḤBB*; CLA *ʔ'JB* 'please', Cairene *'GB*; CLA *ʔMṬR* 'rain', Gulf dialects *MṬR*. In the vast majority of CLA Pattern IV cases, however, in which the augmentation was causative in meaning, the dialects simply eliminated such verbs from their base vocabulary, either by using a preexisting Pattern II verb (such as *XBBR*) or creating one if none existed. The following, for example, are common non-CLA causatives shared by all Gulf dialects:

I	II	compare: CLA IV
ṬL' 'go up, out'	*ṬLL'* 'expel'	*ʔṬL'*

G'D 'sit'	*G"D* 'sit (s'one) down'	*ʔQ'D*
RWY ('see')	*RWWY* 'show'	*ʔR?Y*
T'B 'be tired'	*T"B* 'tire (s'one)'	*ʔT'B*
ṬFY 'go out (fire)'	*ṬFFY* 'extinguish'	*ʔTF?*
DRY 'know'	*DRRY* 'cause to know'	*ʔDRY*
XŠŠ 'enter'	*XŠŠŠ* 'insert'	—

3.6.3.3.3 Intensives/extensives

The intensive/extensive meaning potentials of Pattern II, relatively rarely employed in MSA, where the causative and denominative possibilities of this pattern have been dominant in the coining of neologisms, are still operational in the dialects and carry one of the CLA uses of this pattern into roots where it had not been applied. The following are examples of dialectal verbs that have developed a non-CLA Pattern II "intensive" meaning from a CLA Pattern I:

Bahrain:

DHČ 'laugh'	*DHHČ* 'laugh a lot'
DFG 'spill'	*DFFG* 'spill all over the place'
TRK 'leave'	*TRRK* 'leave in large numbers'
LGṬ 'pick up'	*LGGṬ* 'pick up here and there'
GṢṢ 'cut'	*GṢṢṢ* 'cut in swathes'
HYM 'wander'	*HYYM* 'mill around in large numbers'
BY' 'sell'	*BYY'* 'sell in quantities, sell off completely'

Damascus:

BRM 'twist'	*BRRM* 'twist tightly'
ṢRX 'shout'	*ṢRRX* 'shout and bellow'
RBṬ 'tie'	*RBBṬ* 'tie tightly'
DFŠ 'push'	*DFFŠ* 'push and shove'

Increasingly, the intensive meaning of Pattern II is being lost in "educated" dialectal speech and replaced with Pattern I + adverbial complement. It is still much used by the uneducated, however, especially to lend color to narratives, jokes, and other species of oral performance.[30]

3.6.3.3.4 Applicatives/denominatives

A third major application of Pattern II in the dialects is, as in MSA, the formation of denominative verbs. In the dialects, however, unlike in MSA, such verbs may be intransitive as well as transitive, and foreign borrowings are particularly prone to this kind of development:

Bahrain:
source noun

bagil 'greens'	*BGGL* 'go around selling vegetables'
tabil 'drum'	*ṬBBL* 'drum'
'irs 'marriage'	*'RRS* 'get married'
guṭ'a 'piece'	*GṬṬ'* 'sell by the piece, retail'
zita:t 'quick'	*ZTTT* 'rush'
bi:ma (Pers.) 'insurance'	*BYYM* 'insure'
rang (Pers.) 'color'	*RNNG* 'paint, color'
check (Eng.)	*ČYYK* 'check'
finish (Eng.)	*FNNŠ* 'fire, sack'
foot (measure) (Eng.)	*FWWT* 'measure'

Cairo:
source noun

ze:t 'oil'	*ZYYT* 'oil'
buxa:r 'vapor'	*BXXR* 'fumigate'
bla:ṭ 'paving stone'	*BLLṬ* 'pave'
zarr 'button'	*ZRRR* 'button up'
sergi (Turk.) 'register book'	*SRRK* 'register (mail)'
foul (Eng.)	*FWWL* 'play dirty'
shoot (Eng.)	*ŠWWT* 'shoot (soccer)'

There are relatively few examples in the dialects of the ascriptive use of Pattern II and even fewer that seem to be independent dialectal developments that had no CLA etymon, but the following two interesting examples are taken from (Shiʻi) Bahraini dialect: *BWK* (or *BWG*) 'to steal' (CLA *BWQ* 'cheat') → *BWWK* (or *BWWG*) 'to call someone a thief, to think something an example of theft', as in:

Wife recounting what she said to a husband who had accused her of stealing:

*ma: mahratk ʔilla **tibawwikni** bi lbe:za:tə*
'Is that all you can do, accuse me of stealing the money?'
(lit. not your job except you-accuse-me-of-stealing the money?)

Two jewelers discussing a theft at a relative's shop in Abu Dhabi:
— *issana ibta:g min dikka:na θala:θ milyo:n*
'This year three million was stolen from his shop'
— *šilli **bawwaga***
'What made him think it was a theft?'

(lit. what that made-him-think-it-theft?)
— *ma yaw ʔilla lja:m maksu:ra, w illi fi lamya:z killa ra:ḥ*
'When they got there they found the window smashed and everything in the display cases had gone'

In the remainder of the system of triliteral augmented patterns, the dialects as a whole differ little from CLA, from either the formal or semantic point of view. What differences there are tend to be local rather than general.[31]

3.6.3.3.5 Quadriliteral verbs

Quadriliterals have been another area of growth in the verb morphology of all the dialects, and, exactly like Pattern II to which they are morphologically (and possibly historically) akin, they have a frequentative, intensive, or extensive meaning. Reduplication of C_1 of an existent triliteral Pattern I is one of the most productive procedures. For example:

	PATTERN I	QUADRILITERAL
Bahrain/Baghdad	*LMM* 'collect'	*LMLM* 'collect here and there'
Bahrain	*XMM* 'sweep'	*XMXM* 'sweep a lot'
Bahrain/Baghdad	*GṢṢ* 'cut'	*GṢGṢ* 'snip, chop up'[32]
Bahrain	*ḤWČ* 'weave'	*ḤWḤČ* 'wriggle about'
Damascus	*FRḤ* 'rejoice'	*FRFḤ* 'rejoice a lot'
Damascus/Cairo	*ʔR'* 'knock'	*ʔRʔ'* 'clatter, resound'

In other cases a consonant is inserted after C_1 or C_2, often /w/, /y/, or /r/. In the Bahrain case, a subclass of these formations replaces CLA Pattern IX (inchoatives of color and physical defect):

/w/ insert:

Bahrain	*DXL* 'enter'	*DWXL* 'interweave'
	FŠḤ 'open (legs)'	*FWŠḤ* 'become bowlegged'
	ḤMR 'red'	*ḤWMR* 'become reddish'
	XDR 'green'	*XWDR* 'become greenish'
Damascus	*D'S* 'tread on'	*D'WS* 'tread all over'
	ŠLḤ 'take off (clothes)'	*ŠLWḤ* 'take off and throw around'

/r/ insert:

Bahrain	*FG'* 'break'	*FRG'* 'crack (knuckles)'
Bahrain/Baghdad/	*XBṬ* 'bang (trans.)'	*XRBṬ* 'mess up, ruin'

Damascus	*XMŠ* 'scratch'	*XRMŠ* 'scratch'
	ŠBK 'involve'	*ŠRBK* 'entangle, complicate'
	KDS 'pile up'	*KRDS* 'pile up in heaps'
/y/ insert:		
Bahrain	*ṬRB* 'be joyful'	*ṬYRB* 'perform a wedding dance'
	BXL 'be a miser'	*tBYXL* 'be stingy'

Large numbers of quadriliterals, simple or augmented, have also been formed from foreign borrowings, as has also occurred in the case of Pattern II. Whereas in MSA the borrowed foreign elements have tended to relate to imported abstract or scientific processes of one kind or another (*FLSF* 'philosophize', *NṢRN* 'Nasserize', *ʿLMN* 'secularize' *ʔKSD* 'oxidize', *KBRT* 'vulcanize'), in the dialects they refer to imported concrete objects or patterns of everyday activity, the source language usually being that of a former colonial power or, increasingly nowadays everywhere, (American) English:

tirba:l 'rubber sheet' (Eng. 'tarpaulin') →*TRBL* 'rubberize, line with rubber' (Bahrain)

birwa:z 'frame' (Turk. *pervaz* 'fringe, border') →*BRWZ* 'frame' (Cairo)

baxši:š 'tip' (Turk./Pers. *baxšiš* 'gift') →*BXŠŠ* 'tip' (Damascus)

sirkl 'circle' (Fr. *cercle* 'circle') →*SRKL* 'surround' (Morocco)

bigbus 'big shot' (Amer. Eng. 'big boss') →*tBGBS* 'act big' (Morocco)

karda 'card' (Eng./Fr. 'card'/'*carte*') →*KRDY* 'pay by credit card' (Tunis)

In conclusion, the reader should perhaps be reminded that the description given here of dialectal augmented pattern morphology is based on usage at the informal/uneducated end of the scale of formality. Augmented pattern usages that have no direct correlates in the MSA/CLA lexicon are salient to more educated speakers and tend to be replaced by MSA equivalents, or otherwise neutralized when, for whatever contextual reason, more attention is being paid to the form of what is being said.

Notes

1. Greenberg 1978, 431.
2. Bobzin 1980, 39, 65–68.
3. Fleisch 1979, 231.
4. Blohm 1990, 19, reporting research by Uth (1987), states that 61 percent of Pattern II verbs in common use in MSA are causative in meaning, 11 percent are intensive, 3 percent are causative-intensive, and 25 percent cannot be assigned to any of these categories.

5. Intransitive verbs denoting actions can be passivized if no subject can or need be specified, for example, from *XaRaJ-* 'go out' *XuRiJ-* 'be gone out from', as in *xurija mina lġurfa* 'the room was vacated' (lit. 'it was gone out from the room'); from *ĠaḌiB-* 'be angry' *ĠuḌiB-* 'be caused anger', as in *ġuḍiba ʿalayhi* 'there was anger at him' (lit. 'it was angered at him'). The only Pattern I verbs that cannot be passivized by internal vocalization are those that denote permanent states of being.
6. The internal passive pattern is not normally applied to Patterns VII and IX. Pattern VII is intrinsically passive, and IX denotes states of being.
7. Pattern IV could arguably be placed in the same category as II and III if an unattested historical **-uCaCCiC-* p-stem is assumed to lie behind synchronic *-uCCiC-*.
8. Where, in postpausal position, this results in an impermissible initial consonant cluster (as it does in Pattern I and Group (i) of the augmented patterns), a prosthetic *ʔv-* is prefixed, the vowel being determined by regular rule: in Pattern I, the postpausal vowel is /u/ if the p-stem theme vowel is /u/; otherwise, it is /i/, as it also is in Group (i) of the augmented patterns. In Pattern IV the vowel is /a/. In juncture, the final vowel of the preceding word removes the need for prosthetic *ʔv-*; but if the preceding word ends in a consonant, an /i/ is supplied to make the liaison. Pattern IV is apparently exceptional in that the /ʔ/ of the prosthetic element is stable in juncture. However, it could be argued that the *ʔa-* element in pattern IV is not prosthetic at all, but structural and etymological.
9. Cantarino 1974–75, 1:78.
10. Fleisch 1979, 360.
11. Rabin 1951, 158, citing ancient poetry, gives a number of different treatments of these verbs, related to the tribal affiliation of the poet. The differences he reports have continued into the present era in the modern Arabic dialects. For example, Prochazka 1988, 68 reports that in Saudi Arabia *yawCvC, yo:CvC, yu:CvC, yiCvC*, and *ya:CiC* are each, in a different area of the country, the dialectal paradigm forms of the p-stem. See also Meissner 1901, 154–55, 160, where a similar pattern of variability is noted for the dialects of southern Iraq.
12. There are no verbs with C_I = /y/ that have a Pattern VIII.
13. Pattern VIII without assimilation of the semivowel is described by various medieval grammarians as typical of tribes in western Arabia (Rabin 1951, 158, Fleisch 1979, 358–59) and is reported by Reinhardt (in p-stems only) for sedentary Omani dialects (Reinhardt 1894, 199–200). In the mid-1970s I noted a couple of s-stem examples in texts from the village Shiʿi dialects of Bahrain, for example, *iwtaʿe:na* 'we became aware, realized'. See Holes 2001, 562.
14. Angoujard 1990, 14–15.
15. For all intents and purposes, the morphology of CLA and MSA is the same. The differences between the two are largely questions of usage, style, and lexicon. In morphology, CLA makes us of some patterns that, although theoretically available to the modern writer, have dropped out of common use.
16. The variants of the prefix of the p-stem verbs that have C_I = /w/ is just one case in point: *yaw-, ya:, yay-* and *yi:-* are all attested by Sībawayhi (d. 794) as early tribal forms.
17. Wright 1896–98, 2:292–93.

18. This is true of "real duals," that is, enumerative expressions such as 'two cars', 'two houses'. Most dialects have a separate category of dual expressions, which Blanc terms "pseudo duals." These are nonenumerative, mostly natural pairs such as 'two eyes/hands/ears', etc., and (in the urban dialects) normally require feminine singular concord. See Blanc 1970.

19. But when bound pronoun objects are suffixed, these final vowels are pronounced long.

20. Harrell 1962, 41, 46 for Fez, Rabat, and Casablanca; Singer 1984, 320, 325 for Tunis.

21. All the Ha'il forms are taken from Abboud 1964.

22. Holes 1990, 204.

23. Some of the verbs that have changed category from CLA /a/-stem to the Cairene *i–i* class: *TiLi-*, *SiKiT-*, *HiRiB-*, *ĠiRiB-* have free variants in *a–a* even in the plain colloquial: *TaLá-*, *SaKaT-*, etc. (see Hinds and Badawi 1986) that echo the CLA form. The *a–a* variant on the analysis presented here would be the older CLA-descended (and, in terms of the Cairene system, anomalous) one, and the *i–i* variant a more recent innovation reflecting a (possibly still ongoing) general realignment of stems whereby *a–a* = transitive and *i–i* = medial passive/reflexive.

24. Alternatively, it is possible that the dialectal *u–u* verb class arose via (i) the coalescence of CLA /i/- and /u/-stems to *i–i* dialectal stems (ii) the resulting *i–i* stems developing *u–u* variants in backed and/or labial environments

25. Johnstone 1967, 11.

26. That is, the dialect of uneducated speakers in the Sunni communities of old Muḥarraq town, al-Ḥidd, West Rifāʿ, Budayyaʿ, and a few other enclaves whose inhabitants are entirely or predominantly Sunni and distantly descended from Najdi populations. The Baḥārna (Shiʿi) Arabic-speaking communities speak a different kind of dialect (see Holes 1983a, 1984, 1987 for details).

27. Blau 1966–67, 167–68.

28. ʿAbdurraḥmān Rafīʿ, a Bahraini Sunni poet, has written a comic ode on the problems of getting married rhyming in -*e:t* that achieves much of its humorous effect through the extensive use of verb forms of this "vulgar" type (Rafīʿ 1971, 29–36).

29. In the rest of this chapter, the dialectal examples are taken from the following sources: Bahrain, personal field notes made in 1977–78; Cairo, Hinds and Badawi 1986; Baghdad, Erwin 1963; Damascus, Cowell 1964; Morocco, Heath 1989; Tunis, Singer 1984; Oman, personal field recordings, 1986; Ha'il (Saudi Arabia), Abboud 1964.

30. See Holes 1995a, 61–62.

31. For example, in Bahrain, Pattern X is highly productive and often has an inchoative meaning not common in this pattern in CLA or in other dialects, for example, *stMLL* 'get bored', *stGWY* 'gain in strength', *stHMG or stĠYð* 'get angry', *stḤBB* 'come to love', *stĠBB* 'get pregnant'. In other cases it is equivalent to a CLA reflexive or mediopassive Pattern V or VIII: *stḤFL* 'celebrate' (CLA *ḤtFL*), *stXBR* 'know about, experience' (CLA *XtBR*), *stḤJJ* 'argue, protest' (CLA *ḤtJJ*), *st'YŠ* 'live on, exist' (CLA *t'YYŠ* or *'tYŠ*).

32. In Bahraini, note the difference between *GṢGṢ* 'snip, chop up', which is an "intensive" extension of *GṢṢ* 'cut', and Pattern II *GṢṢṢ* 'cut in swathes, cut a lot', with an "extensive" modification of the Pattern I meaning.

4

Noun Morphology
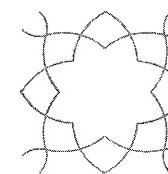
In this chapter we shall proceed in the same way as in the last, dealing first with the derivational aspects of noun morphology ("morphosemantics") and then with the inflections that indicate syntactic function in the sentence ("morphosyntax"). CLA/MSA and the dialects differ from each other little from the point of view of noun derivation, the dialectal patterns representing a subset of those of the CLA parent system and of its modern descendant MSA, although some of these shared patterns have been extended in the dialects in ways that had no precedent in CLA. Because of their similarity, it will be presentationally simpler to deal with the derivational noun morphology of CLA/MSA and the dialects together, noting the differences where appropriate. However, the morphosyntax of the noun (and verb) in the dialects differs considerably from that of CLA/MSA, reflecting the profound changes that have occurred over time in sentence structure as CLA became a purely written medium and the spoken dialects spread, evolved, and displaced the indigenous languages of the pre-Islamic Middle East and North Africa without being exposed, until very recently, to the constraining influence of mass literacy. We see here that the linguistic unity of Arabic at the formal level, over long periods of time and across vast areas of geographical space, resides more in the unchanging morphological principles for derivation and word coinage than it does in any continuity of syntactic structure. However far the dialects have

developed away from the inflected, "synthetic" syntax of CLA, they have hardly altered these morphological principles at all.

4.1 The verbal noun

The Arabic verbal noun corresponds both to the English infinitive in sentences such as 'he likes to eat sweets' and to the gerund in 'eating sweets is his favorite pastime'. As in English, Arabic verbal nouns may refer either to the actual occurrence of an action or to the possibility of its occurrence—for example, 'they tried to kill him' (i.e, they failed) and 'it was a mistake to kill him' (they succeeded)—and both to an activity per se and to its concrete result, for example, 'writing is fun' (process) and 'I don't like his writing' (product). And again, as in English, Arabic noun phrases equivalent to 'his killing' that employ the verbal noun are ambiguous when taken out of context, as to whether 'he' did the killing or was killed. In both MSA and the dialects, however, different structures are often used to disambiguate examples of the last kind (although the MSA and dialectal means of disambiguation are different).

A variety of consonant–vowel patterns are used, largely unpredictably, for the verbal nouns of Pattern I verbs, whereas those that go with the augmented patterns are virtually completely predictable. Some forty-four Pattern I verbal noun types are listed in Wright's grammar of CLA,[1] of which the commonest in modern use are:

CaCaC	'amal 'work', ṭalab 'demand', karam 'generosity'
CaCC	fatḥ 'opening', qatl 'killing' ṣabr 'patience'
CuCC	ḥukm 'judgment', ḥuzn 'sadness' šurb 'drinking'
CiCC	ḥifḏ 'keeping', 'ilm 'knowledge', ḏikr 'mentioning'
CaCa:C	sama:' 'hearing', naja:ḥ 'success', fasa:d 'corruption'
CaCa:Ca	naḏa:fa 'cleanness', ʔaṣa:la 'purity', sala:ma 'safety'
CiCa:C	bina:ʔ 'building', liqa:ʔ 'meeting'
CiCa:Ca	kita:ba 'writing', qira:ʔa 'reading', zira:'a 'farming'
CuCu:C	ruku:b 'riding', wuṣu:l 'arrival', duxu:l 'entrance'
CuCu:Ca	suhu:la 'easiness', ṣu'u:ba 'difficulty'
CvCCa	xidma 'service', 'awda 'return'

Some of these are associated with particular kinds of meaning. CuCu:C, for example, is frequently the pattern employed for intransitive verbs of motion: in addition to those noted above, one notes ruju:' 'return', nuzu:l 'going down', ṭulu:' 'going up', xuru:j 'going out', hubu:ṭ 'landing', etc. CiCa:Ca is associated with professions and crafts, for example, ṣina:'a 'industry', nija:ra 'carpentry',

tija:ra 'commerce', ḥiya:ka 'weaving, knitting', xiya:ṭa 'tailoring, sewing'. CuCu:Ca denotes qualities, physical or abstract, actual or metaphorical: suxu:na 'warmth', buru:da 'coldness, aloofness', xušu:na 'roughness (of material or manners)'. This kind of regular patterning, already recognized by the ancient grammarians, has been extensively exploited by the Arab language academies in recent years in the coining of new words to cover terminology gaps in the sciences.[2] The pattern CuCa:C, for example, which was associated in CLA among other meanings with illnesses and disabilities, as in su'a:l 'cough', zuka:m 'head cold, catarrh', was used to generate many other medical terms, for example, buwa:l 'diabetes' < BWL 'urinate', xuna:q 'diphtheria' < XNQ 'choke', buha:r 'seasickness' < baḥr 'sea', ḥula:q 'sore throat' < ḥalq 'throat', kuta:f 'shoulder pain' < katf 'shoulder'. The pattern CaCaCa:n, associated with verbs of motion but relatively rare in CLA, has also been much applied to provide terms to describe physical processes that involve rapid repetitive action, for example, xafaqa:n 'palpitation (e.g., of the heart)' < XFQ 'vibrate, flap', θawara:n 'agitation, excitement (e.g., of atomic particles)' < θWR 'swirl up, be aroused', nawasa:n 'oscillation' < NWS 'dangle, bob', nabaḍa:n 'pulsation' < NBḌ 'throb', mawaja:n 'undulation' < MWJ 'swell, roll, heave'.

The augmented pattern verbal nouns in MSA are as follows:

II	taCCi:C	tadri:s 'teaching' < DRRS
	taCCiya (weak verbs[3])	tasfiya 'purification' < ṢFFY
III	CiCa:C	difa:' 'defense' < Dv:F'
	muCa:CaCa	muḥa:wala 'attempt' < Ḥv:WL
IV	ʔiCCa:C	ʔiḥba:ṭ 'thwarting' < ʔḤBT
	ʔiCa:Ca (hollow verbs)	ʔida:ra 'management' < ʔDWR
V	taCaCCuC	taqaddum 'progress' < tQDDM
VI	taCa:CuC	tasa:muḥ 'tolerance' < tSMḤ
VII	nCiCa:C	nqila:b 'overthrow' < nQLB
VIII	CtiCa:C	ktiša:f 'discovery' < KtŠF
IX	CCiCa:C	ḥmira:r 'reddening' < ḤMRR
X	stiCCa:C	sti'ma:l 'use' < st'ML
	stiCa:C(a) (hollow verbs)	stifa:da 'benefit' < stFYD

Quadriliteral verb nouns are formed on the pattern CaCCaCa and occasionally CiCCa:C, for example, tarjama 'translation', sayṭara 'control', zalzala/zilza:l 'earthquake', and those of augmented quadriliterals on the pattern taCaCCuC, for example, tamarkuz 'concentration', taqahqur 'retreat, recession'.

In uneducated dialectal speech relatively unaffected by MSA, the verbal noun does not figure very often, other locutions being preferred. It is in fact ques-

tionable whether many of the CLA/MSA forms listed above can be considered part of the dialects at all. In "educated" speech, however, verbal nouns occur more often. On the formal level, Arabian peninsula dialects in general preserve examples of alternative CLA Pattern II verbal noun patterns no longer productive in MSA, viz. *taCCu:C(a), tiCCa:C*, and *CiCCa:C*. These are now very restricted in use and are heard now only from the illiterate old when reminiscing about a preindustrial culture now almost dead.[4] For example:

taklu:f	'causing of nuisance' < *KLLF*
ta'lu:m	'teaching/learning (especially of the Koran)' < *'LLM*
taslu:ma	'wedding dowry' < *SLLM*
tiṣla:ḥ	'mending' < *ṢLLḤ*
tisga:m	'advance paid to pearl divers' < *SGGM*
tiḥla:	'beautification' < *ḤLLY*
tiswa:	'doing, making' < *SWWY*
giffa:l	'close (especially of pearling season)' < *GFFL*
nissa:l	'washing (of a bride's hair in traditional marriage)' < *NSSL*

A more general and important difference between the dialects and CLA/MSA is that those *(ʔi)n-* prefix passives that do not have a CLA etymon (the majority, because these are freely formed in most dialects), and all *ʔit-*prefix passives (in those Egyptian and North African dialects that have them) do not form a verbal noun. In CLA/MSA, the Pattern VII passive verbal noun focuses attention on the result of an action whose agent is unspecified. For example:

nqiṭa:'u l   *ma:ʔ*   *ka:riθa*
being-cut (of) the-water disaster
'The cutting off the water is/was/will be a disaster'
(e.g., the crops will die from the resultant lack)

Here *nqiṭa:'* refers to the result of an action that may have a human agent (e.g., the deliberate draining of the marshes of southern Iraq) or a natural one (e.g., a river drying up as a result of no rainfall). In contrast, the Pattern I verbal noun focuses on the action itself:

lam yuwa:fiq   *'ala: qaṭ'i*   *lma:ʔ*
not 3msg-agree to   cutting off (of)   the-water
'He didn't agree to the cutting off of the water'
(e.g., because it would be a spiteful action)

Passivized nominal constructions of the type exemplified in the first example are not really part of dialectal Arabic, a simple Pattern I verbal noun plus di-

rect object being preferred, although of course phrases of the former kind do occur as "borrowings" from MSA in educated speech. In the plain colloquial of the Gulf, the noun clause in parentheses containing a prefix-passive verb in:

ma: yiṣi:r   *(tinʔakil   halašya)*
not 3msg-is-possible (3fsg-are-eaten these-the-things)
'These things can't be eaten'

is replaceable by a Pattern I verbal noun:

ma: yiṣi:r (ʔakil halašya)
........(eating (of) these-the-things)
'These things can't be eaten'

The verbal noun of the reflexive/mediopassive Pattern VIII is also little used in "plain" colloquial because, as is often the case, there is a noun derived from a Pattern I or II verb (or from no verb at all) with a similar meaning. So the verbal noun corresponding to Pattern VIII *ftakar* 'think' is, depending on whether "process" or "product" is intended, *tafki:r* 'thinking' (Pattern II verbal noun) or the noun *fikr* 'thought', and that corresponding to *štara* 'buy' is *šira:ʔ* 'buying' or *muštara* 'purchase(d item)', which is a passive participle (see below).

4.2 Participles

CLA/MSA has two types of participle: active and passive. Their morphological shape is completely predictable. Both have a variety of syntactic (and pragmatic) uses in both CLA/MSA and dialectal Arabic, although these are not exactly the same in the two kinds of Arabic. The basic difference between the two types of participle is that the active describes the state in which the subject of the verb from which it is derived finds itself as a result of the action or event that the verb describes, whereas the passive refers to the state in which the object or complement of the verb from which it is derived finds itself after the completion of the action/event. The participles have no fixed time reference—this has to be interpreted from the context. Thus from *KTB* 'write' the active participle *ka:tib* describes an entity which is/was/will be 'writing', has 'written' or habitually 'writes', whereas the passive participle *maktu:b* describes an entity that is/was the result of 'writing'. The active participle can function syntactically as a noun, verb, or attributive adjective. For example:

	huwa ka:tib *'aði:m*
noun	he write-A.PART great

'He is a great writer'

huwa   ka:tibuhu

verb   he write-A.PART-it

'He is writing/has written it'

?a:la   ka:tiba

adjective   machine write-A.PART-f

'Typewriter'

whereas the passive participle is often used predicatively as quasiverbal adjective to indicate the result or present relevance of a completed action. For example:

kullu ttaqa:ri:r maktu:ba

all the-reports write-P.PART

'All the reports have been written'

or as an attributive adjective:

ṣadara   baya:n   maktu:b

came-out-3msg announcement write-P.PART

'A written announcement has come out'

Like active participles, all passive participles may function as substantives (e.g., *maktu:b* also means 'letter') but in this case denote the result, concrete or abstract, of an action. In many verbs indeed, this is perhaps their commonest use, e.g., *mafhu:m* 'concept' (= (thing) understood) < *FHM* 'understand', *mantu:j* 'product' < *NTJ* 'produce', *mashu:q* 'powder' < *SHQ* 'crush'. The passive participles of augmented patterns also have an important function in CLA/MSA as nouns of time and place (see below).

The complete CLA/MSA paradigm of participles in common use is as follows:

PATTERN	ACTIVE	PASSIVE
triliteral:		
I	*Ca:CiC*	*maCCu:C*
II	*muCaCCiC*	*muCaCCaC*
III	*muCa:CiC*	*muCa:CaC*
IV	*muCCiC*	*muCCaC*
V	*mutaCaCCiC*	*mutaCaCCaC*
VI	*mutaCa:CiC*	*mutaCa:CaC*
VII	*munCaCiC*	*munCaCaC*
VIII	*muCtaCiC*	*muCtaCaC*
IX	*muCCaCC*	—

X	*mustaCCiC*	*mustaCCaC*
quadriliteral:		
I	*muCaCCiC*	*muCaCCaC*
II	*mutaCaCCiC*	*mutaCaCCaC*

As with the verbal noun, the demands of science, technology, and western material culture in the modern period have led to a large-scale exploitation of the potential of this morphological resource. The active participle pattern has been used for coining nouns denoting culturally or scientifically novel agents, human or mechanical, whereas the passive participle pattern has provided words for nonhuman, man-made entities, concrete or abstract. The passive participles of verbs that already have an inherently passive, mediopassive, or reflexive meaning—principally verbs of Patterns V, VII, VIII, and X—were relatively rarely used in CLA except substantivally as nouns of time or place, and this potential has been revived in the modern period. The following will give some idea of the range of modern participial coinages or meaning extensions:

I

ḌĠT 'press'   → *ḍa:ġiṭa* 'compressor' (A.PART)

'JN 'knead'   → *ma'ju:n* '(tooth)paste' (P.PART)

II

KBBR 'make bigger'   → *mukabbir* 'loudspeaker' (A.PART)

?YYD 'support'   → *mu?ayyid* 'fan, supporter' (A.PART)

XYYM 'make camp'   → *muxayyam* 'camp' (P.PART)

III

Qv:TL 'fight'   → *muqa:tila* 'fighter (plane)' (A.PART)

Qv:WM 'resist'   → *muqa:wim* '(electrical) resistor' (A.PART)

'v:ML 'treat, deal'   → *mu'a:MiL* 'coefficient' (A.PART)

IV

?RSL 'send'   → *mursila* 'transmitter' (A.PART)

?RFQ 'append'   → *murfaqa:t* (P.PART-pl) 'enclosures' (e.g., in a letter)

V

tṬLLB 'demand'   → *mutaṭallaba:t* (P.PART-pl) 'requirements' (e.g., for a university course)

VI

t'v:WN 'cooperate'   → *muta'a:win* (A.PART) 'member of a cooperative society'

VII

nḤDR 'decline'   → *munḥadar* 'slope' (P.PART)

nḤNY 'deviate'   → *munḥana* 'curve' (P.PART)

VIII

LtQY 'meet together'	→	*multaqa* 'crossroads' (P.PART)
StWY 'be even, level'	→	*mustawa* 'level' (P.PART)
JtM' 'assemble'	→	*mujtamaʿ* 'society' (P.PART)

X

stQTB 'attract'	→	*mustaqṭib* 'polarizer' (A.PART)
stWTN 'make one's home'	→	*mustawṭan* 'settlement' (P.PART)

quadriliterals:

LʔLʔ 'shine brightly'	→	*mulaʔliʔ* 'scintillator' (A.PART)
tBRTN 'become Anglicized'	→	*mutabarṭin* 'Anglicized person' (A.PART)

As elsewhere in morphology, the formal differences between the CLA/MSA participle system and that of the modern dialects seem to have originated in the changes that affected the dialectal vowel systems. These changes had a more profound effect on the augmented patterns than on Pattern I.

In Pattern I, the dialectal system maintains the passive/active distinction intact, because here the distinction is carried by two quite different morphological patterns. However, certain concatenations containing vowels or identical consonants that cause surface irregularities in the CLA/MSA system are treated differently in the dialects so that a greater degree of symmetry results. A few examples will be given here. From Pattern I doubled verbs (such as *HTT* 'put'), some dialects generally (Cairo, Damascus, but not the Gulf, North Africa) have derived *Ca:CiC* forms (*ḥa:ṭiṭ* 'putting') rather than CLA/MSA *Ca:CC* (*ḥa:ṭṭ*), thus "regularizing" the doubled verb participle. Pattern I hollow verb passive participles of roots in which C_2 = /y/, such as *mabyu:ʿ* < *BY'* 'sold' (compare CLA *mabi:ʿ*) are permissible in some dialects (Morocco, Jordan, and the eastern Arab World generally) and again regularize what would have appeared as a morphological anomaly. In other dialects (Cairo), this anomaly has been resolved in a different way: hollow verb passive participles are replaced by the *active* participle of the corresponding prefix-passive Pattern VII verb (*minba:ʿ* < *nBY'* 'be sold'). CLA Pattern I verbs in which C_1 = /ʔ/ lost the /ʔ/ in most dialects because it was particularly vulnerable in initial and intervocalic position. For example, *ʔKL* became simply *KL*. In the active participle, the resultant morphological irregularity has been resolved either by substituting /w/ for the missing radical = *wa:kil* (North Africa and Cairo) or by /m/ = *ma:kil* apparently by analogy with the augmented pattern participles (Baghdad and the Gulf). The dia-

lectal developments exemplified illustrate how analogical restructuring has brought "irregular" CLA forms into line with the local paradigm.

In the augmented patterns, there have been wholesale changes. As we have already seen, Pattern IV has been generally lost as a pattern, and most of its semantic coverage shifted to Pattern II. In the rest of the system, phonological changes destroyed the short vowel contrast that uniquely carried the active/passive distinction. In the large-city dialects of North Africa—Casablanca, Fez, Rabat, Tunis, and as far east as Cairo—this distinction has been totally lost. For transitive verbs, the participle is interpretable as active or passive depending on context; for intransitive verbs, it is only active. Although the trigger for these changes was phonological in origin, the concomitant particularization of the augmented patterns described in chapter 3—whereby Pattern II became, even more than it had been in CLA, a causative/denominative transitive pattern, and Patterns V, VII, and VIII reflexive, passive, or mediopassive—meant that, for many verbs, voice distinctions in the participle, as in the finite verb, ceased to have any functional raison d'être. The many verb pairs that developed, one transitive and the other its reflexive/passive correlate, did not both need to have an active and a passive participle. So for example, in Cairene Arabic the form *minaḍḍaf* < *NDDF* (CLA *N ḋ ḋF*) 'to clean (trans.)' can have both an active ('cleaning') and a passive ('cleaned') meaning. Context would normally make clear the desired sense, but where there is a risk of confusion, the Pattern V participle *mitnaḍḍaf* 'having been cleaned' (carrying no active/passive distinction in its voweling but marked as passive by the *-t-* infix) can be used. As in finite verb morphology, the dialects have thus tended to exploit and extend the possibilities of consonantal affixation as an indicator of syntactic function and diminish the role of vowel contrasts. In the case of Pattern III, VI, and IX verbs, their intrinsically conative, reciprocal, and inchoative meanings rendered passive participles redundant. Only in Pattern X is a regular active/passive participial contrast maintained in some dialects, for example, *st'ML* 'use' gives *mista'mil* 'user' and *mista'mal* 'used'.

Syntactically too, participle development in the dialects has been different from that in written Arabic. Whereas in MSA the potential of the participles for noun coinage has been heavily exploited, in the dialects they have become an important element in verb syntax. The verbal value of the active participle was always a part, albeit a relatively minor one, of its range of uses in CLA, but this has become perhaps its main function in the dialects. As we shall see in chapter 6, it sometimes has a "continuous" aspectual meaning and sometimes a "perfective" one, depending on the verb, the context, and the particular dialect;

in some dialects, its perfective use contrasts modally with the s-stem perfective in terms of degree of commitment to the truth value of the proposition being expressed. From the purely morphological viewpoint, however, which is our concern here, the verbal use of the participle is in certain cases distinguished from its nominal uses. Thus in most dialects (Damascus is used here as an example), there is a difference between:

> (noun)
> *hiyye ka:tbit*   *halmaktu:b*
> she write-A.PART (of) this-the-letter
> 'She is the writer of this letter'

in which the genitival relationship between the two nouns is indicated by the -*t* feminine affix and:

> (verb)
> *hiyye ka:tbi*   *halmaktu:b*
> she write-A.PART this-the-letter
> 'She has written this letter'

in which the relationship is verb–object and the -*t* is not pronounced. The masculine forms would, however, not be so distinguished, being in each case *ka:tib*. However, if the feminine form of the participle has a pronoun suffixed to it, it makes no difference if the participle is nominal or verbal, for example, *hiyye ka:tə́bto* 'she is its writer/ she has written it'. Another case of functionally based differentiation is in Baḥārna (= Bahraini Shi'i) dialects where there is a difference in pronoun affixation between:

> (noun + possessive)
> *huwa mu'allimhum*
> he   teach-A.PART-their
> 'He is their teacher'

and:

> (verb + object)
> *huwa mu'alliminhum*
> he   teach-A.PART-them
> 'He has taught them'

where an -*in*- infix obligatorily separates the participle and the pronoun affix.

4.3 Other categories of the singular noun

4.3.1 "Primitives"

By the term "primitive" is meant that category of nouns that are not verbal derivatives and that, unlike those exemplified in 4.3.2 (f)–(i) below may function syntactically only as substantives and not adjectives. Examples are *rajul* 'man', *šams* 'sun', *ma:ʔ* 'water', and *laḥm* 'meat', as well as proper names. Most primitives are grammatically masculine, but some are feminine by form, meaning, or simply convention:

By form
(a) most nouns that have the -*a(t)* suffix, e.g., *sikka* 'alleyway', *ja:dda* 'main thoroughfare', *luġa* 'language'.
(b) those endings in -*a:* or -*a:ʔ, da'wa:* 'claim', *ṣaḥra:ʔ* 'desert'
By meaning
for example, *bint* 'daughter', *ʔumm* 'mother'
By convention
for example, *ʔarḍ* 'earth', *šams* 'sun', *ri:ḥ* 'wind', *na:r* 'fire', *da:r* 'house'; parts of the body that come in pairs, for example, *yad* 'hand', *'ayn* 'eye'; and the names of most countries and cities.

Some other nouns—notably collective nouns (see 4.3.2 (a) and 4.4 below)—are indeterminate as to gender, being sometimes construed as masculine and sometimes feminine.

4.3.2 Derivatives

There are a number of other secondary noun patterns, derived from verbs or other nouns, in which morphological structure is regularly related to semantic function. Syntactically, those in categories (a)–(e) are like primitives in that they can serve only as substantives, whereas those exemplified in (f)–(i) may function either as nouns or adjectives, predicatively or attributively:

(a) Instance nouns
These are formed by added the feminine suffix -*a* to the verbal noun, most productively of Pattern I. Examples from Pattern I are: *ḍarb* 'hitting', *ḍarba* 'a blow'; *qafz* 'jumping', *qafza* 'a jump'; *sariq* 'stealing', *sariqa* 'a theft'; an example from Pattern VIII is *ltifa:t* 'turning round', *ltifa:ta* 'a turn, swivel'. In Baghdadi dialect we have, for example, from Pattern I *dafur* 'kicking', *dafra* 'a kick'; *gaḥḥ* 'coughing', *gaḥḥa* 'a cough'; *čawi* 'burning', *čawya* 'a burn'; and from Pattern II: *ta'bi:s* 'frowning', *ta'bi:sa* 'a frown'; and from Pattern VII *nhina:ʔ* 'bowing', *nhina:ʔa* 'a bow'.

Instance nouns may also be formed from collective nouns in the same way. Thus: *šajar* 'trees', *šajara* 'a tree'; *samak* 'fish', *samaka* 'a fish'; *tamr* 'dates', *tamra* 'a date'.

(b) Nouns of manner

Nouns of manner denote the way in which an action is carried out and are formed on the pattern *CiCCa*. Thus from *MŠY* 'walk' is derived *mišya* 'gait' (compare the ordinary verbal noun *mašy* 'walking'), from *MWT mi:ta* 'manner of death' (*mawt* 'death'), from *JLS* 'sit' *jilsa* 'sitting posture' (compare *jalsa* 'session', *julu:s* 'act of sitting'). This is no longer a productive pattern in either MSA or dialectal Arabic.

(c) Nouns of place and time

There is a very large number of these, many of them neologisms, formed on the patterns *maCCaC*, *maCCiC*, and *maCCaCa* from Pattern I verbs, for example, from *KTB* 'write' *maktab* 'desk, office' and *maktaba* 'library, bookshop'; from *XBZ* 'bake' *maxbaz* 'bakery'; from *ṢN'* 'make' *maṣna'* 'factory'; from *ṬYR* 'fly' *maṭa:r* 'airport'; from *JLS* 'sit' *majlis* 'sitting-room, council'. Some dialect (Bahraini) examples of nouns of place are: *maðʕana* 'summer camp', *maṭraḥ* 'anchorage', *maʔtam* '(Shiʕi) funeral house'. As already noted and exemplified, the passive participles of the augmented patterns are used for this purpose also.

(d) Nouns of instrument

The patterns *miCCaC*, *miCCa:C*, and *miCCaCa* are used, for example, *BRD* 'file' *mibrad* 'file, rasp', *FTḤ* 'open' *mifta:ḥ* 'key, opener', *KNS* 'sweep' *miknasa* 'broom'. This pattern has been particularly heavily exploited in scientific and other modern coinages, for example, *ŠRṬ* 'cut, slit' *mišraṭ* 'lancet', *RQB* 'observe' *mirqab* 'telescope', *ZLJ* 'slide' *mizlaj* 'skate, ski', *ḌRB* 'hit' *miḍra:b* 'mallet', *ṬḤN* 'grind' *miṭḥana* '(food) grinder, mill'. Some dialectal (Gulf) examples are *DWS* 'tread' *midwisa* 'treadle (on a loom)', *FLG* 'split open' *mifliga* 'knife (for opening clams)', *ČWY* 'burn, brand' *mičwa* 'branding iron'.

(e) Nouns of profession/occupation

Just as there are special verbal nouns denoting regular activities and occupations, so there are nominal patterns denoting those who do them. The pattern is *CaCCa:C*. *CaCCa:C* (or *CaCCa:Ca* in its intensified form) was used adjectivally or nominally in CLA to denote a person who carried out an activity to excess, for example, *ʔakka:l* 'glutton' < *ʔKL* 'eat', *kaððða:b* 'liar' < *KðB* 'lie', *'alla:m*/*'alla:ma* 'extremely learned' < *'LM* 'know', *raḥḥa:l*/*raḥḥa:la* 'great traveler, globetrotter' < *RḤL* 'emigrate, set off, roam'. In MSA its main use is as a pattern denoting occupations, for example, *farra:š* 'cleaner', *xabba:z* 'baker', *zajja:j* 'glazier', *najja:r* 'carpenter', *xayya:ṭ* 'tailor', *qaṣṣa:b*/*jazza:r* 'butcher', *baqqa:l* 'greengrocer', *zarra:'* 'farmer', and, interestingly, *qanna:ṣ*, 'sniper' (originally 'hunter'). In the dialects the pattern has also been highly productive, for example,

(Gulf) *naxxa:l* 'date palm grower', *samma:č* 'fisherman', *ḥayya:č* 'sail-maker', *gayya:ð* 'summer pearl diver', *galla:f* 'shipwright', *nahha:m* 'singer (on a pearling dhow)'.

In MSA, the same pattern, usually with a feminine suffix, has been used for coining inanimate nouns of instrument (= nouns that perform a particular action regularly), for example, *dabba:sa* 'stapler', *xarra:ma* '(paper) punch', *'aṣṣa:ra* '(juice) squeezer', *sayya:ra* 'car', *θalla:ja* 'refrigerator', *ġassa:la* 'washing machine', *xalla:ṭ* '(food) mixer', and, more dialectally, *bazza:za* 'baby's bottle, teat' (< *bizz* 'nipple').

(f) Nouns/adjectives describing states/qualities

In MSA there are perhaps a dozen noun/adjective patterns in common use that are derived from the Pattern I stative/inchoative *CaCiC-*/*CaCuC-* s-stem verbs and Pattern IX verbs described in chapter 3, denoting contingent states or permanent qualities. These patterns can be used as adjectives or as (usually defined) nouns whose referent is recoverable from context. For example:

CaCC	*ṣa'b* 'difficult'
CaCaC	*ḥasan* 'beautiful, good (thing/person)'
CaCiC	*fariḥ* 'happy'
CuCC	*ṣulb* 'solid', *murr* 'bitter', *ḥulw* 'sweet'
CaCa:C	*jaba:n* 'coward(ly)'
CuCa:C	*šuja:'* 'brave'
CaCi:C	*kabi:r* 'big, old', *jami:l* 'beautiful', *šari:f* 'noble'
CaCu:C	*ʔaku:l* 'glutton', *ṣadu:q* 'truthful', *ḥasu:d* 'envious'
CaCCa:n	*sakra:n* 'drunk', *ta'ba:n* 'tired', *za'la:n* 'upset'
ʔaCCaC	*ʔaswad* 'black', *ʔaḥmar* 'red', *ʔaxḍar* 'green'

Some of these patterns tend to be associated with specific semantic modifications of the root meaning: *CaCu:C*, like *CaCCa:C*, is an "intensive" pattern, whereas *ʔaCCaC*, as an adjectival derivative of Pattern IX, is limited to certain roots that denote colors or physical defects. Many of them, particularly *CaCaC*, *CaCa:C*, and *CaCi:C*, as well as the active participle *Ca:CiC*, are a source of proper names, for example, *ḥasan* 'beautiful', *jawa:d* 'generous', *rafi:q* 'companion', *ṣa:liḥ* 'honest, true'. These patterns mostly form their feminine by suffixing *-a*, but *ʔaCCaC* has the pattern *CaCCa:ʔ*, for example, *ḥamra:ʔ* 'red' (whence Al-Hambra (= 'the red (one)')), etc.

The pattern *CaCi:C* is also used to derive adjectives from certain transitive verbs and in these cases is used substantivally as the equivalent of a passive participle pattern, for example, (MSA) *qati:l* 'killed (person)' < *QTL* 'kill', *jari:ḥ* 'wounded (person)' < *JRḤ* 'wound', (Bahraini) *nagi:la* 'transplanted (seedling)'

< *NGL* 'transfer, remove'. It is interesting to note here a further instance of the shared morphology of the categories intransitive-stative and passivized-transitive. We earlier noted that a similar coincidence of morphological structure was a consequence of the reorganization of the voweling of Pattern I verbs in Cairene (see 3.6.3.1). And in those Arabian dialects that, unlike Cairene, still have a fully functional internally voweled Pattern I passive verb (e.g., Omani dialects in general) the same process of conflation has happened in a more thoroughgoing way: here, the dialectal passive paradigm is indistinguishable from that of verbs that developed from CLA stative and transitive *agent moyen* verbs, being *CCvC* (where *v* = high vowel determined by consonantal environment) in all cases, for example, *ksir* (passive) 'be broken', *kbur* (stative) 'be/become old', *friḥ* (stative) 'be happy, rejoice', *šrub* (*agent moyen* = subject affected by action) 'drink'; compare CLA *kusira, kabura, fariḥa,* and *šariba.*

Some of the morphological distinctions in the list of CLA example forms have been lost in the dialects because of the short vowel changes and deletions discussed earlier. In many areas, as we have noted, unstressed nonfinal open short vowels have been lost, collapsing the distinctions between *CaCa:C, CuCa:C,* and (the plural pattern) *CiCa:C*; in others (e.g., Baghdad), in which final clusters were broken down by an epenthetic vowel, the *CaCC, CuCC, CaCaC,* and *CaCiC* categories have coalesced into a single *CvCvC* category in which the quality of the vowels is predictable from the consonantal environment rather than the voweling of the CLA etymon. The result has been a blurring of the functional specializations originally associated with these patterns.

Although some pattern distinctions have been lost, others have expanded in compensation. The category *CaCCa:n* has been particularly productive in the dialects and functions as an active participle pattern in verbs denoting temporary states that do not have the *Ca:CiC* pattern. However, a number of verbs have both types of participle, and here *Ca:CiC* and *CaCCa:n* usually have different types of referent, with *CaCCa:n* being predicated of animate beings and *Ca:CiC* of inanimate entities, abstract or concrete. Thus from *BRD* 'be cold' *barda:n* is used to mean 'feeling cold' (of human beings), whereas *ba:rid* means 'cold' as applied to the weather, drinks, or temperament; from *SXN* 'be hot', *saxna:n* means 'feverish', whereas *sa:xin* is 'hot' as applied to liquids, etc. Some examples of dialectal *CaCCa:n* are:

Baghdad: *kabra:n* 'grown big' < *KBR* 'grow big'
            samna:n 'grown fat' < *SMN* 'grow fat'
            ðabla:n 'drooping' < *ðBL* 'droop, wilt'

Cairo: *ʔarfa:n* 'fed up' < *ʔRF* 'become fed up'
            dafya:n 'warm' < *DFY* 'become warm'
            'ayya:n 'poorly' < *'YY* 'fall sick'
Bahrain: *ḥamga:n* 'angry' < *stHMG* 'be angry'
            darṭa:n 'useless' < *DRṬ* 'fart'
            ḥarda:n 'furious' < *HRD* 'become furious'

(g) The pattern *ʔaCCaC (f. CuCCa:)*

As well as denoting colors and physical defects, the pattern *ʔaCCaC* has another, more productive use. Originally in CLA an intensive, and still so used in MSA in a few phrases such as *ʔallahu ʔakbar* 'God is most great!' and (in its feminine form) *ʔadduwalu lkubra:* 'the superpowers', *ʔalḥarbu l'uðma:* 'the Great War', its main use now, in both written and dialectal Arabic, is to express comparative relationships. The pattern can be applied to the active participle *Ca:CiC* and to those categories of adjective listed in (f) above that consist solely of root consonants and long vowels. So from *kari:m* 'generous' is derived *ʔakram*, which, depending on the syntactic environment in which it occurs, can mean both 'more generous (than . . .)' and 'the most generous (of . . .)'. In line with the morphophonological principles outlined in 3.5, adjectives in which C2 and C3 are the same make their comparative form on the pattern *ʔaCaCC*, so from *qali:l* 'few' *ʔaqall* 'fewer' (not *ʔaqlal*); and those in which C3 is a semivowel have a final long /a:/, so *ḥulw* 'sweet' → *ʔaḥla:* 'sweeter'. Adjectives denoting colors and physical defects cannot have this pattern applied to them, because they already are of the *ʔaCCaC* pattern, and neither (in MSA at least) can those of the *CaCCa:n* pattern. The comparative forms of adjectival patterns to which *ʔaCCaC* cannot be applied are formed by a syntactic periphrasis (see 6.1.4).

Some dialects have expanded the range of patterns to which *ʔaCCaC* can be applied so that, for example, in the Gulf it is possible to apply it to *šwayy*, etymologically a diminutive noun meaning 'little' as in:

(Kuwait) *iṭṭo:z   ṣa:r   ʔašwa*
          the-dust became-3msg less
          'The dust storm has abated'

or (Shi'i Bahrain) to the noun *usta:d* 'master' (usually applied to those expert in building, farming, or other manual skills) to generate *ʔastad* 'more skillful'. In Damascene Arabic, borrowings that contain more than three radicals and augmented pattern participles may also form this pattern, for example, *zangi:l* (<

Turk. *zengin*) 'rich', *ʔazangal* 'richer', *mušarṭaṭ* 'ripped', *ʔašarṭaṭ* 'more ripped', as well as some common nouns, for example *rižžaːl* 'man', *ʔaržal* 'more manly'. In some dialects of Palestinian Arabic, there is a contrast between *iḥmar* 'red' and *aḥmar* 'redder'. Such formations are purely dialectal and illustrate once more the flexibility of the dialects in analogical derivation.

(h) Diminutives

The diminutive patterns *CuCayC(a)* and *CuCayyiC(a)* may be applied to nouns and adjectives to denote a small example of something, for example, *jubayl* 'small hill' < *jabal* 'mountain, hill', *kulayb* 'puppy' < *kalb* 'dog', *buḥayra* 'lake' (whence Port. Albufeira) < *baḥr* 'sea', *kutayyib* 'booklet' < *kitaːb* 'book'. In MSA the diminutive is a somewhat marginal form productively, although it survives in many proper names such as the man's name *ḥusayn* 'small beautiful (one)' < *ḥasan* and place names such as *al-kuwayt* 'the small fort' < *kuːt* 'fort'. In some modern (especially Arabian) dialects, however, the diminutive is still in common and productive use, especially in the speech of women, where it may have caritative or pejorative overtones, depending on the context. Words like *bnayya* 'little girl' < *bnayy* 'little boy' + f. suffix *a*, and *ṣbayy* 'little boy' < *ṣabiːy* 'boy' can have either, depending on the actual age and status of the referent, whereas address forms such as *xuwayy* 'little brother' and *xuwayya* 'little sister' are usually terms of endearment.[5]

(i) Relational adjectives and abstract nouns

The relational adjective is formed by attaching the termination *-iːy* to a (usually "primitive") noun, minus any feminine ending. Thus from *šams* 'sun' is derived *šamsiːy* 'solar, photo-', from *baġdaːd* 'Baghdad' *baġdaːdiːy* 'from/of Baghdad', *madrasa* 'school' *madrasiːy* 'scholastic', *ṣinaːʕa* 'industry' *ṣinaːʕiːy* 'industrial, man-made'. Relational adjectives are now formed very freely to cope with the demands of science, technology, the media, and modern life more generally (see chapter 8). The very high productivity of this pattern seems to be a consequence of the general lack in Arabic of processes for word compounding, complex noun phrase formation, or agglutinative morphological processes in general, and the almost daily need, in the press at least, for translation equivalents of English phrases. Relational adjectives provide a simple solution to the translation of compound noun phrases that would otherwise require an awkward and sometimes ambiguous structure or a wordy periphrasis. So, for example, 'artillery and rocket barrage' in newspaper reportage of the Lebanese civil war is rendered:

qaṣf   midfaʕiːy   wa   ṣaːruːxiːy
bombing cannon-REL and rocket-REL

(REL = relational adjective ending), where the adjective *ṣaːruːxiːy* from *ṣaːruːx* 'rocket' has apparently been coined for the occasion to match the (also recently coined) adjective *midfaʕiːy* and enable a straightforward noun–adjective phrase to be formed to render the English noun phrase. 'News blackout' is rendered by a similar structure, this time involving a relational adjective formed from a Pattern IV verbal noun: *taʕtiːm ʔiʕlaːmiːy* ('blackout giving-of-information-REL'); 'countdown' by an adjectivalized Pattern VI verbal noun: *ʔalʕaddu ttanaːzuliːy* ('the-count the-descent-REL'); 'request stop' by a Pattern VIII: *mawqif ixtiyaːriːy* ('stop choice-REL'); 'secessionist movement' by a Pattern VII: *ḥaraka infiṣaːliːya* ('movement being-separated-REL'); 'suicide attack' by a Pattern VIII: *ḥamla intiḥaːriːya* 'attack suicide-REL'[6]). Relational adjectives are also freely formed from plural nouns, whether broken (4.4.2), for example, *barnaːmaj waθaːʔiqiːy* 'documentary program' ('program documents-REL'), *jamʕiːya nisaːʔiːya* 'women's society' ('society women-REL') (but note also *ḥaraka nisawiːya* 'feminist movement') and *kalyuːtaːt rijaːliːya* 'men's underclothes' ('underclothes men-REL'); or suffixed plurals (see 4.4.1), for example, *ʔimaːraːtiːy* ('emirates-REL') 'from the (United Arab) Emirates' as in *muḥaːdaθaːt saʕuːdiːya ʔimaːraːtiːya* 'Saudi–Emirati discussions' and *istixbaːraːtiːy* ('pieces of intelligence-REL') in *taqaːriːr istixbaːraːtiyya* 'intelligence reports'. Hundreds of other such noun–adjective phrases, which provide a shorthand and syntactically convenient way of rendering into Arabic such diverse phenomena as 'sunglasses', 'satellite', 'rubber bullets', and 'shuttle diplomacy' form the stock in trade of media reportage and advertising and thence find their way into everyday speech. Relational adjectives are even formed from interrogative particles: from *kam* 'how much?' we get *kammiːy* 'quantitative', and from *kayf* 'how?' *kayfiːy* 'qualitative'. The relational adjective is also freely applied, with the appropriate phonological modifications, to foreign borrowings: *faransaːwiːy* 'French', *barlamaːniːy* 'parliamentary', *fidraːliːy* 'federal', *kalaːsiːkiːy* 'classical'. In (some of) the dialects, there is a related closed class of relational adjectives in *-aːni*, such as Cairene *awwalaːni* 'first' < *awwal* 'first (part)', *foːʔaːni* 'upper' < *foːʔ* 'above', *taḥtaːni* 'lower' < *taḥt* 'below, under', *guwwaːni* 'inner' < *guwwa* 'inside'.

A large class of nouns is formed by adding the feminine ending *-a* to the relational adjective or simply by adding *-iːya* directly to a nominal base. Such formations generally denote abstractions and first arose in the medieval period when there was a need for terms to translate those of Greek philosophy; more recently, the growth of the social sciences and an international political culture has spawned a huge number of such abstract nouns, which equate to the "-isms" and "-ities" of modern English. So from the "primitive" noun *ʕaql* 'mind' is formed *ʕaqliːy* 'mental' and thence *ʕaqliːya* 'mentality'; from the Pattern VIII ver-

bal noun *štira:k* 'sharing, participation' is formed *štira:ki:y* 'socialist', whence *štira:ki:ya* 'socialism'; from the CLA intensive adjective *miṣda:q* 'much given to truth telling' (apparently revived in this sense for the purpose) *miṣda:qi:ya* 'credibility'. An interesting and very recent example is furnished by *jama:hi:r*, plural of *jumhu:r* 'large crowd', from which are formed the relational adjective *jama:hi:ri:y* 'pertaining to the masses', and from there the noun *jama:hi:ri:ya* 'government by the masses', which is the head noun in the official name of the current Libyan state. These terms were quite deliberately coined by the Libyan government in the 1970s to simultaneously underline the (in its view) extreme populism of the Libyan regime and to formally distinguish it from other regimes in the Arab world that use the terms *jumhu:ri:y* 'republican' and *jumhu:ri:ya*, 'republic', coined from the singular *jumhu:r* 'crowd, throng' in about 1830, when European political ideas first gained currency in Egypt.[7]

4.4 Number

Plural nouns in Arabic refer to countable entities that are more than two in number: as in the verb, the dual is a separate morphological category. The plurals of the various classes of nouns and adjectives described above are formed by two different procedures that, with a tiny number of exceptions, are mutually exclusive: suffixation (also the method for dual formation), and the familiar procedure of interdigitating consonant/vowel patterns on the root consonants of the singular noun. The latter method is the more common. The primary factor that determines whether a noun forms its plural by suffixation ("sound plural") or interdigitation ("broken plural") is the morphological pattern of its singular coupled with its meaning. In some cases, however, there are alternative "sound" and "broken" plurals, or several broken plurals; here, it is usage or syntactic function that decides. The particular broken plural pattern that a given singular takes is unpredictable for the class of "primitive" nouns[8] but is largely or totally predictable for the various classes of derivatives. The number of broken plural patterns in the dialects is fewer than in MSA, partly through the loss of some minor patterns and partly because of the coalescence of some as a result of phonological change. There is also a more predictable correspondence between singular and plural patterns.

There exists a separate category of collective noun that is used, in roots where it exists, to refer to uncountable nouns or to plural entities (usually humans, animals and other classes of living things such as fruit, vegetables, trees, etc.) as an undifferentiated group. For example, *naxl* is a collective used when reference is made to 'palm trees' in general, whereas *naxi:l* means, '(particular) palm trees'.

From the collective is derived the instance noun by the suffixation of the feminine ending -*a*, for example, *naxla* 'a (single) palm tree'. The same procedure is applicable to (some) uncountable nouns: *laḥm* 'meat' is used to refer to meat as an undifferentiated, mass substance, from which an instance noun *laḥma* 'a piece of meat' is formed. The broken plural form, *luḥu:m*, refers to 'meats' considered as separate categories—beef, lamb, etc. In the case of human groups, the singular is usually formed from the collective via the relative adjective: so *'arab* 'Arabs', *'arabi:y* 'an Arab'. In other cases—where the singular is morphologically a derivative adjective/noun, and in the case of some relational adjectives—the collective, somewhat confusingly, is formed by adding the suffix -*a*: thus *ma:rr* 'pedestrian' (active participle from *MRR* 'pass') gives *ma:rra* 'pedestrians', *la: ʔadri:* 'skeptic' (lit. 'I do not know') gives *la: ʔadri:ya* 'skeptics'.

4.4.1 Pluralization by suffixation

Suffixation as a means of pluralizing nouns is limited to certain categories, although a few anomalous plurals suggest that it may in the ancient past have been more widely employed. The MSA suffixes are (m.) -*u:na* (nom.) and -*i:na* (acc./gen.) and (f.) -*a:tun* (nom.) and -*a:tin* (acc./gen.). The masculine plural suffixes are added directly to the noun stem, and the feminine ones after removing the -*a(t)* feminine singular suffix. The dual suffixes are, for both genders and nouns of all types, -*a:ni* (nom.) and -*ayni* (acc./gen.). These are added to the singular noun stem, which in the case of the feminine includes the -*a(t)* ending, for example, from *mudi:r* 'manager', *mudi:ra* 'manageress'.

Table 4.1   CLA/MSA "Sound" Plural and Dual Suffixes

		M.	F.
Dual	nom.	*mudi:ra:ni*	*mudi:rata:ni*
	acc./gen.	*mudi:rayni*	*mudi:ratayni*
Pl.	nom.	*mudi:ru:na*	*mudi:ra:tun*
	acc./gen.	*mudi:ri:na*	*mudi:ra:tin*

4.4.1.1 Masculine suffixes

The masculine suffixes are used to pluralize rational male beings whose singulars fall into the following morphological classes:

(a) Active and passive participles

In addition to all participial forms that refer to rational male beings, most other deverbal adjective/noun patterns listed in 4.3.2 (f), but excluding *ʔaCCaC* adjectives denoting colors, may also form their plurals through suffixation. The

following are a few examples (in bold) of plural participles and adjectives chosen randomly from MSA materials to hand:

ʔalmuwazziʕ**uːna** lwaḥiːd**uːna** (*muCaCCiC/CaCiːC*)
the-distributor-pl the-sole-pl
'The sole distributors'

maː dumnaː   baʕiːd**iːna** kulla lbuʕd   ʕani nnuːr . . . (*CaCiːC*)
what remained-1pl far-pl   all   the-distance from the-light
'As long as we are so far from the light . . . '

ʔakkadati   lfuḥuːṣ   ʔannahum ḥaːmil**uːna** li lfiːruːs (*CaːCiC*)
proved-3-fsg the-tests that-they carrier-pl of the-virus
'The tests have proved that they are carriers of the virus'

ʔalmuraːqib**uːna** wa lmuḥallil**uːna** . . . (*muCaːCiC/muCaCCiC*)
the-observer-pl and the-analyst-pl . . .
'Observers and analysts . . . '

kaːnu muḥtajaz**iːna**   fiː sujuːn. . . . (*muCtaCaC*)
were-3pl being-held-pl in prisons . . .
'They were being held in prisons . . . '

yaqifuːna mašduːh**iːna** ʔamaːma miːzaːjiːyatihim . . . (*maCCuːC*)
3pl-stand amazed-pl before   temperament-their
'They stand amazed at their temperament . . . '

As noted earlier, and exemplified here, participles may function as substantives or as adjectives, attributively or predicatively. Many active participles derived from Pattern I verbs, unlike those derived from the augmented patterns that make their plurals by suffixation only, have both sound and broken plurals whose use depends on meaning and syntactic function. Used substantively to describe a regular occupation or quality seen as permanent, they normally form a broken plural. But when predicating a particular activity or contingent state of a subject—that is, when they are used adjectivally or verbally or elliptically as nouns ('writers (of something)')—they are pluralized by suffixation. Thus from *kaːtib* 'writing, writer' is formed *kaːtibuːna*, which is used in a verbal sense, for example, (with loss of the *-na* ending in the construct noun phrase):

naḥnu kaːtibuː haːðihi rrisaːla
we writing-pl this the-letter
'We are the writers of/have written this letter'

The broken plural of *kaːtib—kuttaːb*—is used where the meaning is rather the profession of 'writer':

haːʔulaːʔi **kuttaːb** mašhuːruːna
these writers famous-pl
'These (men) are famous writers'

The same general principle applies to deverbal adjective/noun patterns derived from Pattern I verbs that have s-stem theme vowel /u/ or /i/, denoting states and qualities: in modern writing, the more "substantival" in sense and syntactic behavior such words are, the less likely they are to pluralize by suffixation. Thus *qariːb* 'near, close' < *qaruba* 'be near' may form a suffixed plural *qariːbuːna*, but this is normally used as a "contingent" modifying or predicative adjective (meaning 'near' in a locative sense) such as *baʕiːdiːna* in the second of the examples cited above. The broken plural, *ʔaqribaːʔ* can be used only as a noun to mean 'near relatives'. Some adjectives are so like nouns (in not designating a temporary, contingent state) that they do not have a suffixed (masculine) plural at all, for example *kabiːr* 'big, large, great' < *kabura* 'be/become old, great, big'. As a broken plural, *kibaːr*, the plural of *kabiːr*, is encountered in MSA mainly as a substantive in noun phrases such as *kibaːru lmasʔuːliːna* 'high officials', *kibaːru rrijaːl* 'great men', *ʔalkibaːr* 'adult (males)'. But there are one or two exceptions—a few *CaCiːC* adjectives tend to employ the suffix plural even when used substantively. The most commonly encountered members of this group are *qaliːl* 'few' and *kaθiːr* 'many', which, when pluralized, are by themselves a way of referring to an anonymous 'few/many people'. For example:

lam yarahu   ʔillaː lqaliːl**uːna** wa   huwa yarḥalu
not 3msg-see-him except the-few-pl and he   3msg-leave
'Only a few people saw him leaving'

xuyyila   ʔila: kaθiːr**iːna** ʔan . . .
imagined-PASS-3msg to   many-pl that . . .
'Many people imagined that . . . '

Even for these words, in older writing, we encounter broken plurals when they are used as substantives, for example, Jurji Zaydān, writing at the turn of the century:[9]

wa min ʔulaːʔika lqalaːʔil   qaːsim ʔamiːn
and of those   the-few-pl Qasim Amin
'Qasim Amin was one of those few men . . . '

(b) Nouns of profession/occupation: *CaCCa:C*

CaCCa:C, it will be remembered, was originally a species of intensive adjective. In MSA, this pattern invariably pluralizes by suffixation: *qaṣṣa:bu:na* 'butchers', *jarra:ḥu:na* 'surgeons', *najja:ru:na* 'carpenters', etc.

(c) Relational adjectives in *-i:y*

The relational adjective, whether used as a substantive or adjectivally as the modifier or predicate of a masculine plural always forms a sound plural, for example, *ʔalinfiṣa:li:yu:na lbalti:qi:yu:na* 'the Baltic secessionists', *kutta:b ʕaṣri:yu:na* 'contemporary writers', *ʔarraha:ʔinu lʔamri:ki:yu:na*, 'the American hostages'.

(d) Comparatives/superlatives of the form *ʔaCCaC*

Suffixed plurals of *ʔaCCaC*, such as *ʔakbaru:na* 'greatest (ones)', may be formed. They are used only as nouns or attributive epithets in definite noun phrases and are favored, perhaps, by writers known for their "Classical" style, as well as occurring in reference to religiohistorical groups, for example, from *ʔaqall* 'less, fewer' *ʔalʔaqallu:na mina lmustašriqi:na* (Taha Hussein) 'the minority among the orientalists', *qurayšuni lʔakramu:na* 'the most noble tribe of Quraysh'. In less consciously "literary" MSA writing, there seems to be a preference for broken plurals of the pattern *ʔaCa:CiC* to be used instead if one exists, for example *ʔalʔaka:rim* 'the noblest ones', *ʔalʔaka:bir* 'the greatest/senior ones', etc.

In addition to those in categories (a)–(d), a small number of "primitive" nouns, only some of which denote rational male beings, may be pluralized by suffixation, for example, *sana* 'year' *sinu:na*, *ʔibn* 'son' *banu:na* (in the general sense of 'boys'), *ʔarḍ* 'land, earth' *ʔarḍu:na*. In MSA all of these have more frequently occurring broken plural forms that have become the norm in most contexts: *sanawa:t* 'years' (*fi: ssanawa:ti lʔaxi:ra* 'in recent years', not *fi: ssini:na . . .*), *ʔabna:ʔ* 'sons' (*ʔabna:ʔ miṣr* 'the sons of Egypt' not *banu: . . .*), *ʔara:ḍin* 'lands' (*ʔalʔara:ḍi: lmuḥtalla* 'the Occupied Territories' not *ʔalʔarḍu:na . . .*). The suffixed plural forms of these words now tend to occur only in fixed literary phrases or religious formulae, for example, in the title of a play by the Lebanese writer Mikha'il Nuʕayma *ʔalʔa:ba:ʔu wa lbanu:na* (*Fathers and Sons*) and on an ornately calligraphed signboard at the entrance to a Bahraini primary school:

madrasatu   ʔabi: bakrini ṣṣiddi:qi   libtida:ʔi:ya li lbani:na
school (of) Abū Bakr   Al-Ṣiddīq the-primary for the-boys
'Abū Bakr Al-Ṣiddīq Boys' Primary School'

to provide an echoic contrast with the girls' school (*li lbana:t*, from the same root) down the road. The only other context in which this particular plural is normally encountered is in traditional tribal names (when, as the first element in a construct noun phrase, it loses its *-na*), for example, *banu: ha:šim* 'the Hashimites'

('sons of Hashim'). Compare also the suffixed "religious" plural in the Koranic *rabbu lʕa:lami:na* 'Lord of the Worlds' (and in dozens more Koranic citations) with the broken plural *ʕawa:lim* in "secular" phrases such as *ʕawa:lim majhu:la* 'unknown worlds'.

4.4.1.2 Feminine suffixes

The feminine suffix *-a:t* is the counterpart of the masculine *-u:na*, *-i:na* and, in MSA, is used to pluralize adjectives and nouns that refer to groups of rational females. (Mixed groups are referred to by the masculine form.) The morphological classes affected are the same as in the masculine case, viz. participles: *ka:tiba* 'a female writer' → *ka:tiba:t*; deverbal adjectives: *kabi:ra* 'a big/old/great woman' → *kabi:ra:t*; nouns denoting occupations: *xayya:ṭa* 'seamstress' → *xayya:ṭa:t*; relational adjectives: *filasṭi:ni:ya* 'Palestinian woman' → *filasṭi:ni:ya:t*; comparatives/superlatives: *kubra:* 'greatest one' → *kubraya:t*. In CLA, and to some extent in self-consciously "literary" styles of MSA, these suffixed plurals may be applied to certain nonrational feminine plurals, for example, *sanawa:t maʕdu:da:t* 'a limited term of years', *ʔallaya:li: dda:misa:t* 'the pitch-black nights'.[10]

In addition to these classes, the *-a:t* suffix is used to pluralize some types of noun, which may be grammatically feminine or masculine in the singular and which refer to concrete or abstract nonrational entities. The most productive classes are participles and verbal nouns. Examples are:

Grammatically feminine (-a)

a.part.:	I	*ša:ḥina*	'truck'	→ *ša:ḥina:t*
	II	*mudammira*	'destroyer'	→ *mudammira:t*
	IV	*muškila*	'problem'	→ *muškila:t*
p.part.:	II	*muʔassasa*	'institution'	→ *muʔassasa:t*
abstract n.:		*ʔimka:ni:ya*	'potentiality'	→ *ʔimka:ni:ya:t*
verbal n.:	I	*xidma*	'service'	→ *xidma:t*
	III	*muḥa:wala*	'attempt'	→ *muḥa:wala:t*
instance n.:		*ḍarba*	'blow'	→ *ḍarba:t*
borrowings:		*ṭa:wila*	'table'	→ *ṭa:wila:t*

Grammatically masculine

a.part.:	II	*muhaddiʔ*	'tranquilizer'	→ *muhaddiʔa:t*
p.part.:	I	*mantu:j*	'product'	→ *mantu:ja:t*
	IV	*mulḥaq*	'addendum'	→ *mulḥaqa:t*
(n. of place)	X	*mustašfa:*	'hospital'	→ *mustašfaya:t*
verbal n.:	II	*taʕri:f*	'definition'	→ *taʕri:fa:t*
	V	*taqallub*	'fluctuation'	→ *taqalluba:t*
	VII	*nʕika:s*	'reflection'	→ *nʕika:sa:t*

concrete n.:	*ḥamma:m*	'bathroom'	→ *ḥamma:ma:t*
diminutives:	*kutayyib*	'booklet'	→ *kutayyiba:t*
borrowings:	*tilifu:n*	'telephone'	→ *tilifu:na:t*

4.4.2 Broken plurals

The "broken" plural, as the name suggests, involves changing the shape of the singular through various morphological processes such as long vowel insertion, consonant gemination, semivowel insertion, and the affixation of consonants additional to those of the root. It is another variation on the theme that, as we saw in chapter 3, lies at the heart of Arabic morphology: the superimposition of templatic consonant–vowel patterns onto triradical roots. Wright[11] lists twenty-nine broken plural patterns as in common use in CLA, most of which are still in use, but perhaps only about half of which are commonly encountered in MSA. Apart from those types described in 4.4.1, all other singular nouns, masculine or feminine, form their plurals in this way, and even among those that do have suffixed plural, as we have seen, there may be "broken" alternatives that have subtly different senses.

Primitive nouns—those that are not verbal derivatives—such as *'ayn* 'eye', *rajul* 'man', *ʔimraʔa* 'woman', *ma:ʔ* 'water', *na:r* 'fire', *baḥr* 'sea', *ba:b* 'door', etc., have broken plural patterns that cannot be predicted from the shape of their singulars. Indeed, many of them have more than one plural, although the alternatives are not usually interchangeable in particular senses or collocations. For example, there are two broken plurals of *ba:b* 'door', *bi:ba:n* (pattern *CiCCa:n*) and *ʔabwa:b* (*ʔaCCa:C*), either of which may be used to pluralize the word when used in this original sense, but in its extended sense of 'chapter, section (of a book)' only the second of these is used; *bayt*, meaning 'tent, house' has the plural *buyu:t* (*CuCu:C*), but it also means (again, via a semantic extension) 'verse of poetry', in which sense its plural is *ʔabya:t* (*ʔaCCa:C*).

Nouns derived from verbs or other nouns are much more predictable. Certain patterns always, or almost always, pluralize in fixed ways; for example, nouns and adjectives of the pattern *CaCi:C* often make their plural on the patterns *CuCaCa:ʔ* and *ʔaCCiCa:ʔ*, for example, *ʔami:n* 'trustee, secretary' *ʔumana:ʔ*, *safi:r* 'ambassador' *sufara:ʔ*, *wazi:r* 'minister' *wuzara:ʔ*, *qari:b* 'relative', *ʔaqriba:ʔ*, *ġani:y* 'rich (man)' *ʔaġniya:ʔ*; nouns of place derived from Pattern I verbs, which in the singular have the shapes *maCCaC*, *maCCaCa*, or *maCCiC* use the pattern *maCa:CiC*, so *maktab* 'office, desk' gives *maka:tib*, *madrasa* 'school' *mada:ris*, and *majlis* 'sitting place, council' *maja:lis*. As in the case of the primitive nouns, a multiplicity of plurals indicates polysemy. Thus the active participle *'a:mil* < *'ML* 'work, do' (pattern *Ca:CiC*) 'a person/thing that works' has three different but

related potential meanings, each of which has its own plural. When referring to human beings, *'a:mil* may mean 'one who (at a particular point in time) is working/works', and in this sense pluralizes by suffixation: *'a:milu:na*, cf. *ka:tib* 'one who is (temporarily) writing/has written' exemplified earlier. But if it means 'one who habitually/gainfully works' (= 'laborer, worker'), its plural is *'umma:l*, on the same pattern (*CuCCa:C*) used to pluralize *ka:tib* in the sense of '(professional) writer'. In another semantic extension of its meaning, *'a:mil* can also be used to refer to nonrational 'things which work/operate' (= 'factor'), and its plural when used in this sense is *'awa:mil* (as in, e.g., *ʔal'awa:mil aliqtiṣa:di:ya* 'the economic factors'), formed on the pattern *Cawa:CiC*, which is regularly used to pluralize any adjective/noun on the agentive pattern *Ca:CiC* when its referent is inanimate, for example, *ja:mi'* 'mosque' ('thing that gathers (people) together') *jawa:mi'*, *ma:ni'* 'objection, obstruction' ('thing that hinders') *mawa:ni'*. Sometimes ambiguities may arise accidentally: for example, *maktaba*, a relatively recently coined noun of place meaning 'library, bookshop' should (and theoretically does) have the plural *maka:tib* according to regular rule; but this pattern was already in use in the same root to pluralize the older word *maktab* 'place where writing is done/taught' (= 'desk, office'),[12] and so a suffixed plural, *maktaba:t* was also coined and in MSA is normally used to distinguish it from the plural of *maktab*.

The core of the system for pluralizing assumes a singular of three radical consonants, extended in certain classes of derivatives by one or two other morphologically significant elements—an initial *m-*, as in the nouns of place, time, and instrument; a long vowel, as in participles such as *Ca:CiC*; or both, as in the passive participle pattern *maCCu:C* and the noun of instrument pattern *miCCa:C*. All of these, and others, are accommodated by the *CaCa:CiC* or *CaCa:Ci:C* patterns in which the first or second consonant is not part of the root: *maka:tib* 'offices', *'awa:mil* 'factors', *maka:ti:b* 'letters', *mafa:ti:ḥ* 'keys', and (pl. of Pattern II verbal noun) *taka:li:f* 'costs'. "Genuine" quadriliterals and quinquiliterals are also accommodated by these patterns (and *CaCa:CiCa*), quinquiliterals losing their final root consonant; thus *tarjama* 'translation' gives *tara:jim*, *di:na:r* 'dinar' *dana:ni:r*, *'ankabu:t* 'spider' *'ana:kib*, *'abqari:y* 'genius' *'aba:qira*. There are a few words in Arabic, however, most of them among the most basic in the language, that have only two radical consonants in their singular form. Among these are *yad* 'hand' < *YD*, *dam* < *DM* 'blood', *(ʔ)ism* < *SM* 'name', *ʔab* < *ʔB* 'father', *ʔax* < *ʔX* 'brother', and a few others. These anomalies have been ironed out morphologically by supplying them with an extra consonant—a semivowel—thus allowing assimilation into the regular system for pluralization, for example, *ʔaydi:* 'hands' (pattern *ʔaCCuC*, notional root *YDY*), *dima:ʔ* 'blood' (pat-

tern *CiCa:C*, notional root *DMY*), etc. Foreign borrowings that fortuitously fit into the Arabic triradical system usually have broken plurals: thus *filim* 'film' has the plural *?afla:m*, *siga:ra* 'cigar, cigarette' (< It. *sigaro*) *saga:yir*. Contrast these with *tilivizyu:n* 'television', for which the only viable plural was by suffixation— *tilivizyu:na:t*. Compound nouns that were originally calqued on foreign words may pluralize regularly, as *ra?s ma:l* 'capital' (= lit. 'head (of) wealth'), which gives *ru?u:s ?amwa:l* by using the normal broken plurals of the constituent items, or be treated as if they were a single noun (presumed *rasma:l*) to give *rasa:mi:l*.

The basic plural system described here for MSA applies also, *mutatis mutandis*, in the dialects. The main formal differences are that, with the general loss of the case system, the *-u:na/-i:na* distinctions in the masculine sound plural and the *-a:ni/-ayni* ones in the dual have been lost. A generalization of the oblique case forms *-i:n* for all masculines, *-e:n* for all duals (a category that has survived in the noun but not the verb[13]), and *-a:t* for all feminine sound plurals is now found everywhere. The number of distinct broken plural patterns is reduced for the same reason as the number of distinct singular patterns: the partial or complete collapse of unstressed short vowel distinctions in open syllable. Some patterns have expanded in scope: for example, *CaCa:Ci:C* in the Gulf dialects is applied to the *CaCCa:C* pattern that denotes professions and habitual characteristics, for example, *xabba:z* 'baker' → *xaba:bi:z*, *farra:š* 'cleaner' → *fara:ri:š*, *šaǧǧa:l* 'hard-working' → *šaǧa:ǧi:l*, etc., rather than *xabba:zi:n*, etc., via suffixation. Foreign borrowings have entered the dialects in profusion, most of them to denote culturally novel items of everyday life. As in MSA, they have been assimilated to the root-pattern system where feasible; otherwise, they pluralize by suffixed *-a:t*. For example:

			NOTIONAL "ROOT"
Cairo:			
?o:ḍa (Turk.)	'room'	→ *?iwaḍ*	*?WD*
firwid (Eng.)	'forward (soccer)'	→ *farawda*	*FRWD*
tanda (It.)	'awning'	→ *tinad*	*TND*
but *rikurdar* (Eng.)	'tape recorder'	→ *rikurdara:t*	
Bahrain:			
bijli (Urdu)	'lamp'	→ *baja:li*	*BJL*
dari:ša (Pers.)	'window'	→ *dara:yiš*	*DRŠ*
bi:b (Eng.)	'pipe'	→ *bya:ba*	*BYB*
but *enjin* (Eng.)	'engine'	→ *enjina:t*	
Morocco:			
karta (Fr.)	'card'	→ *kwariṭ*	*KRṬ*
blaṣa (Sp.)	'town square'	→ *blayiṣ*	*BLṢ*

garsu:n (Fr.)	'waiter'	→ *grasn*	*GRSN*
but *bilanti* (Eng.)	'penalty (soccer)'	→ *bilantiya:t*	

Formally, singular/plural/collective distinctions similar to those in MSA are also preserved in some of the more conservative of the dialects, although functionally the collective seems to be less and less used, with the plural replacing it in "collective" contexts. In the Gulf, for example, *naxla* 'a palm tree' is distinguished from *naxal* 'palm trees' (considered as a whole or as a species) and *naxi:l* 'palm trees' (particular examples) exactly as in MSA. However, in the actual conversational example:

ma: nista:nis   ?illa   taḥt   innaxi:l   wiyya zzara:ri:ʕ
not 1pl-be-happy except under the-palms with the-farmers
'We're only happy when we're under the palm trees with the farmers!'

the plural *naxi:l* is used instead of the expected collective *naxal*. There is also a trend in the dialects toward the development of analytic, word-based singular/plural distinctions. As well as suffixing *-a* (or *-a:ya*) to the collective to form the singular, which is still a productive process, all dialects also use a range of free-standing lexemes, which differ from one dialect to another, meaning 'one', 'a piece' or 'a single example'. So in Cairo, for example, *ḥitta* 'piece' is used as a singulative for certain uncountable nouns, such as *ḥittit laḥm* 'a chunk of meat', *ḥittit ḥadi:d* 'a piece of iron', and *ḥabba* 'seed, grain' for certain vegetables and fruit, for example, *ḥabbit ?u:ṭa* 'a tomato'. In Bahrain one also hears *ḥabba* and its diminutive *ḥabbu:na* used in the same way, for example *ḥabbat jiḥḥ* 'a (single) watermelon', or *ḥabbu:nat raggi* 'a (single) watermelon'. For animals, *ra:s* 'head' is generally used, for example, *ra:s ǧanam* 'a (single) sheep'. The word *na:s* 'people' has developed a similar partitive function in some contexts. In reply to the question, "Did they all go?" the answer may be:

na:s   minhum ra:ḥu,   na:s   la
people of-them went-3pl, people not
'Some went, others didn't'

4.5 Case and definiteness

Both CLA and MSA have the same three-case system for nouns. The cases are customarily referred to in English as nominative, accusative, and genitive, although the syntactic circumstances in which they are used correspond only approximately to those in which the same terms are used to describe European lan-

guages that mark case, such as Latin and German. The largest group of nouns, the so-called triptotes, consists of those that distinguish the three cases by suffixing -*u* (nom.), -*a* (acc.), and -*i* (gen.) to singular and plural nouns of both genders when they are grammatically definite; when indefinite, an -*n* is added. Grammatically speaking, a noun/adjective is "definite" if it has the definite article (*ʔa*)*l*- 'the' prefixed to it (e.g., *ʔal-kitaːb* 'the book'); if it is the first of two nouns in a "construct" relationship in which the second noun is definite (see 6.1.3), for example *kitaːb* in *kitaːbu l-walad* 'the boy's book'; if it has a possessive enclitic suffixed to it, for example, *ʔasġar* 'smaller/smallest' in *ʔasġaru-hum* 'the smallest of them'; or if it is a proper name, regardless of whether it has the definite article prefixed (e.g., *lubnaːn* 'Lebanon', *ʔas-suːdaːn* 'Sudan'). Otherwise, it is indefinite.

A second, smaller group of nouns, the so-called diptotes, has, when indefinite, no final -*n* and differs from the triptotes in having -*u* in the nominative case and no distinction between accusative and genitive (both -*a*). When definite, however, this group behaves exactly like the triptotes, the largest group, in distinguishing the three cases.

The case markings of indefinite duals and masculine and feminine suffixed plurals are given in table 4.1. When they are definite, the only differences from the forms given are in the feminine plural, which like any definite noun loses its nunation, and in the loss of the final -*na* or -*ni* when any of the dual or masculine plural forms occurs as the first element of a "construct" noun phrase.

The suffixation of case endings causes certain phonological changes to occur in nouns the final consonant of whose stem is a semivowel, because, when case endings are suffixed, the semivowel occurs intervocalically, for example, -*ayu*, -*ayun*, etc. Such forms are shortened in accordance with the rules given in 3.5.1.2. Also, as noted earlier in 2.1.5, the conventions for spoken MSA performance require that the -*n* that marks indefiniteness be omitted, along with its short vowel, when it occurs in prepause position, although the extent to which this convention is actually observed in practice depends on the content (and hence style) of discourse—religious versus secular—and, in secular contexts, it seems, on the grammatical function of the case ending in the sentence.

This description treats case endings at the phonological level, as they would occur in the spoken performance of MSA; in writing, short vowels are of course not normally marked, nor does the indefinite marker -*n* ever feature as a full graphemic /*n*/, written as if it were a part of the word to which it is attached. Only in the indefinite accusative of triptote nouns not marked as feminine, phonologically -/*an*/, is the existence of short vowel case marking acknowledged graphologically: a final *ʔalif* is suffixed and has always been so from the

earliest written Arabic. Whatever the historical origin of this anomalous suffixing of *ʔalif*, it is interesting to note that during the recent period of rapid syntactic and lexical evolution of written Arabic, its use other than as a marker of the direct object of a verb, viz. as an adverbial complement of the main verb signifying the manner, time, purpose, and other circumstances appertaining to the event or state the verb describes, has grown. The reason for this, and for many other changes in the syntax and phraseology of modern Arabic, has to do with the dominant influence of European languages in the media and international communications generally. In a language whose possibilities for agglutination are severely limited, this overtly marked case suffix is one of a small number of handy morphological/graphological resources (another is the prefixed preposition *bi-*) that have been heavily exploited in the coining of adverbial complements functionally equivalent to English 'ly', '-wise', and prepositional phrases. So, for example, one reads of production as growing 'relatively' larger or smaller 'in absolute terms' (*nisbiːy-an, ʔiṭlaːq-an*); Husni Mubarak can send his greetings in a cable of congratulation 'as president, on behalf of his people and government' (*raʔiːs-an, wa šaʕb-an wa ḥukuːmat-an*); Saddam Hussein could appeal to the west in 1980 'on humanitarian grounds' (*ʔinsaːniːy-an*) for help in developing his 'peaceful nuclear program'. This use of -*an* adverbial complements is most noticeable in media Arabic that deals with political, economic, and social issues—all fields in which writing in European languages is internationally dominant—and, as a result, Arabic sentence structure, if not calqued directly on European source texts such as press agency reports, nonetheless has come to bear a strong resemblance to what Blau,[14] after Blanc, dubs Standard Average European syntax (or what many indigenous prescriptivists would call "bad Arabic").

As a means of syntactic disambiguation in modern written Arabic, case plays almost no role (inevitably so, because in most cases it is carried by short vowel distinctions that are unmarked), and, despite the importance that the indigenous tradition of grammatical description and language pedagogy attaches to it, it is clear, when one examines ancient textual material, that the functional load of the case endings was no higher in the Classical period than it is now. As was argued in chapter 1, by the time of the composition or revelation of the oldest written Arabic materials we have, the full set of case and mood endings was probably no longer in everyday spoken use (as reflected by the script conventions) and was reserved for "elevated discourse" of one kind or another as a species of rhetorical or poetic flourish. Such a theory seems to fit the facts of the modern dialectal survival of case marking in the dialects of Arabia, the only place where its vestiges survive (see also 7.1.1).

As genuine dialectal survivals of an earlier phase in the language's history, case inflection in definite nouns has totally disappeared from all the modern dialects, and in indefinite noun phrases the suffixing of -*n*, the marker of indefiniteness ("nunation") has all but disappeared, surviving, as we noted in 1.2.3, only in the dialects of Arabia and its periphery as an undifferentiated suffix -/*in*/ in one or two specific types of syntactic structure—most notably noun–adjective phrases in association with certain adjectives (*ze:n* 'good' is the most common in phrases like *bint-in ze:na* 'a good girl' (= 'a nice bride'), *fardat-in ze:na* 'good dates'). Its contexts of use are usually poetic, socially ritualistic, or otherwise formulaic and are typically associated with the old, the uneducated, and others whom the new literate culture has left behind. Paradoxically, nunation also occurs in "educated" spoken Arabic but in a different guise, from a different source, and for a different purpose—as a saliently standard element, like /*q*/, that has the effect of "style raising" the plain dialect, a tactic that even slightly educated speakers habitually engage in when they deem that the context or topic demands a more "correct" or serious tone. At its simplest and most everyday level, this involves the use of a sprinkling of "prepackaged" MSA adverbial accusatives such as *taqri:ban* 'approximately', *maθalan* 'for example', and *fawran* 'immediately' that can be slotted into dialectal sentences without forcing the speaker to make a real switch toward standard syntax. At a more elevated stylistic level, speakers may of course attempt to extemporize correct MSA case endings in certain phrase types.

4.6 Gender

The various categories of feminine nouns in MSA (marked morphologically as such by the -*a(t)* suffix; construed as feminine because of the sex of the referent; or construed as feminine by convention) were noted in 4.3.1. There are also a few words that may be construed as either masculine or feminine and a few patterns, notably the intensive adjective pattern *CaCCa:Ca* and certain plural patterns that end in -*a*, that, although they have the feminine suffix, refer to masculine rational beings and require masculine concord.

In the dialects, the same system of gender assignment applies but with a few individual differences in the category of nouns that are feminine by convention. A particularly interesting case is that of certain human mass nouns such as *na:s* 'people' that can be used either collectively to refer to a group in general or specifically to refer to a group as a collection of particular individuals. In MSA, *na:s* is usually construed as a masculine plural in respect of verbal, adjectival, and pronominal concord. But in some modern dialects, agreement depends on the generality or "individuatedness" of the people being referred to. Used as a collective of general reference, *na:s* is given (like nonanimate plurals) feminine singular agreement in any verb or adjectival predicate (although pronouns that refer to it must be plural!); but when some form of deixis, enumeration, or other means of specification of identity is involved, for example 'those people' (sc. 'the ones over there'/'the ones I was referring to', etc.), the agreement required is normally masculine plural (see 6.1).

Notes

1. Wright 1896–98, 1:110–12.
2. Ali 1987, 142–48.
3. *tiCCa:C* in a few cases, for example, *tilqa:?*.
4. See Wright 1896–98, 1:115 for CLA attestations. Holes 1983b, 28–29 notes further examples from Bahraini Arabic. Similar archaic forms have been noted for southern Oman (Dhofar), Hadramawt, San'aa, many areas of Saudi Arabia, Kuwait, and the eastern Arabian dialects generally.
5. Masliyah (1997) gives four main functions for diminutives in spoken Iraqi Arabic—endearment, contempt, enhancement (e.g., *da:hiya* 'misfortune', *duwayha* 'great misfortune'), and what he terms "mercy" (e.g., *miski:n* 'poor', *musaykin* 'a little poor' (sic), although the example here seems caritative, perhaps better translated as 'poor little dear').
6. The translation of 'suicide' in what in the West is invariably termed 'suicide attack' or 'suicide bomber' depends in Arabic on the ideological stance of the translator/publisher. Writers with a pro-Islamist agenda, when reporting the activities of Hamas or Islamic Jihad in Israel, translate such phrases with a different adjective from a Pattern X verbal noun: *stiša:di:y* 'martyrdom-REL' as in *hamla istiša:di:ya* 'martyrdom attack'. Spokespersons from the Palestinian National Authority, however, steer a middle course and opt for an adjective from a Pattern VII, *nfija:ri:y* 'explosion-REL', calling such activities *'amali:ya:t infija:ri:ya* 'explosives operations' and thus distancing themselves from both the western and Islamist viewpoints.
7. Ayalon 1987, 102–3.
8. Certain common plural patterns, known as "plurals of paucity" were originally used only for collections of (some) nouns if they numbered between three and ten. In MSA and the dialects, this usage survives vestigially in its original sense in phrases involving nouns that occur in small sets or that tend to be referred to in small numbers, such as "days" and "months."
9. Zaydān (n.d.), 105.
10. Both examples are taken from the essay *Al-Nūr wa l-Dayjūr* ("Light and Dark") by the twentieth-century Christian Lebanese writer Mikha'il Nu'ayma.
11. Wright 1896–98, 1: 199–224.
12. The further potential ambiguity of *maktab* as meaning either 'desk' or 'office' is resolved in MSA by the tendency to use the neologism *da:?ira* in the latter sense.

13. Blanc 1970 shows that what is superficially a single dual form in many dialects is in fact semantically (and formally) two different structures: a "true" dual with an invariant morpheme *–e:n, -ayn,* or *-i:n* (depending on the dialect) and a "pseudo" dual, the latter used for a small number of nouns (notably parts of the body in pairs) with presuffixal loss of the *-n.*

14. See Blau 1981b, 6off. and especially 116–41; Blanc 1957, 400.

5

Beyond Root and Pattern:
Pronouns and Deictics
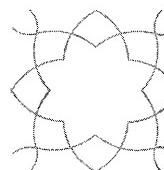
As we have seen, the structure of the vast majority of individual Arabic words can be reduced to one or another of a small set of morphological templates onto which a semantically abstract root meaning is imposed, the resulting combination thereby acquiring a specific, and to a considerable degree predictable, lexical value. But, as in many languages, some of the most fundamental functions of language—personal reference, deixis, and interrogation—are expressed through closed classes of elements outside this main derivational system. In this chapter we will review the structure and main functions of the systems of personal reference, demonstratives, and interrogatives in both MSA and representative dialects, especially insofar as these forms and functions differ.

5.1 Personal pronouns

5.1.1 Forms

In CLA/MSA, there are two sets of pronominal forms: a set of free morphemes that are written as separate words and that generally occur only in the position of grammatical subject (but may be used appositionally in other than subject position) and a set of bound pronominal clitics that can be suffixed to verbs, nouns, prepositions, and particles of various types and that may variously function as the grammatical direct object, indirect object, or possessor of the word to which they are suffixed. Both sets of

Table 5.1   CLA/MSA Pronouns

Free-standing forms:

Person/Gender	Sg.	Dual	Pl.
I	ʔana	—	naḥnu
2m.	ʔanta		ʔantum
		ʔantuma:	
2f.	ʔanti		ʔantunna
3m.	huwa		hum
		huma:	
3f	hiya		hunna

Bound forms:

Person/Gender	Sg.	Dual	Pl.
I	-ni:/-i:	—	-na:
2m.	-ka		-kum
		kuma:	
2f.	-ki		-kunna
3m.	-hu		-hum
		huma:	
3f.	-ha:		-hunna

Note: In the first person, *-ni:* is suffixed to verbs, *-i:* (in some phonological environments its allmorph *-iya*) to nouns and prepositions.

pronouns are marked for the same categories of number and gender as the finite verb, except in the case of the third person dual, in which *huma:* is used for both m. and f., in contrast to the verb forms, in which third person duals are distinguished for gender. The basic forms are given in table 5.1 above (there is some phonologically determined allomorphy that is left out of account).

In all of the modern dialects, the number of these distinctions has been reduced, although the details differ from one geographical area/social group to another. Everywhere, however, the dual–plural distinction has been lost as it has in most other areas of the morphology, with the plural replacing the dual. In the "urban" dialects (e.g., Cairo, Damascus, Jerusalem, etc.) the gender distinction in the second and third person plural has also disappeared, just as it has (as noted earlier) in the verb. However, in many of the Arabian Bedouin-type dialects and in some of those at the periphery (e.g., Oman, whether Bedouin-type or urbanite), which as a group have evolved more slowly, this gender distinction lives on. The effect of various local phonological changes has been to

Table 5.2   Damascene and Ha'il Pronouns

	DAMASCUS		AL-HA'IL	
	Free set	Bound set	Free set	Bound set
1sg.	ʔana	-ni/ -i	ana	-an/ -i
1pl.	niḥna	-na	ḥinna	-na
2msg.	ʔinte	-ak	int	-ak
2fsg.	ʔinti	-ek	inti	-ič
2mpl.			intam	-kam
	ʔintu	-kon		
2fpl.			intin	-kin
3msg.	huwwe	-o	huw	-uh
3fsg.	hiyye	-ha	hiy	-ah
3mpl.			ham	-ham
	hinne	-hon		
3fpl.			hin	-hin

change the loci of the phonological contrasts that carry the morphological distinctions. As an illustration, the pronominal forms for Damascus,[1] an "urban" dialect, and al-Ha'il, central Saudi Arabia,[2] a "Bedouin" one, are given in table 5.2 for comparison with each other and with the CLA/MSA set.

5.1.2 Uses

In "verbal" sentences that do not have an overt noun as subject, whether simple or conjoined (for "verbal" and "verbless" (or "nominal") sentences, see 7.1.1), the free-standing set of pronouns is not usually used in MSA, the verb itself being unambiguously inflected for person, gender, and number. Thus in:

simple sentence:

zur-tu   landan
visited-1sg London
'I visited London'

conjoined sentences, same subject:

zur-tu   landan   wa nazal-tu   fi: funduqin fa:xir
visited-1sg London and stayed-1sg in hotel   luxurious
'I visited London and stayed in a luxury hotel'

conjoined sentences, different subjects:

zur-tu   landan wa 'a:mal-u:-ni: mu'a:malatan ṭayyiba
visited-1sg London and treated-3mpl-me treatment   good
'I visited London and they treated me well'

the pronouns ʔana 'I' and hum 'they' do not appear, because the verb inflections, respectively -tu and -u: indicate the person, gender, and number of the subject (some previously specified third person subject in the case of -u: 'they'). In the dialects, however, there is a tendency, particularly in the first and second persons, for the free-standing pronouns to be used in verbal sentences where they would normally be absent in MSA. The origin of this may have been in the need to disambiguate forms inherited from CLA that were already polysemous and others that subsequently became so as the dialects developed, as a result of the loss of final short vowels. Thus in CLA the p-stem form taktubu means both 'you (m.) write' and 'she write(s)' as do, in the absence of an overt pronominal subject, its successor forms in the dialects, for example, Cairene tiktib, Damascene tiktob; and in most dialects the CLA distinction between the s-stem forms zurtu 'I visited' and zurta 'you (m.) visited' has collapsed through the loss of the final short vowels, for example, Cairene zurt, Baghdadi zirit for both forms. Alternatively, the free-standing pronouns may have begun to be used as a result of substrate influences and the inflectional differences in the verb may have been reduced because of redundancy. Either way, the increased role of the free-standing pronouns as overt subjects in the dialects is one example (of many) in which the dialects manifest an analytic tendency compared to the more synthetic phrase structure of CLA/MSA. If, however, in dialectal Arabic, it is clear from pragmatic or discoursal factors who the agent of an action is, then the free-standing pronoun does not need to be specified: at the beginning of a first person narrative in dialectal Arabic, for example, ana 'I' may be specified as the subject of the verb but becomes less frequent once this first person narrative frame has been established.

In sentences that do not contain a verb or contain a nonfinite (participial) one that is not specified for person, the free-standing pronouns are used in all varieties of Arabic. For example:

(a) As subject of a verbless equational sentence (x in x = y):
CLA/MSA   ʔana mudarris
dialects[3]   ana   mudarris
           I   teacher
           'I am a teacher'

CLA/MSA   hiya 'arabi:ya
dialects   hiya 'arabi:ya
           'She is an Arab'

(b) As subject of a nonfinite verb:
CLA/MSA   hal hum   sa:mi'-u:na   l-xabar
           Q they-mpl hear-A.PART-mpl the-news
Bahraini   hum sa:m'-i:n   il-xabar
Shi'i   they hear-A.PART-pl the-news
           'Have they heard the news?'

(c) As subject of a prepositional predicate:
CLA/MSA   ʔantum 'ala: ḥaqqin fi: ða:lika
           you-mpl on   right in that
Bahraini   intu:n   le:-kum ḥakk fi: da:k
Shi'i   you-pl to-you right in that
           'You are right in that'

(d) As complement of an equational sentence (y in x = y):
CLA/MSA   ʔal-masʔu:lu   ʔanta   ʔiðan
           the-responsible you-msg then
Bahraini   il-masʔu:l   inta   'a:d
Shi'i   the-responsible you-msg then
           'So you are responsible!'

Additionally, the free-standing pronouns are used:
(e) In apposition:
CLA/MSA   ʔantumu   l-baḥa:rina
           you-mpl   the-Bahrainis
           'You Bahrainis'
Bahraini   intu:n il-baḥarna   a'ruf-kum   ze:n!
Shi'i   you-pl the-Bahrainis 1sg-know-you well
           'I know you (Shi'i) Bahrainis only too well!'

(f) To topicalize (and thereby emphasize) a preceding noun:
CLA/MSA   (ʔinna) l-'umma:la   humu llaði:na štak-aw
           (TOP) the-workers   they who   complained-3mpl
Bahraini   il-'umma:l   hum illi štak-u
Shi'i   the-workers they who complained-3pl
           'It was the workers who complained'

(g) To contrastively stress bound pronominal enclitics:
Where contrastive emphasis on an enclitic pronoun in complement or object position is required, the free-standing pronouns are used to "echo" the enclitic and are contrastively stressed. For example:

CLA/MSA	*ḥal katab-u:*		*l-i:*	*ʔana*
	Q wrote-3mpl		to-me	I
Bahraini	*kitbu*		*l-iyyi*	*ana*
Shiʻi	wrote-3pl		to-me	I

'Did they write to *me*?!'

CLA/MSA	*la: yu-wjadu*		*ġayru-na*	*naḥnu*
	not 3msg-be found		other-than-us	we
Bahraini	*ma: miš*		*ġe:r-na*	*ḥna*
Shiʻi	not exist		other-than-us	we

'There is no one except *us*'

(h) In clauses of attendant circumstance:

Free-standing pronouns are used as the subject of a conjoined clause that describes the attendant circumstances surrounding the action in a main clause, whether their subjects are the same or different (see also 7.2.1). For example:

same subject:

CLA/MSA	*qa:m-a*	*wa*	*qa:l-a*	*wa*	*huwa*	*ya-bki:*
	rose-3msg	and	said-3msg	and	he	3msg-weep
Bahraini	*ga:m*	*u*	*ga:l*	*u*	*hu*	*ya-bči*
	rose-3msg	and	said-3msg	and	he	3msg-weep

'Weeping, he got up and said . . . '

different subjects:

CLA/MSA	*xaraj-u:*	*mina*	*l-ġurfati*	*wa*	*naḥnu*	*na-dxulu*
	exited-3mpl	from	the-room	and	we	1pl-enter
Bahraini	*ṭilʻ-u*	*min*	*il-ġurfa*	*u*	*ḥna*	*ndišš*
	exited-3pl	from	the-room	and	we	1pl-enter'

'They left the room as we were entering it'

(i) As a disambiguating device in "equational" sentences (x = y):

A distinctive use of the free-standing third person pronouns in all varieties of Arabic is as the so-called pronoun of separation, where their function is that of a copula marking off (or "separating") subject and predicate in nonverbal sentences in which the complement is a definite noun phrase. Without this pronoun there might be an ambiguity, at least on the written page, in sentences where both subject and complement are defined in the following cases: (i) where the subject and complement are both defined by the definite article (in which case the resulting concatenation could be interpreted as a noun phrase, for example,

ʔal-mudi:ru l-masʔu:lu ʻan ha:ða: 'the manager responsible for this' or as a sentence meaning 'the manager is the person responsible for this'); (ii) where the subject is a proper name (in which case the complement could be interpreted as in apposition—see example (e) above); or (iii) where the subject is a demonstrative and could be interpreted as part and parcel of the noun phrase rather than as the subject of an equational sentence, for example, *ha:ða l-ʔusta:ð* 'this teacher' or 'this is the teacher' (see 5.2.2 below)). But in cases where context makes it obvious that a sentence rather than an appositional or simple noun phrase is intended (very often in extempore speech), the pronoun is omitted even in these cases. Where it is used, it agrees in gender and number with the preceding noun. For example:

CLA/MSA	*muḥammad*	*huwa*	*l-ʔusta:ð*
Bahraini	*mḥammad*	*hu*	*l-usta:d*
Shiʻi	Muhammad	he	the-teacher

'Muhammad is the teacher'

CLA/MSA	*ʔal-mudarris-atu*	*hiya*	*l-masʔu:l-a*
Bahraini	*il-mudarris-a*	*hiya*	*l-masʔu:l-a*
Shiʻi	the-teacher-f	she	the-responsible-f

'The teacher(f) is responsible'

CLA/MSA	*ha:ða:*	*huwa*	*r-rajulu*
Bahraini	*ha:da*	*hu*	*r-rajja:l*
Shiʻi	this	he	the-man

'This is the man'

CLA/MSA	*ʔu:la:ʔika*	*humu*	*l-ʻumma:l*
Bahraini	*hade:la:k*	*hum*	*il-ʻumma:l*
Shiʻi	those	they	the-workers

'Those are the workers'

Uses (a)–(i) are found in all varieties of Arabic. There are a few others that are peculiar to particular dialects, the most important of which are exemplified here:

(j) As a topicalizer

A sentence-initial third person pronoun is used in Cairene (although not, it would seem, elsewhere) as a means of topicalizing and emphasizing a following demonstrative. The order of the demonstrative and pronoun in the unemphatic case is reversed. For example, compare the previous two sentences with:

| *huwwa* | *da* | *r-ra:gil* | *illi* | *ʔult* | *lak* | *ʻannuh* |
| he | this | the-man | which | said-1sg | to-you | of-him |

'*This* is the man I told you about'

humma   do:l   il-'umma:l   ik-kusali:ya
they   those the-workers the-lazy
'*Those* are the lazy workers'

(k) Sentence-initial questioning particle

In the same dialect, sentence-initial *huwwa* is used to introduce yes/no questions in which an element of doubt is being cast on the validity of a proposition. Compare the immediately preceding sentence with the following, which could be a rejoinder to it in a conversation:

huwwa   humma   do:l   ik-kusali:ya
it   they   those the-lazy
'Are *those* (really) the lazy ones?' (implied: "no")

That it is not restricted to sentences with third person subjects is shown by examples such as:

huwwa   ana   magnu:n
it   I   mad
'(Do you think) I'm mad?!' (implied: "I'm not going to do what you want," etc.)
where the force of *huwwa* is something like 'Is it really the case that . . . ?'

5.2 Demonstratives

5.2.1 Forms

The same sets of demonstrative forms function both adjectivally and pronominally in CLA/MSA on the one hand and the dialects on the other. The full set of CLA/MSA forms is in Table 5.3.

It will be apparent from the table that the prefix *ha:-* is a "proximal" marker denoting relative nearness in time or space to the speaker (which, as we shall see, is still used as an independent morpheme with its original presentative function 'see!'), whereas suffix *-(li)ka* is a "distal" marker denoting relative distance from the speaker. The deictic core of the form is an element marked for gender, number, and case: *-ða:-* marks masculine, and a variety of elements, involving either a vowel change (*-ði-*), a consonant change (*ta:*), or both together (*ti* < historical **ti:*[4]) mark the feminine. The common plural element, used only for referring to human beings, is *-ʔu:la:ʔi-*. The dual suffixes *-a:ni* and *-ayni* are inflected for case exactly as they are in the noun.

In the dialects, the phonological shape of the corresponding forms differs from one dialect to another, but the same distinctions of person and number are made as in the dialectal personal pronoun system. Despite local differences of

Table 5.3   CLA/MSA Demonstratives

			PROXIMAL	DISTAL
Sg.	M.		*ha:ða:*	*ða:lika* (or *ða:ka*)
	F.		*ha:ðihi* (or *ha:ði:*)	*tilka*
Pl.	Com.		*ha:ʔula:ʔi*	*ʔu:la:ʔika*
Dual	M.	Nom.	*ha:ða:ni*	*ða:nika*
		Acc./Gen.	*ha:ðayni*	*ðaynika*
	F.	Nom.	*ha:ta:ni*	*ta:nika*
		Acc./Gen.	*ha:tayni*	*taynika*

detail, the demonstrative pronouns in all varieties of dialectal Arabic are also, like the CLA/MSA forms, morphological composites in which the same three basic elements can be discerned:

- a "presentative" morpheme *ha(:)-* (optionally omitted in some dialects); this element turns up in other syntactic contexts, in which it has a function similar to the French *voici* and *voilà*;
- a deictic morpheme denoting the number and gender of the entity being referred to, most commonly: *da* (m.), *di* (f.) *do:l*, or *de:l* (pl.); and
- a morpheme *-k* denoting the distance in space or time from the speaker/text producer, or a zero morpheme denoting proximity.

A typical example of the system is found in the Baḥārna (Shi'i) dialect of Bahrain. In the singular, the proximal demonstrative 'this' is *ha:da* (m.) and *ha:di* (f.), and the distal demonstrative 'that' is *ha:da:k* (m.) and *ha:di:k* (f.), where *ha:-* is the "presentative" particle, *da(:)* and *di(:)* respectively denote masculine and feminine singular, and relative distance is denoted by a zero morpheme for proximity and a final *-k* for distance. In the plural, 'these' and 'those' are *ha:dela:* and *ha:dela:k*, in which the *dela:-* element signifies common plural and the function of the other elements is as in the singular. Most of the "urban" dialects have a similarly symmetrical set of forms (some with a *do:l(a:)-* plural element rather than *dela:-*). The Bedouin dialects as a general rule are more conservative and complex than this (i.e., they are more similar to CLA/MSA), in that they maintain gender distinctions in the plural. In all varieties of Arabic, however, the morphological composition of demonstratives—presentative particle + deictic element + distance marker—is in essence the same.

As noted above, the demonstratives are used both adjectivally and pronominally, in the latter case frequently (particularly in the dialects of the eastern Arab world) as virtual substitutes for the free-standing third person pronouns de-

scribed in 5.1 above. In CLA/MSA, the surface forms of the demonstratives are the same whatever their function, but in some dialects there are differences in their form and syntactic position that correspond to differences in syntactic or pragmatic role (see below).

5.2.2 Adjectival and pronominal uses

In all kinds of Arabic, whenever a demonstrative is used adjectivally, that is, to specify a particular member of a class, its head noun must be grammatically definite. A noun is grammatically definite either by having the definite article prefix, or a possessive pronominal suffix, or by being followed by a definite noun with which it is in a genitive (so-called "construct") relationship. For example:

head noun with the definite article:

CLA/MSA	*ha:ða: l-ʔakl*	
	this the-food	
Baghdadi	*hal-ʔakil* or (more formally) *ha:ða l-ʔakil*	
	this-the-food	this the-food
Cairene	*il-ʔakli-da* or (more formally) *ha:za l-ʔakl*	
	the-food-this	this the-food
	'This food'	

In CLA/MSA, the demonstrative normally precedes a simple definite noun phrase, as in the example above. But it may be postposed, viz. *ʔal-ʔakl ha:ða:*, for contextual reasons such as contrast ('this food (not that)'). The system in the dialects is different: some, such as the Baghdadi, Damascene, and Gulf dialects, characteristically prepose the demonstrative in the "unmarked" case; others, such as Cairene, characteristically postpose it. But, as illustrated in the examples above, all the dialects share a tendency to reduce it morphologically, either to a preposed invariant *ha-*, for example, Baghdadi *ha-l-ʔakil* 'this food' (with m. noun), *ha-s-sana* 'this year' (with f. noun), *ha-l-asa:tiða* 'these teachers' (with pl. noun), or to a postposed element that reflects the gender and number of the head noun, cf. the corresponding Cairene forms *il-ʔakli-da, is-sana:-di, il-asatza do:l*. Where postposing is the norm, reduced forms of this type following nouns ending in -CC or -CaCa may produce syllabically unacceptable strings, as in Cairene. In the case of -CC + C, an epenthetic vowel is inserted to break up the cluster, as in *il-ʔakli-da* above; in the case of -CaCa the final -a is lengthened, *is-sana:-di*, rather than *is-sana-di*. Postposed demonstratives attached to the noun phrase, like negative particles fore and aft of the verb phrase, are phonologically constituents of the noun phrase as far as syllable structure rules and stress assignment are concerned.

If the head noun is made definite by a personal pronoun suffix or is in a "construct" relationship with a following noun, there is a positional restriction on the demonstrative: any demonstrative that applies to the *whole* resulting complex noun phrase (square-bracketed in the glosses below) obligatorily comes after it:

CLA/MSA	*bayt-i:*	*ha:ða:*
Baghdadi	*be:t-i*	*ha:ða*
Cairene	*be:t-i*	*da*
	[house-my] this	
	'This house of mine'	
CLA/MSA	*qaṣru l-maliki*	*ha:ða:*
Baghdadi	*gaṣr il-malik*	*ha:ða*
Cairene	*ʔaṣr il-malik*	*da*
	[palace (of) the-king] this	
	'This palace of the king'	

The postposing of the demonstrative in these examples is a matter of syntactic rule, not of emphasis, as was the case with the definite article + simple noun structures of the general type *ʔal-ʔakl ha:ða:* 'this food' in CLA/MSA and in dialects that normally prepose. If the demonstrative is placed *before* noun phrases of the type exemplified above, an equational sentence results, there being no copula verb 'to be' in Arabic:[5]

CLA/MSA	*ha:ða:*	*bayti:*	*ha:ða:*	*qaṣru lmalik*
Baghdadi	*ha:ða*	*be:ti*	*ha:ða*	*gaṣr ilmalik*
Cairene	*da*	*be:ti*	*da*	*ʔaṣr ilmalik*
	[this]	[house-my]	[this]	[palace the-king]
	'This is my house'		'This is the king's palace'	

Where the complement of the equational sentence is a noun made definite by the definite article, a so-called pronoun of separation (5.1.2 (i)) that agrees in gender and number with its subject and complement has to be inserted between the demonstrative subject and this nominal complement in order to distinguish such sentences from noun phrases in which the demonstrative is functioning adjectivally. Thus in CLA/MSA:

> *ha:ða: l-bayt*
> this   the-house
> 'This house'

but

> *ha:ða: huwa l-bayt*
> this   it   the-house
> 'This is the house'

5.2.3 Presentative and emphatic uses

The original "presentative" function ('see!') of the *ha:-* element of the CLA demonstrative pronouns has been preserved in one form or another in virtually all the dialects and is used rather like the French *voici* and *voilà*. In MSA, presentative forms also exist and are used, for example, in written reports of speech, but they sound rather stilted if spoken, because the speech function of "presentation" is naturally more typical of spontaneous (i.e., dialectal) than scripted or prepared speech. The dialects differ as to the position in which the presentative element occurs:

(a) In Syria/Lebanon, *hayy* (probably derived from the full CLA proximal demonstrative *ha:ði(hi)* 'this (f.)') is used,[6] regardless of the number/gender of what follows:

hayy awla:d-i   ža:y-i:n
this children-my   come-A.PART-pl.
'Here come my children!'

(b) In Cairo, the postposed *ʔa:-* (< historical *ha:-* 'see'[7]) is used with an enclitic pronoun to express the same meaning:

wla:d-i   gay-i:n   ʔa:-hum
children-my come-A.PART.-pl. see-them
'Here come my children!'

In Cairene, the singular forms of the enclitic pronouns that are used in such structures are specific to them: *ʔa:-ho* 'here he is!' and *ʔa:-he* 'here she is!', *-ho* and *-he* occurring only in this type of usage.

(c) In (Shiʻi) Bahraini, preposed *ha:k-* (*ha:-* + *k*[8]) is used with an enclitic pronoun:

ha:-k-hum   ja:y-i:n   awla:d-i
see-distal-them   come-A.PART-pl. children-my
'Here come my children!'

whereas the Bahraini Sunnis use only the "distal" *ka-* plus enclitic:

ka-hum   ya:y-i:n   awla:d-i
distal-them come-A.PART-pl. children-my
'Here come my children!'

In some dialects (Syria and the Gulf), preposed demonstrative + free-standing pronoun constructions have an additional use as a means of emphasizing a

bound morpheme subject and its verbal predicate in rather the same way as we have seen postposed free-standing pronouns are used to emphasize a bound morpheme direct or indirect object. Thus, in Bahraini Arabic the neutral statement *aʻti:k il-flu:s* 'I'm giving you the money', where both the subject and indirect object are bound morphemes, may become:

(i) with emphasis on the indirect object (postposed pronoun):

aʻti:k   il-flu:s   inta
1sg-give-you the-money you-msg
'I'm giving *you* the money' (not him)

(ii) with emphasis on the subject (preposed demonstrative + pronoun):

ha:ðana   a-ʻti:-k   il-flu:s
this-I   1sg.-give-you the-money
'I'm *giving* you the money' (although you seem to think I want something in return)

(iii) with emphasis on both:

ha:ðana   a-ʻti:k   il-flu:s   inta
this-I   1sg-give-you the-money   you
'I'm *giving you* the money' (not him/her/them, and I don't want anything in return, despite what you think)

5.2.4 Textual uses

In MSA, demonstratives of "vague reference" are used paragraph initially to refer anaphorically to the whole of an immediately preceding argument or chunk of text, for example, *kaða:lika* 'Moreover, furthermore' (lit. 'like that'), *maʻa ða:lika* 'Despite that, . . . '. In spoken MSA, particularly that of news bulletins, the phrase *ha:ða:, wa . . .* 'this, and . . . ' is frequently used postpausally as a means of connecting two sections of a report, the (invariably masculine form) *ha:ða:* referring back to the whole of what has just been reported and signaling that something more, but different, is about to be said about the same topic. For example, a television news report on the October 1992 earthquake in Cairo began with a description and film of the death and destruction caused in the city and then, following a cut to shots of the pyramids and the Sphinx, continued (literally translated):

'*This* (*ha:ða:*), and Egyptian officials announced that the High Dam and several ancient Egyptian remains such as the pyramids and the Sphinx had not been affected by the earthquake . . . '

ha:ða: thus performs the dual cohesive role of marking a shift in the information focus but at the same time signaling that what is coming next is topically connected to what has preceded.

5.3 Interrogatives

The form of the interrogative pronouns is another point at which MSA diverges markedly from the dialects. Although there are superficial differences between them, the dialects as a group have developed in similar ways, while MSA continues the CLA system. Table 5.4 gives the MSA forms and those of Cairo, Syria/Palestine, and the Gulf states.

The following observations may be made on the historical[9] and typological relationships between the forms in this table:

(a) Several of the dialectal forms are derivable from the CLA/MSA ones via general phonological changes that affected the dialects. Among others, one may note: (i) the general loss of the glottal stop intervocalically, (ii) the loss of final −*a*, and (iii) the replacement of diphthongs by long vowels, all of which processes operated in the case of *?ayna* → *fe:n/we:n* (probably < *fa ?ayna* 'so where', *wa ?ayna* 'and where' respectively). Other phonological changes were more localized, as in the affrication of the velars, typical of the Gulf and Iraq but not the urban dialects of the eastern Mediterranean, whence *če:f* < *kayfa* and *čam* < *kam*.

In other parts of the system, the changes were not phonological in origin:

(b) *ma:(ða:)* 'what?' (and its derivative *li ma:(ða:)* 'why?' (lit. 'for what?') in CLA and its descendant MSA did not survive in the dialects.[10] The disappearance of interrogative *ma:* almost everywhere was the consequence of the generalization of the homophonous negative particle *ma:* 'not' as a result of the simplification of the CLA system of negativisation (see 6.3) The negative *ma:* 'not', whose main use in CLA had been to negate the s-stem verb, developed in the dialects into the basic negative particle used to negate the p-stem as well. Because both negative and interrogative *ma:* are placed immediately before the verb, ambiguity would have arisen routinely in interrogative and negative sentence structures common in everyday conversation. To replace interrogative *ma:* 'what?', a variety of forms developed that differ slightly from one dialect to another but that are all ultimately abbreviations of the post-Classical phrase:

wa	*?ayy*	*šay*	*hu*
and	which	thing	it

This is the source of the forms *?e:š(hu)*, *we:š(hu)*, and *šu* 'what?' noted in the chart.[11] The 'why?' forms are formed via the preposing of the preposition *li* 'for'

Table 5.4   Interrogatives in MSA and Three Dialects

Interrogatives	MSA	CAIRO	SYRIA/PALESTINE	GULF
What?	*ma:(ða:)*	*?e:h*	*šu:,* *?e:š(hu)*	*š-, šu, šinhu,* *we:š(hu)*
Why?	*li ma:(ða:)*	*le:h*	*le:š*	*šhaggah, šlih* *liwe:š(hu)*
Which (one)?	*?ayy*	*?anhu/* *?anhi/* *?anhum*	*?anhu/* *?anhi/* *?anon*	*?ay(hu)/* *?ay(hi)/* *?ay(hum)*
Who?	*man*	*mi:n*	*mi:n*	*min(hu)*
Where?	*?ayna*	*fe:n*	*we:n*	*we:n*
When?	*mata:*	*?imta*	*?imta*	*mita*
How?	*kayfa*	*iz-zayy*	*ki:f*	*če:f, šlo:n*
How much?/ How many?/ How long?	*kam*	*ka:m,* *?addie:h*	*?adde:š*	*čam, šgadd,* *škiθir*

to the first two of these, with appropriate phonological adjustments (loss of intervocalic /?/): *le:š, liwe:š*. The Gulf (and central Arabian) 'what?' form *šinhu* (variant *we:šinhu*) is also derived from *wa ?ayy šay hu*, but here the CLA case ending of *šay (šay-in)* seems to have been fossilized: unsurprising in dialects that have, as we have noted elsewhere, preserved vestiges of the CLA case system in a limited number of syntactic contexts. Those dialects that have the *we:šhu/šinhu* type of forms for 'what?' tend to have analogous formations in the other parts of their interrogative system: thus Bahraini *we:šhu/šinhu* 'what?', *minhu* 'who?', *?ayhu* 'which?'.

(c) A further general dialectal development, basically an extension of the phrase apocopation noted in (b), has been the evolution of *šay* 'thing' into a flexible interrogative clitic morpheme *š*. This *š* is attached to lexical items indicating size, amount, type, etc., to form compound interrogative pronouns that replaced the morphologically heterogeneous CLA forms. *š* is preposed in some dialects, postposed in others. Typically, Damascene postposes, with forms such as *?adde:š* 'how much?' < *qadd + ?ayy + š*, lit. 'amount which thing'; in the Gulf and contiguous regions where such compound interrogatives are very common, *š* is, as a rule, preposed: *šlih* 'why?' < *š + lih* 'what for-it', *šhaggah* 'why?' < *š + ḥaggah* 'what for-it' *šgadd* 'how much?' < *š + gadd*, lit. 'what amount', *škiθir* 'how much?' < *š + kiθir* lit. 'what amount', *šlo:n* 'how?' < *š + lo:n*, lit. 'what color/type', and many others. Similar formations occur in the related Iraqi dialects, for example, Baghdadi *šwakit* 'when?' < *š + wakit* lit. 'what time'.

(d) In many dialects (generally those of the Levant), a particle *ši:*, etymologically identical to this *š* (< *šay* 'thing'), can be suffixed to statements to turn them into yes/no questions. For example:

'inda-k   flu:s   ši:
with-you   money   thing
'Do you have any money?'

The same *š* (or its allomorph *ši:*) forms the second element of the discontinuous bound negative morpheme *ma: . . . š(i:)* (cf. French *ne . . . rien*, whose lexical origin is analogous). Thus an answer in the negative to the above question would be:

la, ma: 'ind-i:-š
no, not with-me-thing
'No, I haven't'

It is clear that the loss of "interrogative" *ma:*, the generalization of "negative" *ma:*, and the evolution of *šay* 'thing' into a flexible clitic element with complementary interrogative and negative functions are interconnected developments and together provide a further illustration of how the spoken dialects changed and simplified, by making more transparent the "synthetic" language system that was their starting point. In particular, the development of a frequently occurring lexical item of vague reference, such as *šay* 'thing', into a particle with "core" morphosyntactic functions of interrogation and negation is typical. As we shall see in chapter 6, a number of other lexical elements of similarly general reference were also morphologized to form aspectual particles in the verb phrase (see 6.2.2) and genitival particles in the noun phrase (see 6.1.3.3) to replace other "synthetic" elements of CLA morphosyntax.

5.4 Temporal, locative, and manner deictics

The dialectal equivalents of the CLA deictics of time, place, and manner came about by similar processes to those we saw in 5.2 and 5.3: phonological changes affected some OA forms, but in other cases completely new words came into being, derived from the "abbreviation" of post-Classical phrases consisting of a demonstrative clitic and a "content" noun (compare 5.3 (b)–(c) above). Thus from CLA *huna:* 'here' and *huna:k* 'there' developed, via local phonological changes:

	'here'	'there'
Damascus	*ho:n*	*(?)ihni:k*
Baghdad	*hna:*	*hna:k*
Cairo	*hina*	*hina:k*
Gulf	*(?)ihni*	*(?)ihna:k*

Equivalents of 'like this' are more varied, having arisen from the synonymous OA variants **ha:kiða:* and **kiða:*. For example:

	'like this'	
Damascus	*he:k*	(< OA *ha:kiða:*)
Baghdad	*hi:č(i)*	(< OA *ha:kiða:*)
Cairo	*kida*	(< OA *kiða:*)
Gulf	*čiði*	(< OA *kiða:*)

Prefixed by the local negative particle, each of these forms a phrase that is used to make tag questions, for example, Damascus *ma: he:k*, Baghdad *mu hi:č* 'is that not so?', used like French *n'est-ce pas* and German *nicht wahr*.

Dialectal words for 'now' are mostly apocopations of different post-Classical phrases, all of which have the same structure and mean 'this time/hour', the CLA/MSA form *al-?a:n* not having survived in the dialects:

Damascus	*halla?*	< *hal-wa?t* < *ha:ða: l-waqt*	'this time'
Baghdad	*hassa*	< *has-sa:'a* < *ha:ðihi s-sa:'a*	'this hour'
Cairo	*dilwa?ti*	< *da-lwaqt* < *ha:ða: l-waqt*	'this time'
Gulf	*(ha)lhi:n*	< *ha:ða: l-hi:n*	'this time'

The *ha-* element may be freely prefixed to other relational nouns to form deictic adverbs of manner in the same way as we saw earlier that *š-* may be prefixed to form interrogatives, for example, in Gulf Arabic *hallo:n* ('this color'), *haššakil* ('this shape'), *hannamu:na* ('this type') = 'like this'; *halgadd* ('this amount'), *halki θir* ('this amount') = 'this much'. Compare *šlo:n* ('what color') = 'how?', *šgadd* and *škiθir* ('what amount') = 'how much?'.

Notes

1. Cowell 1964, 539.
2. Abboud 1964, 16.
3. "Dialects" indicates that the syntactic structure is common, even if there are slight differences in phonology from dialect to dialect.

4. Brockelmann 1908, 1:317–18.

5. This statement is not strictly true as applied to some of the urban dialects, where a copula verb has developed from a verb with the basic meaning 'to become' and is used as an alternative to the kind of structure exemplified. Thus in Cairo one can also say:

da   yib?a   be:t-i
this 3msg-remain house-my
'This is my house'

6. Among other particles, notably *le:k* and *sa'*- (Cowell 1964, 564–65).

7. Brockelmann 1908, 1:241.

8. This seems to be a development of the presentative *ha:* 'see!' plus "distal" -*k* and occurs in this form in Morocco also (Fischer 1959, 163). Alternatively, it could be a secondary development of the related structure *ha:-* + second person enclitic: *ha:k* (msg.), (f. *ha:č*, pl. *ha:kum*), meaning 'Here you are!', used when handing something over to someone. The -*k*- in sentences such as that exemplified would then be a kind of "ethic" dative.

9. See Singer 1958 for a comprehensive historical treatment of the development of the dialectal interrogatives.

10. An exception is found in the dialects of Omani sedentaries, which have *mu* 'what?' < *ma:* + *hu* lit. 'what it', as in *mu smak* 'What's your name?'

11. The Cairene forms *?e:h* and *le:h* seem to have been formed in the same way as the others cited, via the loss of the final -*š* element.

6

Syntax and Semantics I: Phrase Structure
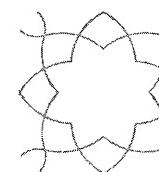
This chapter describes the syntactic structure of the commonly occurring types of noun and verb phrase in Arabic; chapter 7 is concerned with clause and sentence structure. In both chapters, the focus is on what is "core"to both MSA and the modern dialects, but a good deal of attention is also paid to areas of structure in which the dialects as a group have followed a different path of development from that of MSA.[1] In the noun and verb phrase, the dialectal developments, unplanned and unconscious, did not simply involve the long-term simplification of the CLA morphosyntactic system via loss of the case and mood inflections, collapse of verb conjugations, etc., but a great deal of compensatory innovation—the morphologization of verbal and other elements to form a system of modal/aspectual prefixes and auxiliaries in the verb phrase, the development of a periphrastic genitive in the noun phrase, and, in the clause, the fixing of word order—all of which have parallels in the evolution of modern European languages. Since the mid-nineteenth century, however, MSA has evolved from the moribund literary traditions of Classical Arabic in a conscious and (relatively speaking) controlled response to the need for a flexible modern written idiom. Although written MSA continues (more or less) to observe at the level of surface syntax the rules laid down by the grammarians for "correct/eloquent" (*faṣi:ḥ*) Arabic some twelve centuries ago, it has undergone massive changes in the way that semantic content is mapped onto

syntactic form, which seem to have accelerated with the information technology explosion since the mid 1970s. Here the influence of the foreign-language media has been pervasive, with the result that much journalistic and expository "secular" Arabic writing has begun to approximate to what Blau,[2] after Blanc, has termed Standard Average European syntax. Some of the most obvious tendencies in this regard are illustrated in the section of chapter 8 devoted to "media Arabic."

6.1 The noun phrase

6.1.1 Determination of nouns

In MSA, common nouns, for example, *rajul* 'man', *kita:b* 'book', *wuṣu:l* 'arrival', *ḥubb* 'love', may be made grammatically definite (see also 4.5) in three ways: by prefixing to them the definite article, by adding a pronominal suffix, or by using them as first element in the construct state when the second element is definite (see 6.1.3. below).

6.1.1.1 Definite article

The proclitic definite article *(ʔa)l-* 'the' may be prefixed. Orthographically, *(ʔa)l-* is never written as a separate word. It serves two broad discoursal/semantic functions. On the one hand, as in many languages, it is used to *specify* an individual example of a category, and on the other, to refer *generically* to a whole category. Thus, typically, in the first case, it specifies a particular 'man', 'book', etc., as the same one as has already been referred to previously in a text or conversation:[3]

> ka:na   fi: baġda:d rajul baxi:l   wa ka:na   **r-rajul** ...
> was-3msg   in   Baghdad man miserly and was-3msg the-man ...
> 'There was in Baghdad a miserly man, and **the man** ...'

There is also a category of common nouns (in some cases with modifying adjectives) that, when used to denote unique natural phenomena or particular named entities or events, always have the article, for example, *ʔal-ʔarḍ* 'the earth', *ʔal-qamar* 'the moon', *ʔal-ḥarbu l-ʕuðma:* 'The Great War', *ʔadduwalu l-kubra:* 'the superpowers', *ʔal-wila:ya:tu l-muttaḥida* 'the United States'. As an indefinite noun, *ʔarḍ* means 'land, territory' in a quantifiable sense, and *qamar* means simply any 'moon' or 'satellite'.

In the case of a few temporal expressions, the "specifying" use of the article is extended to that of contextual circumstance: for example, *ʔal-yawm* 'today' (lit. 'the day') = 'on the day we are now in', *ʔas-sa:ʕa* 'at the time' (lit. 'the hour') = 'at the time in which the action described took place'.

In contrast to the above, the article may be used when referring *generically* to actions, states, or properties of all the individual members of a class, or to actions, qualities, substances, etc., in the abstract where these are not individuated. For example:

> *ʔal-ja:r*   qabla d-da:r   wa r-rafi:q   qabla   t-ṭari:q
> the-neighbor before the-house and the-companion before the-road
> 'Choose your neighbor before your house and your traveling companion before your road' (proverb)

> *ʔal-jadi:d fi: ṭ-ṭibb*
> the-new   in the-medicine
> 'What's new in medicine' (title of BBC Arabic Service program)

> *ʔan-nu:r wa d-dayju:r*
> the-light and the-dark
> 'Light and Dark' (title of book)

> *qami:ṣ mina l-quṭn*
> shirt of the-cotton
> 'A cotton shirt'

Other uses of the article follow from these general principles. So, for example, the article is usually used with active/passive participles and adjectives when these are being used as nominals, because in this case they are making generic reference to a class of agents or patients. For example:

> *ʔal-qa:dimu:na   wa l-muġa:diru:na*
> the-arrive-A.PART-pl and the-leave-A.PART-pl
> 'Arrivals and Departures' (lit. 'the arrivers and the departers')
> (airport sign)

> *ʔal-ḥa:ṣil ...*
> the-happen-A.PART
> 'In short, ...' (lit. 'the thing that happens')

> *mina l-maʕlu:m   ʔanna ...*
> of   the-know-P.PART that ...
> 'It is known that ...' (i.e., 'it is of the class of known things that ...')

> *ʔas-saʕi:d   huwa llaði: ...*
> the-happy he   who
> 'A happy man is he who ...'

The article is normally used with the verbal noun, whether it is being used generically or is describing an action predicated of a particular individual:[4]

generic:

ʔaṣ-ṣala:tu xayrun mina n-nawm
the-prayer better than the-sleep
'Prayer is better than sleep' (from the morning call to prayer)

ʔat-tadxi:n xaṭar ʕala: ṣ-ṣiḥḥa
the-smoking danger to the-health
'Smoking is a danger to health' (public sign)

ʔal-ḥaya:
the-life
'Life' (title of daily newspaper)

specific:

sa:fara ʔila: ʔamri:ka: li l-ḥuṣu:l ʕala: ṣaha:da
travelled-3msg to America for the-getting of certificate
'He went to America to get a (university) degree'

lam ʔasmaḥ lahu bi d-duxu:l
not 1sg-allow to-him with the-entering
'I did not allow him to enter'

la: tastaṭi:ʕu ð-ðaha:b
not 3fsg-be-able the-going
'She cannot go'

There is a tendency to define nominals with proclitic *(ʔa)l* where these indicate amounts, especially when used in a relative or comparative sense. For example:

li l-mazi:d mina l-maʕlu:ma:t ʔal-mura:jaʕa lada: . . .
for the-increase of the-information the-application to . . .
'For more information please apply to . . . '

yaʕtaqidu l-kaθi:r mina n-na:s ʔanna . . .
3msg-believe the-many of the-people that . . .
'Many people believe that . . . '

The proclitic definite article in the dialects, generally of the form *il-*, is used in the same ways and for the same purposes as *(ʔa)l-* in MSA.

6.1.1.2 Pronominal suffixation

The second main means of making common nouns definite is by suffixation of a pronominal enclitic directly to the noun (which is marked for case according to its syntactic function in the sentence). The relationship between the noun and the enclitic may be that of:

(i) noun-possessor
e.g., *qaṣruhu* (*qaṣr-u-hu*) 'his palace'
palace-NOM-his

(ii) part-whole
e.g., *kulluhu* (*kull-u-hu*) 'all of it/him'
all-NOM-him

(iii) verbal noun-agent
e.g., *ziya:ratuhu* (*ziya:rat-u-hu*) 'his visit'
visit-NOM-his

(iv) verbal noun-patient
e.g., *ʔiʕda:muhu* (*ʔiʕda:m-u-hu*) 'his execution'
execution-NOM-his

mawtuhu (*mawt-u-hu*) 'his death'
death-NOM-his

Because Arabic verbal nouns derived from Pattern I transitive verbs are not marked for voice—*qatl* can mean 'killing' or 'being killed'—there is a potential ambiguity in a form such as *qatluhu*. This is resolved by context or by an amplifying complement. For example:

qatluhu (li) r-rajul
killing-his (to) the-man
'His killing (of) the man'

The use of the preposition *li*, whose original sense is benefactive, in structures of this kind seems to have become particularly common in media Arabic. Thus, during the 1990–91 Gulf War, one routinely read of *ġazwu l-ʕira:q li l-kuwayt* (lit. 'attack of Iraq to Kuwait') 'Iraq's invasion of Kuwait', where *li* occupies the same syntactic slot as 'of' in the equivalent English phrase. In older and nonmedia styles of written Arabic an equivalent structure without *li* seems to be more com-

mon, which possibly suggests that the modern popularity of the use of *li* in the media phrase is a European influence.

There have been significant "analytic" developments of structures (i) and (iii) in the dialects, to which the noun-plus-enclitic structures illustrated here are functionally similar. These are dealt with below in 6.1.3. on the so-called construct state or phrase.

6.1.1.3 First element in the defined construct state

Any noun that occurs as the first noun in the construct state in which the second is grammatically definite (i.e., has the definite article prefixed or has a pronoun suffix) is also definite. For example:

> *ziya:ratu r-rajul*
> visit the-man
> 'The man's visit'
> *mawtu binti-hi*
> death daughter-his
> 'His daughter's death'

Where the second noun becomes definite by virtue of itself being the first element in a grammatically definite construct state, the first noun in the phrase is also definite (and so on, theoretically at least, ad infinitum):

> *mawtu binti 'ammi-hi*
> death daughter uncle-his
> 'His cousin's death'

> *xabaru mawti binti 'ammi-hi*
> news death daughter uncle-his
> 'The news of his cousin's death'

6.1.1.4 Proper nouns

Proper nouns—the names of particular individuals, places, etc.—are inherently grammatically definite, although by no means all are morphologically marked as such. Thus, whereas many place names have proclitic *(?a)l-* (or its dialectal equivalent), for example, *?al-'ira:q* 'Iraq', *?as-su:da:n* 'Sudan', *?al-qa:hira* 'Cairo', *?al-jaza:?ir* 'Algiers/Algeria', others, for example, *'uma:n* 'Oman', *miṣr* 'Egypt', *lubna:n* 'Lebanon', do not. Recently borrowed and transliterated place and country names do not usually have *(?a)l-*, for example, *?amri:ka:* 'America',

ba:ri:s 'Paris' (although note the older arabicization with the article *?an-namsa:* 'Austria' (probably via Turkish *Nemçe* < Hungarian *németség* 'the Germans')). Given names are definite by virtue of their personal specificity and do not normally have *(?a)l-*, but in some contrastive contexts *(?a)l-* is used (cf. English 'The Peter I told you about isn't him'), as it is in deictic ones with other proper nouns that do not otherwise require the article (cf. 'This dear England, which . . . ').

6.1.2 Agreement in noun–adjective phrases

In MSA, attributive adjectives generally follow the nouns they modify (the major exception being comparative constructions, for which see 6.1.4 below) and agree with them in number, gender, case, and definiteness. This principle of agreement applies also, *mutatis mutandis*, to the dialects, although, taken as a whole, they are somewhat simpler in the sense that they make fewer category distinctions, because of the loss of case, loss of the dual form of the adjective, and loss of feminine plural agreement in many urban and some recently urbanized Bedouin dialects of the northern Gulf.

6.1.2.1 Definiteness

Grammatically definite nouns also require their attributive adjectives to be overtly marked as definite by the prefixing of *(?a)l-*.[5] For example (MSA examples):

indefinite phrases:	definite by procliticization:
rajulun ġabi:yun	*?ar-rajulu l-ġabi:yu*
man stupid	the-man the-stupid
'a stupid man'	'the stupid man'
bintun ġabi:yatun	*?al-bintu l-ġabi:yatu*
girl stupid-fsg	the-girl the-stupid-fsg
'a stupid girl'	'the stupid girl'

definite by encliticization:
bintuhu l-ġabi:yatu
girl-his the-stupid-fsg
'his stupid daughter'

proper nouns:	
'i:sa: l-kabi:ru	*?al-qa:hiratu l-kubra:*
Isa the-great	the-Cairo the-greater-fsg
'Isa the Great'	'Greater Cairo'

6.1.2.2 Gender and number

In all varieties of Arabic, singular nouns are masculine or feminine (by form, meaning, or convention—see 4.3) and require masculine and feminine adjectives respectively, whether the adjective is part of a noun–adjective phrase or a predicate. There are a few examples of singular nouns that are feminine by form (ending -a) but have exclusively male reference, such as *xali:fa* 'caliph', that require masculine agreement and a few that are masculine by form but female in reference, for example, *ḥa:mil* 'pregnant woman', which require feminine agreement. There is also a category of collectives in both MSA and the dialects that denote certain national, tribal, or occupational groups and that end in the same normally feminine ending -a but that follow the normal plural noun agreement patterns, for example, MSA *baya:rita* 'inhabitants of Beirut' (pl. of *bayru:ti:y*), Bahraini Arabic *bawayga* 'robbers' (pl. of *ba:yig*).

As far as plural nouns are concerned, the prime factor that determines agreement is whether the noun's reference is human or nonhuman. If the modified noun is human, the system is one of so-called strict agreement in which the adjective is marked for both gender and number. The following diagram illustrates this:

ADJECTIVE

m	sg.	*rajul*	*ġabi:y*	'a stupid man'
NOUN [+ human]	pl.	*rija:l*	*ʔaġbiya:ʔ*	'stupid men'
f.	sg.	*bint*	*ġabi:ya*	'a stupid girl'
	pl.	*bana:t*	*ġabi:ya:t*	'stupid girls'

But if the noun is nonhuman, its plural form, regardless of the gender of its singular, requires a feminine singular adjective ("deflected" agreement):

m.	sg.	*kita:b*	*qadi:m*	'an old book'
NOUN [- human]	pl.	*kutub*	*qadi:ma*	'old books'
f.	sg.	*madrasa*	*qadi:ma*	'an old school'
	pl.	*mada:ris*	*qadi:ma*	'old schools'

The dual forms simply involve suffixing -*a:ni* (nom.) or -*ayni* (acc./gen.) to the singular noun and its modifying adjective, for example, *rajula:ni ġabi:ya:ni* 'two stupid men', *madrasata:ni qadi:mata:ni* 'two old schools'.

This agreement system, nowadays the norm in all expository, nonliterary writing, is a simplification of the CLA (roughly pre-tenth century) one, which it is thought was a reflection of the situation in the Old Arabic dialects. In the Arabic of the pre-tenth century, it was common[6] for broken nonhuman plural nouns to take "broken" plural adjectives as an alternative to fsg. or (if the adjective had no broken pl. form) f. suffixed pl. In MSA, this type of plural concord in nonhuman nouns can still be found, mainly in writing whose authors consciously cling to the Classical tradition, and in fixed phrases borrowed from it. Thus Mikha'il Nu'ayma, a twentieth-century literateur who frequently resorted in his moral philosophic essays to a rather mannered and archaic style, could write in 1966 of 'dark nights' as *laya:lin da:misa:t* (f. suffixed pl. adj.) rather than *laya:lin da:misa* and of 'bygone eras' as *'uṣu:r xawa:lin* (broken pl. adj.) rather than *'uṣu:r xa:liya*;[7] the ancient Arabian odes that are supposed to have been suspended from the Ka'ba in seventh-century Mecca are conventionally referred to using the Classical concord system as *ʔalqaṣa:ʔidu ssab'u ṭṭiwa:lu lja:hili:ya:t* 'the seven long pre-Islamic odes' (with broken pl. and suffixed f. pl. adjectives), rather than as *ʔalqaṣa:ʔidu ssab'u ṭṭawi:latu lja:hili:ya* (with fsg. adjectives). Some of the modern dialects continue to employ something akin to the old Classical agreement system, but, increasingly, educated speech throughout the Arab world reflects the simplified norms of MSA. As in many other areas of syntax and morphology, the more conservative, Classical-like dialects are those of the relatively recently settled Bedouin of central Arabia and the neighboring areas. Here, the Classical system of "strict" agreement for both human and nonhuman plurals is largely preserved (although the dual has in all cases fallen in with the plural), at least by those not exposed through schooling to the standardizing effects of literacy,[8] although there are many local variations.[9]

A noteworthy development in many dialects, urban and rural, is that the agreement pattern varies according to the speaker's perception of the individuated versus generic reference of the noun, whether human or nonhuman.[10] For example, the word *na:s* 'people', and other words denoting human groups, frequently has "deflected" (fsg.) agreement where it is being used to refer to 'people in general', but "strict" (mpl.) agreement where reference is to a countable, known, or otherwise previously specified group of individuals, for example, 'the people who live in my street'. There seems to be a saliency hierarchy at work: the more individuated, particularized, and countable the reference, the more likely the noun is to require "strict" (i.e., plural) agreement.[11] In MSA, however, such human nouns tend to have strict agreement whether their reference is individuated or generic. A similar pattern of concord in the dialects can be seen where the head noun is nonhuman. Cowell[12] gives the following examples (transliterated here according to the system used in this book) for Damascene Arabic, in which the verb form varies with variation in the semantic reference of the same nonhuman noun:

(i) *ilkutub   ma: bihimmu:*
the-books not *b*-3pl-interest-him
'The books don't interest him'

(ii) *ilkutub   ma: bithimmu*
the-books not *b*-3fsg-interest-him
'Books don't interest him'

In (i), where *ilkutub* has individuated reference (e.g., 'the books I lent him last week'), the verb is plural; in (ii), where it has generic reference (= 'books in general'), the verb is feminine singular. The MSA equivalents of these sentences would not make this distinction, requiring the feminine singular verb in both cases.

6.1.2.3 Case

In MSA, attributive adjectives are obligatorily marked for the same case, nominative, accusative, or genitive, as the nouns that they modify. This does not mean, of course, that in all cases the phoneme(s)/grapheme(s) that mark the agreement are the same, because the noun and adjective may belong to different declensions, in which there are differences in how the same grammatical case is marked (see 4.5).

6.1.3 The construct phrase

The so-called construct state, or phrase (in Arabic *ʔidaːfa* 'annexion'), involves the juxtaposition of two nouns that may have one of a number of syntactic relationships. Formally, the second, amplifying, noun, must, in MSA, be in the genitive case. The first, "annexed" noun must always be grammatically indefinite (i.e., lacking a proclitic *(ʔa)l-* or an enclitic pronoun); the amplifying noun may be definite or indefinite. Constructs consisting of strings of more than two nouns are possible ('the X of the Y of the Z of the . . . '), although in practice a string is rarely longer than three. As in the case of two-noun constructs, only the final noun of a multinoun construct may carry a marker of definiteness. Where the first term of the construct is not singulative, but a coordinated phrase, that is, 'the X and the Y of Z', it is considered good (i.e., literary) style to turn the phrase 'the X of Z and its Y' (where 'its' refers to 'Z'). However, in media MSA it is now very common to find the coordinated alternative. This preference is probably partially because of the influence of analogous coordinated structures in English and French, which have affected the development of the Arabic media style (see chapter 8). A similar shift seems to be occurring in prepositional phrases: 'taking the children to and from school' should in "good" Arabic literary style be ' . . . to school and from it' but the BBC Arabic Service, in a

piece called "Should Houswives Be Paid for Housework?" in 1984 rendered this phrase *tawsiːlu lʔatfaːl min wa ʔilaː lmadrasa*, mirroring the English structure.

6.1.3.1 Types of construct

The semantic relationships that may exist between the terms of a construct are broadly six in number:[13]

(i) identity
for example,   *madiːnatu dimašq* 'the city of Damascus'
              fariːdatu lḥajj 'the duty of pilgrimage'

(ii) partitive/quantitative relationship
for example,   *finjaːlu qahwa* 'a cup of coffee'
              ṭaliːʕatu θθawra 'the vanguard of the revolution'

(iii) possession
for example,   *sayyaːratu lmudiːr* 'the boss's car'

(iv) agent relationship
for example,   *wuṣuːlu lmalik* 'the king's arrival'

(v) object relationship
for example,   *ʔiʕdaːmu lmujrim* 'the execution of the criminal'

(vi) limitation of the first term by the second
for example,   *baʕiːdu lmadaː* 'long-term' (lit. 'long of the duration')

Some additional comment is in order here. First, the "partitive" construct is used in many cases where in European languages a noun–adjective phrase would be normal. For example:

two-term construct:
kibaːru lmasʔuːliːn
big-pl the-officials
'the important officials' (lit. 'the big (ones) of the officials')

three-term construct:
fiː muxtalifi ʔanḥaːʔi lʕaːlam
in various parts   the-world
'in various parts of the world' (lit. 'in the various of the parts of the world')

Here, what might be thought of as attributive adjectives, *kibaːr* and *muxtalif,* function exactly like the noun *finjaːl* 'cup' in *finjaːlu qahwa* 'a cup of coffee'.

Second, the "possessive" function of the construct is a very broad and vaguely defined category, for example, *dawwaːru ddaːxiliːya* 'the Interior Ministry roundabout' is strictly a relationship of proximity rather than of possession, whereas *saːʕaːtu ddawaːm* 'working hours' (lit. 'hours of the routine') is one of con-

stituency: 'the working day has (= 'consists of') hours' seems to be the underlying logical relationship.

Third, in category (vi), although formally identical to the others, the function of the second term is dissimilar in that it specifies the semantic scope of the first. Often, category (vi) constructs correspond to compound adjectives in English, for example, from *ṭawi:l* 'long, tall', we get:

ṭawi:lu ššaʿr	'long-haired' ('long of the hair')
ṭawi:lu llisa:n	'impertinent' ('long of the tongue')
ṭawi:lu rru:ḥ	'long-suffering' ('long of the spirit')
ṭawi:lu lbaːʿ	'mighty' ('long of the arm')

Category (vi) constructs behave syntactically like ordinary adjectives when used attributively, that is, if the head noun is definite, *l-* must be prefixed. The bracketed example below indicates the immediate constituents in such cases:

[ʔaššaːbbu [ṭṭawiːlu ššaʿr]]
NOUN   ADJECTIVE
the-youth the-long-(of)-the-hair
'the long-haired youth'

If the definite article were not prefixed to the adjective, the meaning would be 'the youth is long-haired'.

If the first term of a construct is a dual or a masculine plural formed by suffixation (4.4.1.1), the final *-n* of the suffix and its vowel are dropped. Thus *ʔal-mudarrisuːna* 'the teachers', but *mudarrisuː lmaʿhad* 'the institute teachers'. The same basic principle applies to noun-plus-enclitic structures, thus *ʿaynaːni* 'two eyes', but *ʿaynaːka* 'your(m.) two eyes' via deletion of the *-ni* of the dual suffix and the suffixing of *-ka* 'your'.

6.1.3.2. Constructs and attributive adjectives

Where a construct phrase is modified by an adjective, the adjective must, according to the rules of MSA, follow the whole phrase, even if the adjective modifies the first term (or 'head noun'). For example:

[[maraḍu naqṣi lmanaːʿati]   lmuktasabu]
illness   lack   the-immunity the-acquired
'Acquired Immune Deficiency Syndrome (AIDS)'

Here the adjective *muktasab* 'acquired' is marked as masculine and nominative because it modifies the masculine head noun *maraḍ*, even though *manaːʿa* 'immunity', the immediately adjacent noun, is feminine and genitive. Where the terms of the construct are all of the same gender and number, however, the resulting phrase may be ambiguous prepausally, because the final disambiguating case vowel is not sounded. For example:

šaʿru ššaːbbi   ṭṭawiːl (pause)
hair the-youth the-long

could be parsed

[[šaʿru ššaːbbi] ṭṭawiːl] (missing final vowel = *-u*)
'the youth's long hair' (where *ṭawiːl* modifies *saʿr*)

or

[šaʿru [ššaːbbi ṭṭawiːl]] (missing final vowel = *-i*)
'the tall youth's hair' (where *ṭawiːl* modifies *šaːbb*)

As we have seen with the AIDS example, this problem does not arise in cases where the gender and number of the terms are different. Nonetheless, although a phrase such as

[[madrasatu lbanaːti] lʔinkiliːziːya]
school-fsg the-girls the English-fsg

can only be understood as meaning 'the English school for girls' and not 'the school for English girls' because of the number marking on the adjective, in contemporary MSA an alternative type of structure involving a prepositional complement is often employed in order, it would seem, to reduce the functional load on the morphological marking of the adjective. Thus one tends to encounter:

ʔalmadrasatu lʔinkiliːziːya   li lbanaːt
the-school-fsg the-English-fsg for the-girls
'The English school for girls'

instead of the bracketed version. This, of course, is another example of the move away from a "synthetic" type of structure whose interpretation depends on inflection to a more analytic type in which syntacticosemantic relationships are

lexicalized. In media and political styles of MSA, such prepositional structures have become very common in the coining of equivalents for complex noun phrases such as Organization of Petroleum-Exporting Countries (OPEC) (see 8.3). But even in simple possessive relationships where there is no adjective and where one might have expected to find a simple construct, there seems to be an increasing tendency to use *li*, for example, *ka:tim li ṣṣawt* '(gun) silencer' (lit. 'hider for the sound'), *dawwa:ma li l'unf* 'spiral of violence'. In media MSA, *li* seems in fact to function very much like English 'of', even in contexts where there is no discernible possessive relationship between the nouns that it links, for example (from a recent magazine editorial) *?attadaffuqu lha:?il li rrasa:?il* = 'the huge flow *of* letters', where 'flow' is a measure of quantity. This development of the function of *li* parallels tendencies in clause nominalization (see the examples in 6.1.1.2 above).

In the case of a few fixed phrases, the rule that demands that attributive adjectives be placed at the end of constructs is consistently ignored. The most commonly encountered of these is 'secretary general' in, for example, 'Secretary General of the United Nations':

[?ami:n 'a:mm]u l?umami   lmuttaḥida
secretary general the-nations the-united

where the adjective *'a:mm* should theoretically be placed at the end of the phrase. The explanation is that the adjective in the phrase *?ami:n 'a:mm* refers to an essential, not a contingent, component of the meaning; accordingly, the phrase is treated as if it were a unitary noun.

6.1.3.3 Constructs in the dialects

The phrase types illustrated in 6.1.3.1–2 also occur in the dialects, but in order to express certain kinds of semantic relationship—chiefly those we have categorized as "partitive" and "possessive," but also "agent"—they have developed an alternative, analytic, structure. This involves linking the first and second terms with a particle, which differs from dialect to dialect but seems in most cases to have originated in a noun meaning 'right', 'belonging', or 'property'.[14] In some cases (*bita:', ma:l*), the particles may optionally be marked for gender with suffixed *-at* or *-it* where the first noun is feminine. Generally speaking, this construction is used only to express alienable possession; inalienable possession is expressed by the construct phrase, as in MSA (e.g., *ša'r iṣṣa:bb* 'the youth's hair').

For example:

Cairo	*il?a'la:m bita:'it iṣṣa'b*
	the-flags POSS-PTCLE-f the-people
	'the people's flags'
Amman	*iljisir taba' ilmuxa:bara:t*
	the-bridge POSS-PTCLE the-intelligence-services
	'the Intelligence Services' flyover' (i.e., the one near their building)
Baghdad	*dihin ma:l ša'ar*
	oil POSS-PTCLE hair
	'hair oil'
Bahrain	*bu:k ḥagg ilflu:s*
	wallet POSS-PTCLE the-money
	'money wallet'
Rabat	*ssiya:ra dyal 'ammi*
	the-car POSS-PTCLE uncle-my
	'my uncle's car'

These particles are also used to form alternative possessive constructions via the suffixing of the second element of the construct, which then stands grammatically in apposition to the first. This dialectal construction often corresponds to the MSA noun + enclitic pronoun structure (see 6.1.1.2 above). For example:

MSA	*?a'la:m-u-na:*	Cairo	*il?a'la:m bita:'it-na*
	flags-NOM-our		the-flags POSS-PTCLE-our
		'our flags'	
MSA	*bayt-u-ha:*	Gulf	*ilbe:t ma:l-ha*
	house-NOM-her		the-house POSS-PTCLE-her
		'her house'	
MSA	*mudarris-u:-ka*	Amman	*ilmudarrisi:n taba'-ak*
	teachers-NOM-your		the-teachers POSS-PTCLE-your
		'your(m.) teachers'	

Analytic constructions are particularly favored in the dialects in two similar types of noun phrase:

(a) in phrases where the first element of a definite construct has a modifying adjective that would, according to the rules of normal MSA, have to be placed after the whole of the construct phrase. Thus, in the Gulf dialects, the phrase equivalent to 'my friend's big house' would usually be:

[ilbe:t   ličbi:r]   ma:l   ṣadi:gi
the-house the-big POSS-PTCLE friend-my

rather than a dialectalized version of the MSA structure:

[[baytu   ṣadi:qi:]   lkabi:r]
= [[be:t ṣadi:gi]   ličbi:r]
house   friend-my the-big

The latter is a possible grammatical structure in the dialects but would be more likely to be interpreted in the absence of a context as meaning 'my elderly friend's house', that is:

[be:t [ṣadi:gi ličbi:r]]

(b) especially in phrases where the first element is definite by virtue of an en-clitic pronoun *and* it has an attributive adjective. Thus whereas in the Gulf dia-lects 'her house' can be *be:tha* (synthetic MSA-like construction) or *ilbe:t ma:lha* (analytic), 'her big house' is more likely to be rendered by the analytic *ilbe:t ličbi:r ma:lha* (lit. 'the big house belonging her') than by the MSA-like *be:tha ličbi:r* ('her house the big').

6.1.4 Comparative constructs

The comparative/superlative adjective—there is no distinction in Arabic—that was originally an intensive (see 4.3.2(g)) can be used in noun phrases like an ordinary adjective, when, as expected, it follows and agrees with its noun in gender, number and case. For example:

From *kabi:r fi: ssinn* 'old' (lit. 'big in age'):

msg.   *ʔarrajulu lʔakbaru   fi: ssinn*
            the-man the-bigger   in the-age
            'the oldest man'

fsg.   *ʔalbintu lkubra: fi: ssinn*
            the-girl the-bigger in the-age
            'the oldest girl'

mpl.   *ʔarrija:lu lʔakbaru:na fi: ssinn*
            the-men the-bigger   in the-age
            'the oldest men'

fpl.   *ʔalbana:tu lkubraya:tu fi: ssinn*
            the-girls   the-bigger   in the-age
            'the oldest girls'

But there is a much-used alternative: the comparative adjective becomes the first (annexed) term of what is grammatically a construct phrase, with the noun functioning as the second (amplifying) term. If the entity being referred to is singular, the adjective is usually masculine and the noun is an undefined singular or a defined plural, for example, *ʔakbaru rajulin fi: ssinn* or *ʔakbaru rrija:li fi: ssinn* 'the oldest man', *ʔakbaru bintin fi: ssinn* or *ʔakbaru lbana:ti fi: ssinn* 'the oldest girl' (but note, from a September 2002 newspaper article *kubra: muduni l-janu:b* 'the largest (one) of the southern cities'). There is, however, a difference in the type of semantic reference of these alternatives. Where the amplifying term is in-definite, the phrase is interpreted as classificatory. Thus *ʔakbaru rajulin* implies 'the oldest man' (as opposed to, say 'the oldest woman'). However, *ʔakbaru rrija:li* is identificatory, implying 'the oldest of the men' (is, e.g., Ahmad).

If the entity referred to is plural, the adjective usually remains masculine sin-gular, but number and gender agreement is sometimes encountered, for example, *kubraya:tu lbana:t fi: ssinn* 'the oldest ones of the girls', in order to avoid ambi-guity. In contemporary MSA, this type of construction is the usual means of ex-pressing comparative/superlative attributes, whereas the noun + adjective phrase seems increasingly to be reserved for "frozen" collocations, where a noncom-parative meaning is to the fore. Compare, for example:

CONSTRUCT PHRASE	NOUN + ADJ PHRASE
ʔaʿḍamu ḥarbin	*ʔalḥarbu l'uḍma:*
greater war	the-war the-greater
'the greatest war'	'the Great (i.e., First World) War'
ʔakbaru dduwali	*ʔadduwalu lkubra:*
bigger the-countries	the-countries the-bigger
'the biggest countries'	'the superpowers'
ʔaqṣa: ššarqi	*ʔaššarqu lʔaqṣa:*
farther the-east	the-east the-farther
'the farthest (point) east'	'the Far East'

Note, however, that the *CuCCa:* comparative/intensive form is not freely formable from all masculine *ʔaCCaCu* forms and is absent from the dialects.

Some adjectives cannot form *ʔaCCaC* comparatives, either because their morphological shape contains nonroot consonants (e.g., augmented stem participles) or because their basic form is identical with the comparative form (colors). In such cases a semantically colorless *ʔaCCaC* form such as *ʔakθar* 'more' < *kaθi:r* 'many', *ʔaqall* 'less' < *qali:l* 'few', or *ʔašadd* 'stronger' < *šadi:d* 'strong' is combined with a verbal noun in the indefinite accusative case that specifies the parameter(s) of the comparison (in Arabic *tamyi:z* 'accusative of specification'). Thus from the participle *muntij* 'productive' is derived *ʔakθaru ʔinta:jan* 'more productive' (lit. 'more as regards production'); from *ʔaswad* 'black' *ʔaqallu sawa:dan* 'less black' ('less as regards blackness'). This construction, which is freely incorporated into construct phrases with the mensurative element acting as the annexed term and the specifying verbal noun(s) placed after the amplifying term, has become very common in contemporary journalistic MSA (see 8.3.3). Note, for example, the following phrase, in which there are two parameters of comparison:

> *ʔakθar-u   mudun-i-ha: taqaddum-an   wa ḥayawi:yat-an*
> more-NOM cities-GEN-its advancement-ACC and vitality-ACC
> 'the most advanced and vital of its cities'
> (lit. ' . . . as regards advancement and vitality')

In the dialects, the comparative constructions are basically the same as in MSA. The type of comparison described in the preceding paragraph, however, is expressed in the plain vernacular by a participle rather than a verbal noun, so that, for example, 'more advanced' would be *mutaqaddim ʔakθar* (lit. 'advanced more').[15] The other main difference is that the categories of word to which the *ʔaCCaC* pattern can be applied has expanded well beyond those of MSA/CLA (see Gulf and Damascene examples in 4.3.2(g)).

6.1.5 Mensurative constructs and numbers

With the exception of the decades starting with twenty, and hundreds, thousands, and millions, MSA numbers have both a masculine and a feminine form. The gender agreement rules for combining these with enumerated nouns, which are sometimes singular, sometimes plural, sometimes genitive, and sometimes accusative, are extraordinarily complex: only in the prepared reading aloud of written MSA do many educated Arabs themselves feel confident of getting them correct. Indeed, in a recent study of Egyptian television MSA usage, it was found that one of the small number of features that consistently distinguished "religious" style from "secular" style was that in the latter the use of the much simpler dialectal system of enumeration (or a heavily dialectalized version of the

MSA system) was acceptable (and favored by listeners), although in the former it was not.[16]

'One' and 'two' are relatively straightforward. 'One' *ʔaḥad*, f. *ʔiḥda:* usually functions as the first term of a partitive construct, for example, *ʔaḥadu lkutub* 'one of the books', or as a pronoun meaning 'someone'. In enumeration, for example, 'one book', 'one' is *wa:ḥid*, f. *wa:ḥida*, which conforms morphologically to the adjectival pattern for ordinal numbers. It is used as an attributive in a normal noun-adjective phrase, for example, *kita:b wa:ḥid* 'one book'. 'Two', *ʔiθna:ni*, f. *ʔiθnata:ni*, is used like *wa:ḥid* but after dual nouns. For example:

> *kita:b-a:ni θna:ni*
> book-d   two-m
> 'two books'

Beyond this, the MSA system becomes very complicated. The basis of it is:

(a) 3–10. Syntactically, these are a subvariety of the construct phrase. The enumerator is the first term, and the enumerated noun, which is indefinite, plural, and genitive, is the second. A "gender polarity" rule operates: if the singular form of the enumerated noun is masculine, its enumerator is feminine, and vice versa. For example:

> *ʔarbaʕat-u   ʔawla:d-in*
> four-f-NOM   boys-GEN
> 'four boys'
> *ʔarbaʕ-u   bana:t-in*
> four-m-NOM   girls-GEN
> 'four girls'

If these constructions are made definite (e.g., 'the four boys'), the definite article is prefixed to the second term of the construct:

> *ʔarbaʕat-u   lʔawla:d-i*
> four-f-NOM the-boys-GEN
> 'the four boys'

or an appositional phrase may be used:

> *ʔalʔawla:d-u   lʔarbaʕat-u*
> the-boys-NOM the-four-f-NOM
> 'the four boys'

The first of these two definite constructions can also be construed as a partitive meaning 'four of the boys', however, and where this is the meaning, another construction is sometimes preferred that is more transparent:

> ʔarbaʿat-un   mina lʔawlaːd-i
> four-f-NOM of   the-boys-GEN
> 'four of the boys'

(b) 11–19. The enumerator and its noun are not here in a construct relationship: all elements in the enumerator and its noun are always casemarked as indefinite accusative, and the noun is singular. However, the enumerator, unlike the noun, is not marked with the indefinite marker -*n*. As far as gender agreement is concerned, the "tens" (in 'three ten', 'four ten', etc., which is the way Arabic forms these numbers) agree in gender with the enumerated noun in the normal way, but the "units" show the same phenomenon of "gender polarity" as in 3–10. The exceptions to this are 11 and 12 ('one ten', 'two ten') in which the 'one/two' element, like the 'ten' element, agrees in gender with its noun in the normal way! Syntactically, the enumerated noun functions as an accusative of specification (6.1.4). For example:

> xams-a   ʿašrat-a   bint-an
> five-m-ACC ten-f-ACC girl-ACC
> 'fifteen girls' ('fifteen as regards girl')

> xamsat-a   ʿašar-a   walad-an
> five-f-ACC ten-m-ACC boy-ACC
> 'fifteen boys' ('fifteen as regards boy')

(c) 20–99. The system here is similar to 11–19. The main difference is that the decads, unlike "ten" in 11–19, cannot be marked for gender (whereas the "units" have the same "gender polarity"), and the number is declined regularly according to the three-case system. The actual form of the numbers is on the pattern 'one-and-twenty', 'two-and-twenty', etc. As in 11–19, the enumerated noun, which again is always singular, is syntactically an accusative singular of specification. For example:

> θalaːθ-un   wa ʿišr-uːna   bint-an
> three-m-NOM and twenty-NOM girl-ACC
> 'Twenty-three girls'

> θalaːθat-un   wa ʿišr-uːna   walad-an
> three-f-NOM and twenty-NOM boy-ACC
> 'Twenty-three boys'

(d) 100s. The hundreds form construct phrases with the enumerated noun, as 3–10 do, but they require the enumerated noun to be in the genitive singular, not plural. In expressions such as 'three hundred', the "unit" (here "three") shows gender polarity to "hundred" by always being masculine (*miʔa* 'hundred' is feminine). For example:

> miʔat-u   bint-in
> hundred-NOM girl-GEN
> 'a hundred girls'

> miʔataː   bint-in
> hundred-d-NOM girl-GEN
> 'two hundred girls'

> θalaːθ-u   miʔat-i   bint-in
> three-m-NOM hundred-GEN girl-GEN
> 'three hundred girls'

There are one or two further complications, but from the above examples it can be appreciated that the system for enumeration in CLA/MSA is complex and irregular when compared with how the case and agreement systems work in the rest of the language. The dialectal systems are all much simpler, although there are some differences between Maghrebi (western) dialects and Mashreqi (eastern) dialects on one hand, and, within each of these areas, between Bedouin and urban dialects on the other.

Typical of the number system in the Mashreqi "urban" dialects as a whole is Damascene, in which the CLA system has been reduced and the remaining forms functionally reallocated. Here, the masculine/feminine marking of numbers is preserved only from 3 to 10 and has developed a new and purely functional/pragmatic raison d'être: the feminine forms are used for counting ('one, two, three, four . . . ') and identification ('come in number three, your time is up'), whereas the masculine forms are used for mensuration ('six boys')[17] regardless of the gender of the noun, which, as in CLA/MSA, is always in the plural. Thus 'the number four' is *ʔarbaʿa* (= the equivalent of CLA/MSA feminine form), whereas mensurative 'four' is *ʔarbaʿ* whether it is 'four boys' or 'four girls'. Similarly, from 11 on, the numbers are no longer marked for gender, but, as in 3–10,

a new functional distinction has emerged, whereby for counting and identification, a "short" form of the dialectal numeral is used, for example, *ḥdaʿš* 'eleven', *tnaʿš* 'twelve', but in mensurative phrases one uses the "long" form, for example, *ḥdaʿšar bint* 'eleven girls', *tnaʿšar walad* 'twelve boys'. The numbers from 11 require the noun to be in the singular, as in CLA/MSA, but because of the loss of the case system in all the dialects, the resulting form has become syntactically indistinguishable from the "construct phrases" of 3–10: thus *xams banaːt* 'five girls', *tnaʿsar bint* 'twelve girls'. Two of anything can be expressed in CLA/MSA and the dialects, as we have already noted, by adding a suffix. This is *-eːn* in Damascene and a phonological variant thereof in other dialects. However, Damascene has developed in this area of number morphosyntax, as elsewhere, a functional distinction: as well as *binteːn* 'two girls', one can also, following the 3–10 system, say *tneːn banaːt*, which carries somewhat more emphasis; and a combination of the two systems, *binteːn tneːn* is even more emphatically '*two* girls'.[18] "One" is simply an adjective *waːḥid* 'one', f. *waḥda*, placed after the noun, as in MSA. But in the dialects, *waːḥid* frequently carries no suggestion of mensuration (i.e., of 'one' as opposed to 'more than one') but seems to have become functionally equivalent to an indefinite article, so that the deictic use of the article in *irriʒʒaːl* 'the man' (i.e., 'the one previously referred to') contrasts with the nondeictic *riʒʒaːl waːḥid* 'a man' (i.e., 'some man or other'). Like the development of the periphrastic construct, this lexicalization of the definite/indefinite distinction is a further example of analytic drift in the evolution of the dialects.[19]

The main difference between the "urban" type of system exemplified by the dialect of Damascus and the Bedouin dialect type is the general preservation in the latter of "gender polarity" in mensurative constructs involving the numbers 3–10. In this, as in many other aspects of morphosyntax, the Bedouin dialects are more conservative and more closely resemble the CLA system.[20]

6.2 The verb phrase

We shall consider here the relationship of form and meaning in the verb, specifically the choice of s-stem/p-stem/participial forms of the verb in relation to various semantic properties of the verb and the sentence:

(a) aspect and factuality
(b) mood and modality
(c) time and tense

Nonverbal phrases that behave syntactically as if they were verbs and serve a verbal function will subsequently be considered.

The basic structural elements of the verb phrase in CLA/MSA are the verb itself—the only obligatory element—the auxiliary verb (*KWN* 'to be'), the modal/aspectual particles (*qad* and its variant *laqad*, and *sa-* and its variant *sawfa*), the negative particles (*laː, maː, lam, lan, laysa*), and the bound pronominal object enclitic morphemes. The canonical shape of the verb phrase can be summarized as:

(AUX) (PTCLE) (NEG) VERB (PRON ENC)

where parentheses indicate that an element need not necessarily be present. The verb may be a p-stem (of which there are several morphological subsets representing different moods), an s-stem, or an active participle. The selection of any one of these verb types restricts (in some cases prevents) the selection of specific elements at the AUX, PTCLE, and NEG points of structure. Further selectional restrictions operate between some combinations of elements representing AUX, PTCLE, and NEG, their functions being mutually exclusive.

The overall structure of the verb phrase in the dialects is similar, although the elements available for selection at every point of structure are different, as is, in some dialects, their ordering within the phrase (e.g., a discontinuous NEG . . . NEG element). Different selectional restrictions also apply. In some dialects (Cairene, notably), an extra element that we might term prepositional complement, which covers certain preposition + indirect object structures, should be included as part of the verb phrase on phonological grounds.

6.2.1 Aspect and factuality

The s-stem/p-stem distinction was historically not one of tense but of verbal aspect—although, synchronically, as we shall see, it is evolving in both MSA and the dialects toward a tense system.[21] In all varieties of modern Arabic, the s-stem/p-stem distinguishes actions and states conceived of as completed/factual (s-stem) from ones conceived of as noncompleted/ongoing/notional (p-stem).[22] The usual negative correlate of the s-stem in MSA, *lam* + base set p-stem, behaves to all intents and purposes like the negative of the s-stem itself, that is, *lam yaḍrib* is synonymous with *maː ḍaraba* 'he hasn't/did not hit', as the negative of *ḍaraba* 'he has hit/hit'. Except where indicated, any remarks concerning the MSA s-stem also apply to its normal negative, *lam* + base set p-stem.

Although it is true that actions that occurred and were completed at a time anterior to the time of utterance normally require the s-stem, this seems to be primarily because of their completed aspect and only incidentally because they

occurred, relative to the time of utterance, in the past. That "pastness" is not central to the meaning of the s-stem is shown by its use in a range of nonnarrative, nonpast contexts. Among the commonest are the following:

(a) In all conditional clauses, even those of the "open" type (e.g., 'If he goes' in 'If he goes, I'll go too'), the action/state in the main clause is envisaged as dependent on the prior occurrence of another action/state (or, more precisely, the completion of its onset). For this reason, an s-stem verb is used to express the condition even though the action to which it refers is in the future, relative to the time of utterance. The s-stem is also frequently used in the related category of timeless quasiconditionals of the 'whatever/whoever/whenever/however' type (see 7.2.2.4 for conditional clauses in general). For example:

> hum ʔumma wa:ħida maħma: xtalafat   ʔajna:suhum
> they nation one   whatever differed-3fsg races-their
> 'They (sc. the Arabs) are one nation however different their (component) races are'

These observations apply to MSA and also to the Bedouin-type dialects, although they are becoming less true of the urban dialects, in which the s-stem/p-stem distinction has evolved further toward a true tense system, and p-stem verbs (= "nonpast") are often used in "open" conditionals.

(b) The s-stem of verbs of emotion and cognition is often used with little or no implication of pastness. In CLA/MSA and the Bedouin dialects, these verbs retain a punctual/dynamic as well as a durative/static value, for example, ʕarafa (static value 'to know'/dynamic 'to realize'), fahima (static value 'to have an understanding'/dynamic 'to grasp'), fariħa (static value 'to be happy'/dynamic 'to rejoice'). The s-stem form is used not only when the sense is dynamic and the action conceived of as having (just) occurred, but also when the sense is of an ongoing state. Compare, for example, the schoolboy's dialectal reply to his teacher fihimt! (s-stem, dynamic value) with English present tense '(Now) I understand!', but also ʕaraft ha:ða rrajil (s-stem, static value) with 'I know (= 'have recognized') this man', kariht tanis iṭṭawla (s-stem, static value) with 'I hate (= 'have come to hate') table tennis'. In the urban dialects, and "educated" speech more generally, static semantic value in verbs of cognition and emotion is now more commonly expressed by a p-stem verb or a participle (see below).

(c) Performative verbs like wa:faqa 'to agree' and qabila 'to accept' are also commonly used in the s-stem, as in the Jordanian passenger's gabilt 'I accept' in response to a taxi driver's suggested fare at the start of a journey. Here a personal commitment on the speaker's part has been entered into (completed event), but no "pastness" is involved, because the commitment and its expression are to all intents and purposes simultaneous.

(d) Optatives involving exhortations to God exist in great variety and profusion for every conceivable kind of social occasion, situation, and phatic need. They are extremely commonly used in all forms of spoken Arabic, especially perhaps that of the more traditional, nonwesternized sections of Arab communities. In written CLA/MSA, s-stems are invariably used for this, although in the dialects alternatives with a p-stem are common too. Again, the Bedouin dialects preserve the original s-stem system more consistently. For example (Bahrain):

> ħayya:k   allah
> gave-life-3msg God
> 'May God keep you alive!' (general greeting/leave-taking)

> bayyaḍ   allah wajhak
> whiten-3msg God face-your
> 'God whiten your face!' (expression of thanks)

In all varieties of Arabic, the p-stem form of the verb marks the noncompletedness of an action or state. In MSA, it has, like the s-stem, no particular time value out of context. In given contexts, its aspectual value and time coloring may be any of the following:

(a) to describe general, timeless truths ('Water *boils* at 100°C', 'I *like* ice-cream')
(b) habits and other iterative, nonpunctual processes, past or present ('I *get up* at six every day', 'I *used to watch* [the soccer team] Arsenal when I *was living* in London')
(c) duratives ('He *is/has been/was writing* a letter')
(d) circumstantial clauses ('We left him *digging* the garden')
(e) statements about the future (with the optional future particle *sa-* or *sawfa*) ('We *will arrive* tomorrow')
(f) subordinate noun clauses (the "a-set" of the p-stem—see 3.4) in which the action expressed by the dependent verb is in the realm of the notional, possible, or desirable, not the factual ('He wanted her *to go*', 'I thought she *would leave*').

A p-stem verb would also, in the dialects, be one of the possibilities—in some examples the only one—for rendering the meanings expressed by the italicized verb forms above. But in virtually all the dialects there is a system of aspectual and modal prefixes (differing slightly from dialect to dialect) for particularizing the different aspectual and time values covered by the MSA prefixless p-stem. These developments are discussed in detail in 6.2.2–3. In MSA, by contrast, in the absence of other elements that make the relative timing of actions

clear—time adverbials, the optional future markers *sa-* or *sawfa*, or dependent clauses (such as conditionals)—the only variety of p-stem or s-stem verb form that is intrinsically time marked is that of the s-stem of verbs with dynamic value, for example, *katabtu risa:latan* 'I wrote a letter'. Because the verb translated by 'wrote' (a) denotes a dynamic action and (b) is in the s-stem, indicating completion, it is necessarily understood as having occurred in the past relative to the time of utterance.[23]

Besides the s- and p-stems, the participles have a verbal function (see 4.2) in all varieties of Arabic, although a particularly important and evolved one in the dialects. In MSA, the active participle, although morphologically a species of noun/adjective and hence carrying the case marking of a noun, may, like a verb, govern an object. Like the finite verb, it has no intrinsic time marking, taking its temporal coloration from context. However, the basic semantic value of the verb, dynamic or static, is relevant to the contextual interpretation: in MSA, the active participle of dynamic verbs often has a future meaning relative to the time of utterance, whereas that of static verbs usually indicates a situation coexisting with the time of utterance. For example:

'dynamic' verb:
?alḥarb wa:qiʿa   la: maḥa:la
the-war happen-A.PART   no avoiding
'War will break out, inevitably'

'static' verb:
?inni: muṣaddiquka
PTCLE-I   believe-A.PART-you
'I believe you'

In both of these examples,[24] a p-stem verb could just as well have been used (in the first with the future prefix *sa-*), and in the second, where the verb is (in J. R. Searle's terms) a "representative" speech act, an s-stem could also be substituted without affecting the meaning. But in the absence of a time marker such as an adverb, only the participle of the first, "dynamic" verb admits, out of context, of a present-time *and* a future-time interpretation.

In the dialects, the active participle, unlike in MSA, frequently has a perfect meaning, indicating the inception of a state or action in the past that has continued (or whose consequences continue) up to and including the time of utterance. So, the Cairene dialectal sentence:

ana miš ʿa:rif   irragil   da
I   not know-A.PART the-man that
'I don't know that man'

should be construed as meaning 'I have at no time in the past made that man's acquaintance, and (therefore) I don't know him now'. In all dialects, however, the semantics of the verb—"stative" (of which a subgroup are verbs of perception and cognition) and "dynamic" (of which a subgroup are verbs of motion) play a key role in determining whether perfective, concomitant, or (in some cases) future meaning is implied in particular cases by the use of a participle. The participle is usual in the dialects for verbs of perception and cognition when used as a description of a concomitant state, as in the example above. By contrast, the active participle of verbs of motion is commonly used in most dialects to describe concomitant (or with the requisite adverb such as *bukra* 'tomorrow' future) action, rather than perfective: one says, for example, *we:n ra:yiḥ* (or a local equivalent) to an acquaintance one bumps into in the street to mean 'Where are you going?', not 'Where have you been?'. Further, in many dialects, there is an opposition between the s-stem and the participle of "dynamic" verbs such that the s-stem is used where an action is being viewed as a completed episode and the participle where the state resulting from the action is the focus. So, for example, the first person s-stem of 'to eat', Moroccan *kli:t*, Baghdadi *?akalit*, Bahraini *kale:t* 'I ate/have eaten' is a neutral statement of fact, such as might occur in the context of a narrative; the corresponding active participle forms *wa:kil* (Morocco), *ma:kil* (Baghdad, Bahrain), which have the same translation, would be an appropriately polite reply to an invitation to lunch when the speaker had already eaten and was no longer hungry.[25] In most dialects, in the case of most "dynamic" verbs, there is a regular three-way p-stem/s-stem/active participle set of oppositions, such that, for example:

huwwa mrattib   ilhudu:m
he   tidy-A.PART   the-clothes
'He has tidied up the clothes'

would be typically used by someone (in this case a Cairene) returning home and finding evidence that the cleaner has been. Here, the use of the participle contrasts with both the p-stem verb *birattib*[26] ('He's tidying the clothes' (implication: and I can see/hear him doing it)) and the s-stem *rattab* ('He tidied the clothes'), which would typically occur in the context of a report or narrative:

'He tidied the clothes, swept the floor, and left at midday'. There is, however, a further important semantic distinction between participle and s-stem use, essentially modal in nature, that follows from this and that relates to the degree of speaker commitment to the proposition made.[27] Consider the following:

(1)   *za:rhum ḥa:zim ʔams u xabbarhum ʿan kull ma: jara*
          s-stem   s-stem
(2)   *za:yirhum ḥa:zim ʔams u mxabbirhum ʿan kull ma: jara*
          A.PART   A.PART

In both cases, the English translation is 'Hazim visited them yesterday and informed them of all that happened'. Common to both is that the speech situation is one of narration and that X is telling Y about something that happened. The difference is that whereas (1) is straightforward narration of fact, the use of the s-stem forms guaranteeing the reliability of the assertion (possibly, for example, because X personally witnessed the events described), the use of the participles in (2) suggests reservations: X may not have witnessed the events, or know for certain that they occurred, and cannot therefore be held responsible for their veracity. Of forty-eight Jordanians asked to indicate which of

(1) *sami:r ḍarab muni:r* (s-stem) ⎫
                                    ⎪
and   ⎬   'Samir hit Munir'
                                    ⎪
(2) *sami:r ḍa:rib muni:r* (A.PART) ⎭

suggests that the speaker was an eyewitness of the event, forty-six chose (1) and only two (2). Despite some minor differences in participial usage, this modal distinction of propositional factuality carried by the s-stem and participle is general throughout Egypt, Syria, and most of the Levant. This use of the participle has been considered here for the sake of clarity of exposition, although it properly forms part of modality (6.2.2 below)—a point that underlines the fact that formal categories in Arabic frequently perform a wide range of functions. These are syntactic and semantic in the case of the active participle, which can function as a noun, adjective, or verb and in the latter case have a variety of both aspectual and modal uses, depending on context.

Some aspectual distinctions are expressed in Arabic by distinctions in the morphosemantic system of the verb. We noted in chapter 3 that the value of some of the augmented patterns of the verb, in relation to the simple pattern,

could be considered, at least in part, aspectual in their modern usage. Thus Pattern VI often conveys an idea of an action occurring over a period of time or nonpunctually: compare *rajaʿa* 'to come/go back' versus *tara:jaʿa* 'to retreat (e.g., an army)', 'to go out (sea)'; *saqaṭa* 'to fall' (e.g., the apple/France fell) versus *tasa:qaṭa* 'to fall' (e.g., tears, snow, rain fell).

A number of other aspectual distinctions are lexicalized and idiomaticized in both MSA and the dialects, although differently. The idioms consist of verb + verb strings, in which both verbs are inflected in the normal way and cover such aspects as inceptive ('to start to . . . '), ingressive ('to be on the point of . . . '), iterative ('to keep on . . . '), durative ('to be in the process of . . . '), terminative ('to stop doing something'), etc. The first "aspectual" verb of the string usually has a literal meaning from which its aspectual force is clearly derived, for example, (Gulf) *ga:m* 'to get up' (to 'up and . . . ') = 'to start to . . . ', *garub* 'to be near to' = 'to be on the point of', *ga:ʿid* 'sitting' (A.PART) = 'to be in the process of . . . '. A number of other types of meaning, although not strictly aspectual, are also expressed by verb strings of the same asyndetic structure: in MSA, for example, 'to still do something', 'to no longer do something', 'to nearly do something', 'to scarcely be able to do something'. In the dialects, certain verbs and verbal elements have developed fixed idiomatic usages in which they are used in uninflected form to express aspectual meanings derived from their literal meanings. For example, in Cairo *baʔa* 'to remain' and in Jordan, Arabia, and the Gulf *ṣa:r* 'to become' can be used to express durative perfective meaning. For example, (Gulf):

ṣa:r   li: sa:ʿa u ana wa:gif   ihni
became-3msg to-me hour and I   stand-A.PART here
'I've been standing here an hour'

6.2.2 Mood and modality

"Mood" usually refers to the inflectional marking of verb paradigms, which are often labeled indicative, subjunctive, etc. That is, it refers to a set of morphological markers, the use of which is required by syntactic rule. "Modality" is often reserved for semantic distinctions covering such categories as ability, possibility, obligation, etc., and further subdivisible into "deontic" and "epistemic" modality. In practice, as in MSA, mood and modality are intimately bound. Mood inflection is always marked in the p-stem and is frequently a morphological correlate of modal meaning of one kind or another, expressed lexically elsewhere in the sentence. There is no mood marking in the s-stem, although a modal "slant" can be given by lexical elements, for example, in epistemic state-

ments about completed events, such as in the Arabic equivalent of 'He must have arrived yesterday'.

We noted in chapter 3 that in MSA there are three subsets of the p-stem verb—the base set ("jussive"), a-set ("subjunctive"), and u-set ("indicative"):

(a) The u-set's main use is where a p-stem verb occurs in a declarative or interrogative main clause in which the proposition stated or questioned is presented as factual, whatever its aspectual value and time reference—past, present, or, after the future particles *sa-* and *sawfa*, future. It is also used after the particle *qad*, 'it may be that . . .', for example, *qad yaku:nu lissan* 'he may be a thief', which expresses an epistemic possibility.

(b) The main uses of the a-set are in (i) syndetic subordinate noun clauses where the action is viewed as in the realm of the notional or possible but, at any rate, nonfactual. Typically, such clauses are governed by verbal or nonverbal phrases that give a modal slant on the meaning of the governed verb: 'it is possible/probable/necessary, etc., that . . .', 'it must be/there is no doubt that . . .', etc.; or by verbs of desire, perception, or cognition: 'to want/think/believe/hope/feel, etc., that . . .'; and (ii) other types of subordinate clauses, such as those of purpose, result, and time, where the action is in the future relative to the main clause action; and (iii) future volitional negatives with *lan*, for example, *lan nastaslima* 'we will never surrender'. The common factor in (i)–(iii) is that the proposition is presented as in the realm of the nonfactual at the time of utterance. However, subordinate clauses where the p-stem verb refers to a *fact* require the u-set, not the a-set. Thus the MSA equivalent of 'he is writing' in 'It's probable that he is writing a book' is an a-stem verb:

> *mina lmuḥtamali   ?an ya-ktub-a   kita:ban*
> of   the-probable that 3msg-write-SUBJ book

But if the main clause presents 'his (being in the process of) writing a book' as a known fact, for example, 'I know he is writing a book', 'it pleases me that he is writing a book', the complementizer *?anna* 'that' is used, requiring a noun- or pronoun-initial subordinate clause and a u-set verb:

> *?a'rifu   ?annahu ya-ktub-u   kita:ban*
> 1sg-know that-he 3msg-write-INDIC book
> 'I know that he's writing a book'

> *yasurruni: ?annahu ya-ktub-u   kita:ban*
> 3msg-please-me that-he 3msg-write-INDIC book
> 'It pleases me that he's writing a book'

Of course, if 'his writing a book' is presented as being potentially or actually an accomplished fact, the subordinate clause verb is s-stem and may have either VSO or SVO word order, preceded by the appropriate complementizer: *?an* for VSO, *?anna* for SVO:

> *mina lmuḥtamali   ?an kataba   kita:ban*
> of   the-probable that wrote-3msg book

or

> *mina lmuḥtamali   ?annahu kataba   kita:ban*
> of   the-probable that-he wrote-3msg book
> 'It's probable that he wrote a book'
> *?a'rifu   ?an kataba   kita:ban*
> 1sg-know that wrote-3msg book

or

> *?a'rifu   ?annahu kataba   kita:ban*
> 1sg-know that-he wrote-3msg book
> 'I know that he wrote a book'

(c) As we noted earlier, one of the main uses of the base set of the p-stem (the "jussive") in MSA has nothing to do with uncompleted action: with *lam* 'not', it is the commonest type of verb phrase for expressing the negative of completed action, its affirmative correlate being the s-stem verb. Its specifically modal uses in MSA are in (i) optatives of the 'may we/he/ she . . . !' type, (ii) with *la:* in the negative imperative 'don't . . . !', and (iii) in conditional clauses instead of an s-stem (or indeed in the main clause in conditional sentences). Again, what unites these usages (i)–(iii) is the futurity and/or nonfactuality of the action they refer to, with respect to the time of utterance.

In the dialects, the three-way p-stem/s-stem/active participle distinction is basically aspectual in character; morphologically marked varieties of the p-stem verb carry distinctions that are chiefly ones of mood. This is evidently an oversimplification: we have already observed that the active participle can take on a modal meaning in opposition to the s-stem in the appropriate speech context, and subvarieties of the p-stem can also mark aspect rather than mood in some syntactic frames, notably in the governed verb in verb + verb strings noted at the end of 6.2.1 and exemplified further below. But as an overall description of how types of verbal meaning are mapped onto form in the dialects, the generalization seems valid.

The formal markers of mood in the dialects are different in that the u-set/ a-set/base set inflectional endings of CLA collapsed historically into a single dialectal paradigm that superficially resembles the CLA/MSA base set (the "jussive"). However, something akin to the CLA/MSA mood distinctions has evolved through the development of a set of p-stem proclitics attached directly to the verb ("attached" in the sense that they are part of the same phonological word as the verb for purposes of stress, vowel deletion, and other suprasegmentally determined processes). Although there are sometimes considerable differences of detail from one dialect to another, this means of mood marking is now virtually universal in the urban dialects in all parts of the Arab world, although in the Bedouin dialects, particularly of the Arabian peninsula, it is less fully evolved. Like the particles used in the periphrastic noun phrase, these proclitics are historically decayed lexical or other elements,[28] although in some dialects, the process of decay has not reached the point where their lexical origin and force has been totally lost.[29]

Modal meanings of ability, possibility, obligation, wanting, intending, etc., are, as in MSA, for the most part carried by main clause verbs and other freestanding lexical elements. Verbs governed by such modal elements, for example, 'you must *go*', 'he couldn't *answer*', 'I may *arrive* earlier', and 'I would like *to leave*' carry in their dialectal Arabic p-stem equivalents a proclitic mood marking. This is largely predictable—in all the cases cited, for example, it would be zero (= "subjunctive"). But this predictability is not total, as we shall see below: in embedded clauses, mood marking can be used in the absence of lexical modal elements to indicate modal meanings.

In many dialects of the eastern Mediterranean, the proclitic mood-marking system consists of a three-way distinction between the *b(i)-* prefix, the *ḥa-* (or *raḥ-*) prefix, and zero prefix. The distribution of these three dialectally shared forms across various types of meaning differs (mainly between Egypt and the Levant) in some particulars, but the functioning of the system is broadly similar. The examples cited below are from the dialect of Cairo.

In main clauses, the *b(i)-* prefix most often equates to the MSA u-set ("indicative") of the p-stem—that is, it marks factual statements containing verbs of varied aspect and is also used in predictions about the future. For example:

byiʔra
b-3msg-read
'He reads/can read' (= ability)
byiʔra   kti:r
b-3msg-read much

'He reads a lot' (= habit)
byiʔra   kta:b
b-3msg-read book
'He's reading a book' (= continuous action)
byiʔra   lkita:b   bukra
b-3msg-read the-book tomorrow
'He'll read the book tomorrow' (= prediction about the future)

The zero-marked p-stem also occurs in main clauses, when it provides an essentially modal set of contrasts with the *b(i)-* prefix verb in statements and questions. Compare the following pairs:

(a) statements:

byiʔra   lkita:b   bukra
b-3msg-read the-book tomorrow
'He'll read the book tomorrow' (= factual prediction)
yiʔra   lkita:b   bukra
3msg-read the-book tomorrow
'He is to read the book tomorrow' (= instruction)

(b) questions:

byiʔra   e:h
b-3msg-read what
'What is he reading?' (= inquiry as to fact)
yiʔra   e:h
3msg-read what
'What is he to read?' (= inquiry as to obligation)
bitišrab   ša:y
b-2msg-drink tea
'Are you drinking tea?' (= inquiry as to current activity)
or 'Do you drink tea?' (= inquiry as to habit)
tišrab   ša:y
2msg-drink tea
'Would you like to drink tea?' (= inquiry as to volition)[30]

The zero-marked p-stem is also used in prohibitions and in quasi-imperatives of various kinds: 'Mind you don't . . . !', 'It would be better if we . . . '. In positive commands a variant of it is used: the p-stem without the second person prefix:[31]

(c) imperatives:

?i?ra   lkita:b   da
read-msg the-book this
'Read (m) this book!'
ma: ti?ra:š   ilkita:b   da
not 2msg-read-NEG the-book this
'Don't read (m) this book!'
?iw'a   ti?ra   lkita:b   da
be-careful-2mg 2msg-read the-book this
'Be careful not to read this book!'
ma: yihimmiš,   yi?ra   lkita:b   da
not 3msg-concern-NEG 3msg-read the-book this
'Never mind, let him read this book!'

Other major uses of the zero-marked p-stem in main clauses are with verbs of cognition, perception, and feeling, and with performative verbs (with the exception of 'say') such as 'advise', 'warn', 'thank', etc., although there is some interspeaker variation here between the zero- and *b(i)*-marked forms in certain registers of speech (see chapter 9).[32] There is a similar fluctuation between the zero and *b(i)*–p-stem in sentences of "timeless" relevance such as "Water boils at 100°C."

The *ḥa*-prefix p-stem is used in Cairene (and other dialects, although its phonological shape differs) for what might be described[33] as "proximate intent" and involves a stronger commitment on the part of the speaker (or the imputing of one to a third party) than is the case with the "future" use of the *b(i)*- prefix. So, for example, in contrast to *byi?ra lkita:b bukra* 'He'll read the book tomorrow', which is no more than an uncommitted prediction, *ḥayi?ra lkita:b bukra* 'He's going to read the book tomorrow' indicates somewhat more certainty that this is the subject's real intention. Similarly, in the interrogative, *ḥatišrab ša:y* means 'Are you going (now) to drink tea?' or even 'Are you on your way to drink tea?' An even stronger degree of commitment is betokened by the use of *'a:yiz* (Cairo), *bidd-* (Levant), which are fully lexicalized modals of volition or necessity and govern a zero-marked p-stem: *huwa 'a:yiz yi?ra lkita:b* 'He wants (or needs) to read the book' (see below).

This last example brings us to the expression of mood in embedded clauses of various kinds. Recall (6.2.1) that, in both MSA and the dialects, some types of *aspectual* distinction are lexicalized in verb + verb strings where the first verb expresses aspect, and the second, embedded verb, which describes the action or state, is marked as "subjunctive" (a-set p-stem in MSA, zero-marked p-stem

in the dialects), for example, (Cairene examples) *ṣa:r yi?ra* (inceptive) 'he began to read', *tann yi?ra* (continuous/habitual) 'he carried on reading'. Recall also that verbs in clauses that are the object of the main clause verb or occur in various types of dependent adverbial clause (result, purpose, time, concession, etc.) and thus denote actions and states that are incomplete or in the future relative to the time of the main verb (and are hence nonfactual) also require an a-set p-stem ("subjunctive") in MSA. A parallel form–function correspondence occurs in the dialects, which require the dialectal "subjunctive," the zero-marked p-stem, in such sentences. For example, (Cairene examples):

MAIN VERB   —OBJECT CLAUSE—
na:wi   yi?ra   lkita:b
intend-A.PART 3msg-read the-book
'He is intending to read the book'
ḥa:wil   yi?ra   lkita:b
tried-3msg   3msg-read the-book
'He tried to read the book'
xa:f   yi?ra   lkita:b
feared-3msg   3msg-read the-book
'He was afraid to read the book'
MAIN VERB   —PURPOSE CLAUSE—
ga   badri ḥatta yi?ra   lkita:b
came-3msg early so that 3msg-read the-book
'He came early so that he could read the book'
MAIN VERB   —TIME CLAUSE—
ḥayiddilhum   xabar ?ablima yi?ra   lkita:b
ḥa-3msg-give-to-them news before 3msg-read the-book
'He'll tell them before he reads the book'

In all the above cases there is no choice as to the mood marking of the p-stem verb in the object or adverbial clause. But where the p-stem verb is in a noun clause that is the grammatical subject of the main verb, it can be marked for mood exactly as a p-stem verb in a main clause can, with appropriate differences in the interpretation of the meaning. For example:

MAIN VERB —NOUN CLAUSE—
(a) *b(i)*- proclitic:
yis'idni   ?innuh byi?ra   kul yo:m
3msg-please-me that-he b-3msg-read every day
'It pleases me that he reads every day' (= indicative, factual)

(b) zero proclitic:

yis'idni   ?innuh yi?ra   kul   yo:m
3msg-please-me that-he 3msg-read every day
'It would please me if he would read every day' (= subjunctive, nonfactual)

(c) *ḥa-* proclitic:

yis'idni   ?innuh ḥayi?ra   kul   yo:m
3msg-please-me that-he ḥa-3msg-read every day
'It pleases me that he's going to read every day' (= proximate intent)

If a p-stem verb occurs in a noun clause that is the subject (or object) of a modal element, its mood marking in the dialects depends on the type of modal meaning intended. Generally speaking, where the meaning is not epistemic (i.e., not inference- or knowledge-based) but is a straightforward statement of non-factual possibility, permission, ability, obligation, etc., the p-stem verb in the subject or object noun clause is zero-marked ("subjunctive"). For example:

MODAL —NOUN CLAUSE (S or O)—
la:zim   yi?ra   lkita:b   da
necessary 3msg-read the-book this
'He must read this book' (= obligation)

yimkin   yi?ra   lkita:b   da
3msg-be-possible 3msg-read the-book this
'It's possible for him to read this book' (= possibility)

yigu:z   luh   yi?ra   lkita:b   da
3msg-be-permitted to-him 3msg-read the-book this
'He may read this book' (= permission)

yi?dar   yi?ra   lkita:b   da
3msg-can 3msg-read the-book this
'He can read this book' (= ability/capacity)

But where the modal element is used epistemically, as, for example, *la:zim* frequently is, its semantic scope covers the whole of the proposition expressed by the noun clause that is its grammatical subject. In this case, depending on the sense, the *b(i)-* or *ḥa-* proclitic can be used (as, for that matter, can an s-stem). Compare *la:zim yi?ra lkita:b* 'He must read the book' (deontic meaning, zero proclitic) with:

la:zim   byi?ra   lkita:b   da
necessary *b*-3msg-read the-book this
'He must be reading this book' (= 'It must be the case that he is . . . ')

la:zim   ḥayi?ra   l-kita:b   da
necessary *ḥa*-3msg-read the-book this
'He must be going to read this book' (= 'It must be the case that he is going to . . . ')

la:zim   ?ara   lkita:b   da
necessary read-3msg the-book this
'He must have read this book' (= 'It must be the case that he has . . . ')

Among other minor modal uses of the p-stem in the dialects, two are worth mentioning here, both of which are connected to the notion of potentiality, although in different ways. When predicated of an indefinite antecedent, the p-stem verb may acquire a contextual sense of potentiality. This usage is particularly frequent in yes/no questions and negatives. In dialects that have an indicative/subjunctive contrast, such p-stems are treated like others that denote nonfactual or unrealized actions and states—they require the zero-marked subjunctive. For example:

Amman: *fi: ba:s nirkab   'ale:h ballah*
             existent bus 1pl-ride on-it please
             'Is there a bus we can take, at all?'

Cairene: *ma: 'andi:š   ḥa:ga ?adifha ?ila ma: ?ult*
             not with-me-neg thing 1sg-add to what said-2msg
             'I don't have anything to add to what you said'

We noted in chapter 3 that Pattern VII of the verb, *nC1C2C3*, has an inherently passive meaning. Some dialects have this pattern as the regular pattern for passivizing, whereas others have *tC1C2C3*, and yet others have both: thus Cairene *?itšarab, ?inšarab*, Gulf *inšarab* 'it was drunk'. The *b(i)*–p-stem of this verb pattern is often used in a "potential" modal sense of 'can be . . . ed'/'to be . . . ed'. For example:

Cairene: *ilbe:t   da   ma: byinna:m   fi:h*
             the-house this not *b*-3msg-be-slept in-it
             'This house can't be slept in' (e.g., it's too noisy)

Damascene: *ilbalad   btinša:f   min ho:n*
             the-town *b*-3msg-be-seen from here
             'The town is visible/can be seen from here'

Bahraini: *halma:y   ma: yinširib*[34]
             this-the-water not 3msg-be-drunk
             'This water is undrinkable/can't be drunk'

6.2.3 Time and tense

We have already made the point that historically the p-stem/s-stem distinction was basically one of aspect, not tense, if it is now evolving toward a tense system. Other morphological oppositions in the dialectal verb, like that between the active participle and the s-stem, may be aspectual or modal depending on the semantics of the verb and the pragmatics of situation. Within the p-stem, morphological distinctions are essentially modal. How then are distinctions of time made?

The main means of marking time is the auxiliary verb *ka:na/yaku:nu* 'to be'. We noted earlier that, out of context, unadorned p-stem verbs were inherently non–time marked and that only the s-stem of semantically dynamic verbs was necessarily to be construed as referring to the past, relative to the time of utterance. The effect of placing the appropriate person[35] of the s-stem of *ka:na* before any verb is thus to overtly mark the action/state it refers to as past, relative to the time of other actions/states not so marked. Compare the following MSA examples:

verbless	auxiliary
Pana mudarris-un	*kuntu mudarris-an*[36]
I teacher-NOM	was-1sg teacher-ACC
'I am a teacher'	'I was a teacher'
huwa fi: ljayš	*ka:na fi: ljayš*
he in the-army	was-3msg in the-army
'He is in the army'	'He was in the army'

p-stem

yaPkulu llaḥm	*ka:na yaPkulu . . .*
3msg-eat the-meat	was-3msg 3msg-eat . . .
'He eats/is eating/will eat the meat'	'He used to eat/was eating/ was going to eat . . . '

s-stem ("dynamic" verb)

Pakalta llaḥm	*kunta Pakalta . . .*
ate-2msg the-meat	were-2msg ate-2msg . . .
'You ate/have eaten the meat'	'You had eaten . . . '

s-stem ("stative" verb)

ta'ibtu kaθi:ran	*kuntu ta'ibtu kaθi:ran*
got-tired-1sg much	was-1sg got-tired-1sg much

'I am/have got very tired' participle	'I was/had got very tired'
Pattaḥqi:qa:tu ja:riyatun	*ka:nati ttaḥqi:qa:tu ja:riyatan*
the-inquiries run-A.PART-NOM	were-3fsg the-inquiries run-A.PART-ACC
'Inquiries are continuing'	'Inquiries were continuing'

ka:na, and its negative counterpart *lam yakun*, has an anteriorizing effect on any verb to which it is preposed, whatever its aspectual value. Thus (future) *sayaPkulu llaḥm* 'he will/is going to eat the meat', *ka:na sayaPkulu llaḥm* 'he would/was going to eat the meat'; (inceptive) *ja'ala yaPkulu . . .* 'he started to eat', *ka:na ja'ala yaPkulu . . .* 'he had started to eat'; (iterative) *yaðallu yaḥfiru . . .* 'he keeps on digging', *ka:na yaðallu yaḥfiru . . .* 'he would keep on digging', etc. As we noted earlier, some "mental state" verbs may be used with either a punctual-dynamic or durative-static value: *'arafa* 'he grasped, got to know' or 'he knew, was in the know about'. In MSA, it is noticeable that, in past-time reference, the durative-static value of such verbs is now often expressed by *ka:na* + p-stem rather than the simple s-stem—for example, there is a tendency for *'araftu* to be used in the sense of 'I realized, got to know (that) . . .' and *kuntu Pa'rifu* in the sense of 'I was (already) aware of the fact (that) . . .' The same is true of stative verbs: *kuntu farḥa:na* 'I felt happy' rather than the s-stem *fariḥtu*, which is reserved for the dynamic value 'I rejoiced'. This tendency is possibly caused by the influence of the dialects (and European languages) in which the static value of such verbs is expressed by a p-stem or a participle.

The particle *qad*, which we have already seen can be preposed to the p-stem verb to give a modal meaning of possibility, can also be placed before the s-stem verb, when it also has time-related functions. Before the s-stem, *qad*, or its more emphatic variant *laqad*, gives a sense of proximate past, for example, *qad Pakaltu llaḥm* 'I have (just) eaten the meat'. More generally, *qad* is used to denote actions that occurred in the past but whose consequences are of pragmatic relevance to the current discourse topic, for example, *kasara rijlahu* 'he broke his leg' (past event stating) but *qad kasara rijlahu* 'he has broken his leg/his leg is broken' (past event with present relevance, e.g., reason for his not being able to play next week); *raPaytu lPahra:m* 'I saw the pyramids' (past event, narrative style) but *qad raPaytu lPahra:m* 'I have seen the pyramids (in my time)'—the "experiential" past.[37] *qad* is also used with the s-stem in the related function of describing the anterior attendant circumstances of events described in the main clause, for instance in a narrative. For example:[38]

yaðharu   mašli:niya: wa qad ḥalaqa   liḥyatahu
3msg-appear Mashliniya and *qad* shaved-3msg beard-his
'Mishliniya appears, having shaved his beard'

In this last type of "circumstantial" function, *qad* is often used in conjunction with *ka:na* (placed between *ka:na* and the s-stem verb) to refer to an event anterior to the main past time (s-stem) narrative and explain or otherwise amplify it, rather like an English pluperfect (past in the past). The following translated passage[39] illustrates the "backgrounding" function of *ka:na qad*: 'In those years I got to know (s-stem) Yaqoub Sarruf, who *was over sixty* (*ka:na qad taja:waza ssitti:n*) and whose scientific bent *had come to dominate him* (*ka:nat qad ṭaġat ʿalayhi*). He did not value (*ka:na* + p-stem) literature or philosophy . . .'. The relative clauses both provide parenthetical explanatory material to the main narrative verbs 'got to know' and 'did not value'.

The p-stem of the verb 'to be', *yaku:nu*, also has auxiliary modal and temporal functions, which, as one might expect of a p-stem in the light of the discussion so far, are in the area of as yet unrealized, or nonfactual action. Before an s-stem, *yaku:nu* has the following senses:

(a) *yaku:nu* preceded by *qad*: 'may have done X' (possibility)

qad ʔaku:nu waṣaltu   baʿda yawmayni
qad 1sg-be arrived-1sg after days-two
'I may have arrived in two days time'

(b) *yaku:nu* optionally preceded by *sa-*, *qad* optional before the s-stem verb: 'will have done X'

(sa) ʔaku:nu (qad) waṣaltu   baʿda yawmayni
FUT-1sg-be (*qad*) arrived-1sg after days-two
'I shall have arrived in two days'

The above summarizes the main syntactic means available for locating the action of the verb in relative time, if need be. Reading contemporary written Arabic, one has the impression that the use of auxiliaries to form "compound tenses" is much more widespread than was true in the writing of the early nineteenth century and certainly, going back further, when compared to medieval prose. However, still today, in some literary genres, it is by no means the case that verbs are always, or even usually, overtly marked for tense. Narratives and descriptions are the most striking example of this tendency. Generally speaking, once a past

time frame has been established at the beginning of a section of a story—usually a paragraph—subsequent overt time marking, morphologized in the verb in languages such as English and French, is felt to be redundant, and subsequent main clause p-stem verbs and participles within the paragraph will have no auxiliary verb to mark tense. In adverbial clauses of time ('After he had done that, he did this'), sequence of tenses is also left unmarked, being similarly redundant. Reported speech, whether actual or "interiorized," is not normally put in the past tense, either. The following translation of the opening paragraph of the short story *Naðra* (*A Look*)[40] by the modern Egyptian writer Yusuf Idris illustrates these points:

It was (s-stem (1)) strange that a little child like her should be asking (p-stem (2)) an adult like me whom she didn't know (p-stem (3)), simply and innocently, to adjust (p-stem (4)) the position of what she was carrying (p-stem (5)), and what she was carrying (p-stem (6)) was (s-stem (7)) really complicated: on her head was balanced (p-stem (8)) a tray of roast potatoes, and on top of this tray rested (p-stem (9)) a large tin basin full of baked pastries. In spite of her careful grip, the basin had slipped (s-stem aux. + *qad* + s-stem (10)), but she hung on to it (s-stem (11)) like grim death, to the point where what she was carrying (p-stem (12)) was (s-stem (13)) threatened with disaster (lit. 'falling to the ground').

Commentary:

The opening s-stem verb *ka:na* (1) marks the time of the story as past, relative to the time of the narration. Within this past time frame, verb (2), which is in a subordinate noun clause governed by 'was strange that . . . ' describes an ongoing, incomplete action, hence it is p-stem a-set—"subjunctive." The following verbs (3, 5, and 6) are in relative clauses and provide relevant circumstantial information about the protagonists in the story without taking the action forward—hence the use of the p-stem. Verb (4) reports the girl's speech: also p-stem. Verb (7), the s-stem copula *ka:na*, signals a change: although its function is descriptive, it advances the narrative by presenting the idea that what the girl was carrying was 'complicated'. This new topic is then amplified in the rest of the paragraph. P-stem verbs (8 and 9) then describe the nature of the complexity, whereas (10) fills in what we earlier termed the "anterior attendant circumstance" of the narrative thread. The s-stems (11) and (13) advance the action, whereas the p-stem (12) sandwiched between them in the relative clause is again descriptive.

What is noticeable here is that all the p-stem verbs are non-dynamic/descriptive and none are time marked by *ka:na*. Past time is marked by three cop-

ula *ka:na* + participle/adjective verb phrases, at the beginning, middle, and end of the paragraph. Although they have an inherently nondynamic sense ('was strange', 'was complicated', 'was threatened'), their past time marking signals that they represent stages in the narrative development. Of the lexical verbs, only (10), the pluperfect of anterior circumstance, is overtly marked for past time relative to the narrative. We shall make further observations on the contrasting functional uses of the p-stem and s-stem verbs in MSA, and the related question of word order, in chapters 7 and 8.

In the dialects, syntactic resources similar to those in MSA exist for expressing relative time. Thus in Cairene, to give just a few of the possibilities, we find:

ka:n + *b(i)*- proclitic:
ka:nu   by?u:lu . . .
were-3pl *b*-3pl-say
'They were saying/used to say'

ka:n + *ḥa*- proclitic:
ka:nu   ḥay?u:lu . . .
were-3pl *ḥa*-3pl-say
'They were going to say . . . '

ka:n + s-stem:
ka:nu   ?a:lu . . .
were-3pl said-3pl
'They had said . . . '

proclitic + *yiku:n* + s-stem:
byku:n   xallaṣ   iššuġl
b-3msg-be finished-3msg the-work
'He will (usually) have finished the work' (e.g., by 7 P.M.) (future perfect habitual)

ḥayku:n   xallaṣ   iššuġl
ḥa-3msg-be finished-3msg the-work
'He will have (definitely) finished the work' (e.g., by 7 P.M.) (future perfect)

ḥayku:n   byxallaṣ   iššuġl
ḥa-3msg-be *b*-3msg-finish the-work
'He will be finishing the work' (e.g., when I arrive) (future continuous (or habitual)) etc.

"Compound tense" forms of this type are used in all urban educated speech. This is not, however, the case in many rural and Bedouin dialect areas, especially

in the speech of the less educated and the elderly, whose dialect shows a less fully evolved system of proclitics and far less use of the auxiliary *ka:n*.[41] In this sense, uneducated rural dialectal usage resembles what occurs in much premodern Arabic prose writing. Let the following (translated) extracts from conversations[42] serve as an example of the differences between educated and uneducated speakers from the same community in the use of auxiliaries. In the first, an illiterate housewife ('F'), about forty years old, is telling an educated acquaintance ('T') about her early hardships; in the second, a female bank clerk, about twenty years old, is also talking about her childhood. Both women are from Bahrain, although of widely different educational and social backgrounds.

Extract 1

T: You lived (s-stem (1) the life of an orphan . . . didn't you have (no verb (2)) anyone (sc. to look after you) or anything?

F: No. I only had (no verb (3)) my mother and my brothers. My father had died (a.participle (4)).

T: So where did you use to go (p-stem (5)) (sc. to work), the palm groves or somewhere?

F: Of course I used to go (p-stem (6)) to the palm groves! Like I was just telling (p-stem (7)) you, the Ahmad family had given us (s-stem (8)) palm trees, so we became (s-stem (9)) date sellers, I mean, my two brothers. My mother worked (p-stem (10)) as a hired help and I was (no verb (11)) at home cleaning (p-stem (12)) the house. When I had cleaned (p-stem (13)) the house I would take (p-stem (14)) the date basket and go (p-stem (15)) and shake down (p-stem (16)) the ripe dates. After we came back (p-stem (17)) from this, we'd pick up (p-stem (18)) our weaving, we would sit (p-stem (19)) and weave (p-stem (20)). We would weave (p-stem (21)) anything—date sacks, mats. When you wanted (p-stem (22)) a date sack, or two or four, I would make them (p-stem (23)) for you.

T: Right . . . How did you have (*ka:n*. + p-stem (24)) your children? Where did you go (p-stem (25)) when you were about to deliver? (s-stem "ingressive" + p-stem (26))?

F: I had them (p-stem (27)) at home . . .

Commentary:

Once the past time frame has been established by the past time reference of T's s-stem in (1), her follow-up question (2), and the whole of F's response describing the habitual work and domestic patterns of her family at that time, are described in a series of non–time-marked, aux.-less p-stem verbs, or verbless

"equational" sentences ((3), (6), (7), and (10)-(23)). Apart from these, there are two examples of the explanatory "anterior of attendant circumstance" ((4) and (8)). In written Arabic, both would require *ka:n* as aux. + s-stem, and (8) would require *qad* as well; in educated Bahraini speech *ka:n* would tend to be used in both cases. F, however, omits *ka:n* here as elsewhere, because the past-time context already established, combined with her use of the simple p-stem for the "past habitual" ensures the correct relative-time interpretation of the s-stem, without it having to be marked as pluperfect by *ka:n*. In (24), by contrast, T, the educated speaker, uses *ka:n* for a "past habitual," even though speech context (and the fact that F is, in Bahraini village terms, past child-bearing age), makes it redundant; F, the uneducated speaker, replies with a simple p-stem.

Extract 2

Before, we used to travel (*ka:n* + p-stem (1)) abroad, we used to travel (*ka:n* + p-stem (2)) abroad a few years ago, when we were (*ka:n* (3)) kids. We were still living (*ka:n* + p-stem (4)) in Saudi Arabia a few years ago, like. We used to go (*ka:n* + p-stem (5)) to Cairo, but now we spend (p-stem (6)) all our time working. When I was (*ka:n* (7)) small we came back (s-stem (8)) to Bahrain, in 1970, it's now seven years since we came back (s-stem (9)) from Saudi, I was (*ka:n* (10)) quite young. We had (*ka:n* (11)) no freedom, I mean at school they made us (s-stem (12)) wear veils, we had to wear (p-stem (13)) them, and there were (no verb (14)) no cinemas . . . '

Commentary:

As in extract 1, the topic is the speaker's childhood, and a clear past time frame had already been set by the question to which this speech is a response.

Nonetheless, this more educated speaker uses *ka:n* as aux. or as a past-tense copula verb in (1), (2), (3), (4), (5), (7), (10), and (11): all cases where, in a similar speech context and text type, the uneducated speaker left it out. Only in (14), an equational sentence, is there an example of the omission of the time-marking copula, s-stem *ka:n*. The redundant use of auxiliaries to mark relative time is, it would appear, one of the characteristics of the syntax of contemporary written and urban spoken Arabic, as compared with older (more "orate"?) styles of writing and the speech of the present-day uneducated. In both of these, syntactic economy is the norm, with more reliance on context to make things clear that otherwise might not be. It seems possible that the exposure of literate educated speakers to languages such as English, with its elaborate tense system, is one of the reasons for this stylistic differentiation.

6.3 Negation

Although they share some of the same negative particles, the negation systems of MSA and the dialects differ. Here we describe in outline the structure of the two types of system and the main differences between them.

In MSA, the negative particles are *lam, lan, ma:,* and *la:*. There is also a negative verb *laysa*, 'to not be'. Orthographically, all of these are separate words and precede the element, verb or noun, that they negate. Their use is syntactically determined:

(a) <u>*lam*</u> occurs only before the base set p-stem and forms the negative analog of the s-stem: thus s-stem *daraba* 'he hit/has hit', but *lam yadrib* 'he didn't hit/hasn't hit'. *lam* is marked as an MSA item and is not normally used in spoken dialectal Arabic, except when a conscious effort is being made to speak "correctly," for example, where the context is so formal that speech begins to come under the influence of written norms, or perhaps to a foreigner who speaks Arabic but who it is thought will not understand the dialect.

(b) <u>*lan*</u> occurs only before the a-set p-stem (the "subjunctive"), which combination functions as an emphatic future negative, often with volitional overtones, for example, *lan nastaslima* 'we shall not/shall never surrender' (in a defiant political speech of Nasser's at the time of the Suez crisis). The future particle *sawfa* can also be used in combination with the general negative *la:* to express a future negative. This is regarded as unacceptable by purists but is much used nonetheless. Contemporary usage suggests that this construction does not have the emphatic force of the construction with *lan*: *lan yaqbala ða:lika* 'he will never (in a million years) accept that' suggests greater certainty, perhaps, than the more neutral *sawfa la: yaqbalu ða:lika*. Like negation with *lam*, negation with *lan* (or *sawfa la:*) is typical of prepared written, rather than extempore spoken, Arabic.

(c) In CLA, negative *ma:* could negate the u-set p-stem verb and was especially likely to do so in direct speech and in the reporting of actions contemporaneous with the utterance, as opposed to habitual or future action, where *la:* was normal: *ma: yadribu* 'He's not (now) hitting' versus *la: yadribu* 'He doesn't (usually) hit/won't hit'. *ma:* was also used for the negative of completed actions expressed by s-stems. This was particularly likely in first person direct speech where the speaker was certain of the truth of what he was saying: *ma: darabtu* 'I didn't/haven't hit', whereas there was a tendency for reports of third-party action, where the speaker might be less certain or less committed to the truth of the proposition, to be expressed by *lam* + base set p-stem (see 8.3.4). Neither of these semantic–affective distinctions[43] in the use of negative particles is any

longer observed to the same extent or in the same way in contemporary MSA. Nowadays, especially in documentary nonliterary writing, the u-set p-stem, whatever its aspectual or discoursal value, is in the overwhelming majority of cases negated by *la:*, not *ma:*, so that *la: yaḍribu* is used to mean 'he doesn't (usually) hit' as well as 'he is not (now) hitting/won't hit', and *lam* + base set p-stem is similarly dominant in all persons of the verb for the negative of completed actions.[44] Where *ma:* does occur with the p-stem in MSA, however, there is an echo of its CLA function—it is found virtually exclusively with stative verbs, predominantly those of cognition and emotion such as 'believe', 'know', 'suspect', 'think', and 'like'. When used with these verbs, it is far more common in the first than the third person. In other words, with these verbs, *ma:* + p-stem is associated with a more personal commitment to a (negative) proposition.[45]

One may speculate on the reasons for the recession of *ma:* as a particle of verbal negation in MSA. Historically *ma:* became, and remains to this day, the normal means of negating the verb, whether p-stem or s-stem, in all the spoken dialects (see below), whereas "interrogative" *ma:*, which the dialects also inherited from CLA, has almost completely disappeared from them.[46] As a consequence of this rearrangement, negative *ma:*, which had been a fully functioning part of the CLA negative system, seems to have become tainted as "incorrect" and has been marginalized in contemporary MSA, whereas interrogative *ma:*, which dropped out of the dialects, was seen as "correct" and has survived in MSA (see 5.3).

In certain verbless constructions, however, negative *ma:* is still used in MSA:
(i) with pronominal subject and predicate introduced by preposition *bi*:

ma: ʔana bi jundi:yin
not I   soldier-GEN
'I am not a soldier'

(ii) with preposition *min* 'of, from':

ma: min šakkin   fi: ʔan . . .
not of doubt-GEN in that
'There is no doubt that . . .'

(d) *laysa* is morphologically and syntactically a verb, but it is an anomalous one in two respects: it only has s-stem forms, yet does not refer to completed action, and it has none of the semantic functions of verbs described in 6.2. From the point of view of sentential function, two main uses and one minor use of *laysa*

can be distinguished: principally to negate an adjectival (or nominal) attribute ('X is not (a) Y') and to negate the existence of something ('there is not . . . '), and less commonly as an alternative to *ma:* in the negating of p-stem stative verbs, usually in the first person ('I don't know . . . ', 'I don't like . . . ', etc. see (c) above). In the first main use, the negated attribute may either be governed by the preposition *bi* (+ genitive) or be in the accusative; in the second, *laysa* is itself the predicate ('does not exist'), so the thing whose existence is denied is in the nominative. In both these types of structure, as in the third, *laysa* is in free variation with negative *ma:*. For example:

lastu bi jundi:yin
not-1sg soldier-GEN

or

lastu   jundi:yan
not-1sg soldier-ACC
'I am not a soldier'

= *ma: ʔana: bi jundi:yin*

laysa   huna:ka šakkun   fi: . . . = ma: min šakkin fi: . . .
not-3msg there   doubt-NOM that . . .
'There is no doubt that . . .'

lastu ʔuḥibbu llawna   lʔasmara = ma: ʔuḥibbu llawna . . .
not-1st 1sg-like the-color the-brown
'I don't like the color brown'[47]

If either of the first two sentence types is to be marked as in the relative past, *laysa* is replaced by *lam* + base set p-stem of *KWN*, which behaves syntactically like *laysa*:

lam ʔakun   jundi:yan
not 1sg-be soldier-ACC
'I was not a soldier'

lam yakun   huna:ka šakkun   fi: . . .
not 3msg-be there   doubt-NOM that . . .
'There was no doubt that . . .'

laysa may also be used to negate the whole of a proposition or an adverb that stands in its stead and is equivalent to 'It is not the case that . . . ':

ʔalaysa   nnabiːyu   qad maːtɑ[48]
Q-not-3msg the-Prophet *qad* died-3msg
'Hasn't the Prophet died?' ('Is it not the case that . . . ?')

ʔalaysa   kaðaːlika
Q-not-3msg like-that
'Is that not so?' (= French *n'est-ce pas?*)

Just as *maː* is now becoming uncommon in nonliterary styles of MSA as a
verbal negative particle, so it seems to be receding in favor of *laysa* in verbless
sentences,[49] and for the same reasons: the dialects have generally preserved vari-
ants of the construction with *maː*, whereas *laysa*, except in a few archaic dialects
of Arabia, is a saliently nonspoken (and therefore to most users' minds more
"correct") form.

(e) *laː* is the negative particle with the widest range of uses in MSA and is
historically the oldest—*lam, lan,* and *laysa* being ultimately derived from the fu-
sion of *laː* with other elements. In MSA, *laː* negates the u-set ("indicative") and
a-set ("subjunctive") p-stem verb. It is used with the zero-set p-stem for negative
imperatives, as well as with the s-stem in negative exhortation. For example:

laː yaḍrib-u   lwalad
not 3msg-hit-INDIC the-boy
'He doesn't hit/isn't hitting the boy'

yajuːzu   ʔan laː yaḍrib-a   lwalad
3msg-is-possible that not 3msg-hit-SUBJ the-boy
'It's possible/permissible for him not to hit the boy'

laː taḍrib-ɸ   lwalad
not 2msg-hit-JUSS the-boy
'Don't hit the boy!'

laː samaḥa   llaːh
not permit-3msg God
'God forbid!'

laː is used as the particle of general denial of propositions, equivalent to En-
glish 'no':

hal yuriːdu   nuquːdan? — laː
Q 3msg-want money-ACC no
'Does he want money?'—'No'

It is also used in equivalents to English "no + substantive" constructions, in
which case the governed noun is marked as indefinite accusative but without
nunation:

laː ʔilaːha: ʔillaː llaːhu
not god-ACC except the-God-NOM
'There is no god except God'

laː is also used as a correlative negative, in which the first negative may be
any negative. For example:

maː ʔana bi jundiːyin   walaː   bi mudarrisin
not I   soldier-GEN and-not   teacher-GEN
'I am neither a soldier nor a teacher'

The dialectal negative systems as a whole are simpler than the CLA/MSA
system and show a large measure of similarity one to another, all apparently
having undergone the same reductional tendencies.[50] The basis of the dialect
negation systems is a division of predicates into two types: verbal (i.e., s-stems
and p-stems with or without modal proclitics but including also certain prepo-
sitional phrases that have developed a verbal function—see below) on the one
hand, and the rest (i.e., nominal, adjectival, adverbial, prepositional, and whole
sentence) on the other.[51]

In all dialects, most types of verbal predicate are negated by an immediately
preverbal *maː*, with or without a morpheme -*š* suffixed to the verb phrase. In
some (e.g., Palestinian), the -*š* seems to be merely optional, whereas in others
(Moroccan), its absence or presence is associated with differences in the cate-
gorical/noncategorical nature of the complement.[52] Nonverbal predicates are
negated by a particle that is variously *muš/miš* (Egypt, Libya, parts of Yemen,
and Morocco), *mu* (Levant generally), *mu/mi/mub/muhub/hub* (Gulf states), and
maši (Morocco, Algeria, Tunisia, and parts of Yemen).[53] Etymologically, these
forms are all cognates and all reduced composite forms, derived from different
combinations of a closed set of elements: (1) *maː* 'not' (2) *hu/hi* 'it, he, she', (3)
bi a semantically empty prepositional element, and (4) *ši* 'thing'. Thus Cairene
combines 1 + 2 + 4, Levantine 1 + 2, etc.

The negative *laː* is used for 'no' in answer to yes/no questions, occurs in a
few fixed phrases such as *walaː ḥaːga* (Cairo), 'nothing' and in some dialects is
retained, as in MSA, for negative imperatives, negative optatives, and coordi-
nated negatives. Thus, in the dialect of the Gulf *maː truːḥ* 'you don't go' con-
trasts with *laː truːḥ* 'don't go!', whereas in Cairene and Moroccan there is no such

distinction in the mode of negation—both have *ma: . . . š(i)* for both types of form—the meaning distinction being carried by the presence of a modal pro-clitic in the first and its absence in the second: Cairene *ma: btruḫši* 'you don't go', *ma: truḫši* 'don't go!'. The negative functionals *lam, lan,* and *laysa,* all essential elements of the MSA system, are not used at all in relaxed informal dialect speech anywhere.

Certain formally prepositional phrases, when used to express existence, pos-session, and volition, have acquired in the dialects a quasiverbal function and are treated syntactically as if they were verbs rather than prepositions. So *fi:* 'there is/are' (lit. 'in it'), *'ind* 'with', and *il-/li* 'to', to which pronouns are suf-fixed with the sense of 'I/you/he, etc., have' (lit. 'with/to me/you/him, etc.) are generally negated with *ma: . . . (š),* not *muš* or one of its cognates: *ma: fi:š fayda* 'it's no use', *ma: 'indi:š ma:ni'* 'I have no objection'. The same is true of the gen-eral Levantine *bidd-* 'to want', which is etymologically a prepositional phrase *bi widd-* 'in the desire of . . . ' to which enclitic pronouns are suffixed: *ma: biddi:š ašu:f ilfilim* 'I don't want to see the film'.[54] The prepositions that are negated in this way also retain a locative function, however, and are negated according to the general principle already given: contrast the treatment of *'and* when used possessively (with "verbal" negation) in *ma: 'andi kta:bak* 'I don't have your book' (lit. 'not with me your book') and its treatment when used locatively (with "non-verbal" negation) in *kta:bak mu 'andi* 'Your book is not with me (e.g., not at my house)'.[55] There have been other general dialectal developments also, such as the use of *'umr* 'life' with the negative to mean 'never, no more': *'umruh ma: ra:ḫ landan* 'he has never been to London' (lit. 'his life he didn't go . . . ').

A brief general comment on the use of these different negation systems is in order here. In written Arabic, except where an attempt is being made to imi-tate ordinary speech, as in newspaper cartoons and folk poetry (see chapter 2), the MSA system is naturally the norm. As we have already briefly noted, how-ever, there is a distinct tendency in much documentary MSA to avoid negative forms that are shared with or superficially similar to those of the dialects. Chiefly, this means avoiding the negative with *ma:* (even though such structures are perfectly "correct" by the canons of CLA) and insisting on modes of nega-tion that are saliently nondialectal, *lam, lan,* and *laysa* being the most obvious. In educated speech, however, there is much context-related variation. Like the deictic functionals (pronouns and demonstratives), the negative particles are part of the dialectal "bedrock" and less susceptible to replacement by MSA equiva-lents than lexical content items. Like the use of word-final short vowels to mark mood as opposed to the use of word-initial proclitics, or the use of the inter-nally voweled passive as opposed to a passivizing prefix, the use of the MSA

or dialect negation system is heavily marked as a "correct," "formal" feature and carries a commensurate stylistic significance—it signals that the speaker is mak-ing a deliberate effort, for whatever reason, to avoid his or her normal patterns of speech (see chapter 9).

Notes

1. A detailed comparative study of the evolution of several major syntactic categories in four geographically dispersed dialects—Moroccan, Egyptian, Syrian, and Kuwaiti—can be found in Brustad 2000.
2. Blau 1981b, 31 and 60ff; Blanc 1957, 400.
3. Case endings are marked in example sentences where this is relevant to the grammati-cal explanation or in examples where such endings are invariably pronounced (e.g., re-ligious texts).
4. In this latter case, the defined verbal noun can always be replaced by a noun clause in which the subject of the verb is, of course, specified for person.
5. The marking of adjectives as definite when used as attributes of defined nouns may have its origins in appositional juxtaposition—that is, what we now perceive as a noun + ad-jective structure was originally an apposition of two nominals, so that, for instance, *ʔar-rajulu l-ġabi:yu* 'the stupid man' may have originally been parsable as 'the man, the stu-pid one'. See Beeston 1970, 44–45.
6. Belnap and Shabaneh 1992, 260.
7. Nu'ayma 1966, 12.
8. See Abboud 1964, 74–76, Ingham 1994, 61–68 for the agreement system of Najd, cen-tral Saudi Arabia; and Owens and Bani-Yasin 1987 for a quantified study of the diffu-sion of the MSA agreement system in the spoken Arabic of rural northern Jordan.
9. For example, Holes 1990, 155–56 notes four systems in operation in different areas of the Gulf coast.
10. See, for example, the discussion in Belnap 1993, especially 106–10, in which the influ-ence of this semantic variable on patterns of agreement in Cairene Arabic is illustrated.
11. Barlow 1988, 306–7.
12. Cowell 1964, 424.
13. Beeston 1970, 46–47.
14. This structure is one of the main analytic features that differentiate the Arabic dialects typologically virtually without exception from CLA/MSA, and it is at least 700 years old. Blau records it for medieval Christian and Judaeo-Arabic (Blau 1988, 52, 70, 181). It may well have originally arisen in those areas of the eastern Mediterranean area where, as we noted in chapter 1, Arabic was for many centuries in contact with Aramaic, all of whose dialects had an analytic genitive of this type (Brockelmann 1908–13, 2:244–45).
15. Blau 1976, 162 notes that this construction also occurs in MSA in (his examples suggest (CH)) dialogue and reported speech in novels. The use of this construction may be an attempt to alleviate the impression of artificiality and stiffness of "speech in MSA" and

lend a dialectal flavor to dialogue in the type of literary context where dialect is normally avoided (see 9.5).

16. Badawi 1973, 140–41.

17. A vestige of the "gender polarity" system survives in the case of one or two nouns that commonly occur as the noun in number expressions ("day," "month," and "thousand") but has now acquired a phonological function: these nouns begin with a glottal onset, but this is dropped in juncture and a -*t* (originally that of the feminine form of the number) is used to make the liaison. This -*t* is phonologically part of the following noun, for example, from (pause forms) *ʔiyya:m* 'days' and *ʔa:la:f* 'thousands' we get *xams tiyya:m* 'five days', *saba' ta:la:f* 'seven thousand'.

18. Cowell 1964, 367. Damascene and many Syro-Mesopotamian dialects have alternative dual forms for certain words that designate parts of the body. Where, for example, 'eyes' in the plural is being referred to, the -*e:n* dual suffix functions as a plural marker, not as a dual: *arba' 'e:ne:n* 'four eyes'; but where the reference is to two eyes *as a pair* (in contrast to 'one'), a –*t* "singulative" suffix is added to the noun: *'e:nte:n* 'two/ both eyes'. See Blanc 1970 for a discussion of what he terms "true dual" (*'e:nte:n*) and "pseudo-dual" (e.g., *'e:ne:n*) forms.

19. In Baghdadi Arabic, an indeclinable form *fadd* (< *fard* 'individual') is used to mean 'a/an', for example, *'indak fadd galam* 'Have you got a pen?', *ha:y fadd fikra ze:na* 'That's a good idea'. See Erwin 1963, 355ff.

20. Johnstone 1967, 64 for the dialects of the Gulf, for example.

21. Eisele 1999, in a study of tense and aspect in Cairene spoken Arabic, provides a review of work on these topics in all forms of Arabic and argues that the traditional view of an aspectual system or (the view I am putting forward here) a historically aspectual system that is evolving into a tense/aspectual system, but at different speeds in different varieties of Arabic, is flawed by being "limited in scope: too little data, flawed data, or a reliance on basic terms and frameworks which incorrectly describe this data" (Eisele 1999, 24). The charge of "too little data," however, could equally well be leveled at Eisele's own analysis; in the dialects of some areas that he leaves out of account—the Arabian peninsula and the Gulf, for instance—the system used by the noneducated still looks more like an aspectual one than one in which the s-stem/p-stem distinction is primarily one of tense. Ingham's view of the Najdi dialects, based on an analysis of very large amounts of text in many subgenres gathered from uneducated speakers, is that "the Najdi dialects still preserve the original aspect-centred system of Classical Arabic, although it can be shown that structures showing a new tense-based system can occasionally be found within the macrostructure." (Ingham 1994, 87, and 87–115 more generally).

22. Using comparative evidence from older Semitic languages, Moscati et al. (1964, 133–34) and Voigt (1988, 102ff.) argue that s-stems and p-stems were prehistorically in complementary distribution: the s-stem expressed states ('to be big') while the ancestor of the p-stem expressed actions ('to hit'). Only later through a long process of analogical systematization did s-stem verbs develop p-stems and vice versa. The origin of the s-stem itself may have been nouns to which personal pronouns were suffixed (which then developed into the inflections we see today).

23. Beeston 1970, 79.

24. The first example is from an article written in the Egyptian newspaper *al-Akhbār* on 15 January 1991, the day *before* the bombing of Iraq began, and hence the participle here can only have had a future, predictive sense. The second is a recent literary example, taken from Cantarino 1974–75, 2:409.

25. See, for example, Caubet 1986, 109ff. for a detailed account of the s-stem/participle aspectual distinction in Moroccan Arabic; Holes 1990, 190ff. for Gulf Arabic; Woidich 1975 for Cairene; and Mitchell 1952 for Cyrenaica, Libya.

26. The *b(i)*- prefix is discussed later in this section.

27. The data cited here are taken from Mitchell and El-Hassan 1994, 16. See also Henkin 1992, 438ff.

28. For example, according to Cohen 1975, 136, the proclitic *qa*-, which denotes continuous action in the Jewish Arabic dialect of Tunis, derives from the active participle *qa:'id* 'sitting'. This *qa*- or *qa:'id* occurs with a similar aspectual value in other dialects as geographically widespread as Cairene, Iraqi, Sudanese (Agius and Harrak 1987), and Gulf. The Damascene proclitic *'amm*-, which has a similar function to *qa-/qa:'id* had an analogous morphosemantic origin in *'amma:l*, a species of adjectival participle derived from the verb *'imil* 'to do, work' (= 'doing', 'working').

29. This is observable today in some Gulf dialects. The p-stem of the verb 'to want' may be prefixed asyndetically to other p-stem verbs to give a sense of 'proximate intention' as well as 'wanting'. Thus *ʔabba ru:ħ* 'I want/am going to go', and *tabba tru:ħ* 'you want/are going to go' (Shi'i Baħārna dialects of Bahrain). In other dialects (among them the Sunni Arab dialects of Bahrain, but Gulf littoral dialects generally) the 'want' verbal element has decayed to a proclitic *ba-/bi-* and has lost its volitional sense, viz. *baru:ħ* 'I'm going to go', *bitru:ħ* 'you're going to go', whereas the lexical verb 'to want' from which this particle was derived retains its original sense; that is, in these dialects there is a contrast between *baru:ħ* 'I'm going to go' (intention) and *ʔabbi aru:ħ* 'I'd like to go'. This Gulf "intentive" *ba-/bi-* is quite different from the Egyptian/Levantine *b(i)*- form discussed later in this chapter, which has a different origin (according to Blau 1981a, 121–22, it is ultimately to be traced to the preposition *bi* 'in, at, with' and first appears in Christian Arabic texts of the ninth century A.D.) and a quite different modal use.

30. The Levantine dialects in general differ from Cairene in having the *b(i)*- prefix in both these senses, so that in Damascus and Jerusalem the equivalent sentences would be the same and, out of context, ambiguous as between factual inquiry and invitation.

31. The initial consonant cluster that results from the deletion of the p-stem prefix in some verbs is resyllabified in the urban Mashreqi dialects (not in the Maghrebi ones) by a prosthetic syllable *ʔi*-, the phonetic value of whose vowel varies from dialect to dialect.

32. Doss 1987 claims that some Cairene speakers of what she calls Spoken Literary Arabic (see 9.3.2), a somewhat dialectalized variety of MSA used in extempore speaking in the media, now have a tendency to generalize the *b(i)* + p-stem verb to all contexts where a p-stem verb is used in written MSA.

33. Mitchell and El-Hassan 1994.

34. Bahraini, like the other Gulf littoral dialects, does not (at least natively) have an "indicative" *b(i)*- proclitic (see n. 29 above). The zero-marked form is the usual one in contexts where in Cairo and the cities of the Levant the *b(i)*- proclitic is used.

35. Where a nonsubject noun is extraposed to the beginning of the sentence (see chapter 7), *ka:na* agrees with it and not with the "underlying" subject. For example:

> *ka:nat*   *ha:ðihi lmar?atu   zawjuha: ya'malu   ta:jiran*
> was-3fsg this   the-woman husband-her 3msg-work   trader-ACC
> 'This woman's husband was working as a trader'

36. Nominal complements of *ka:na/yaku:nu* are construed in the accusative case. A number of other verbs, all of which denote states or changes of state, behave similarly: 'to remain', 'to become', 'to still be', 'to no longer be', 'to almost be', etc.

37. In its dialectal manifestation, *gad* is still used in this sense in the dialects of the Arabian peninsula, for example, (Bahraini example) *ma: gad riht issa'u:di:ya* 'Have you never been to Saudi Arabia?'.

38. Cantarino 1974–75, 1:69.

39. Ziadeh 1964, 219.

40. Idris n.d., 12.

41. See Abboud (1964, 38–39, 100) and Ingham (1994, 97ff) on Najdi Arabic, in which the absence of the auxiliary *ka:n* as a tense marker from "pure" dialect speech is specifically commented on. Informants regarded this use of *ka:n* as a "classicism" (i.e., a written form).

42. The first extract is from a text published earlier (see Holes 1984, 40–45 for the full text and translation), and the second is part of a large unpublished corpus of Bahraini material.

43. " . . . der primäre Unterschied zwischen *lam yaktub* und *ma: kataba* nicht Tempus, sondern in der persönlichen Anteilsnahme des Sprechers an der Aussage, also im Affektgrad zu sehen ist . . . " ["the primary difference to be seen between *lam yaktub* and *ma: kataba* is not tense, but rather in the personal commitment of the speaker to the statement as well as the degree of affect . . . "](Wehr 1953, 31).

44. A recent analysis of negative usage in a sample of some 3,067 pages of post–World War II literary and nonliterary MSA texts produced 3,094 examples of *lam* + base set p-stem as against 386 of *ma:* + s-stem to express negative perfect or past, and 3,414 examples of *la:* + u-set p-stem as against 130 of *ma:* + u-set p-stem to express negative nonpast indicative (Rammuny 1978, 261).

45. Ibid., 252.

46. Interrogative *ma:* survives as the normal means of asking 'what?' in the sedentary dialects of Oman.

47. Example taken from Rammuny 1978, 256.

48. Example taken from Cantarino 1974–75 1:125.

49. In his more than 3,000 pages of text, Rammuny 1978, 254–56 reports only 72 instances of *ma:* in negative equational sentences, compared with ten times that number, 715, of *laysa* to express the same type of meaning ('X is not Y').

50. Again, a qualification needs to be entered in regard to some of the Arabian peninsula dialects, where vestiges of the CLA negation system have been preserved that were totally lost in the cities of the eastern Mediterranean. For example, *le:s* 'it is not' (cf. CLA/MSA *laysa*) still occurs as a fossilized third person only form in Bahrain in the speech of nonliterates in some communities.

51. "Equational" sentences that in Arabic do not have a verb ('X = Y') have alternative structures in many dialects, for example, Cairene *ma:* (. . . *š*), in which case the subject pronoun is negated: *mantu:š kwaysi:n* (< *ma:* + *?intu* + *š*), or with *muš* when the predicative adjective is negated: *?intu muš kwaysi:n*, in both cases meaning 'You're not nice'.

52. Harrell 1962, 154 contrasts *ma: žbert flu:s* 'I didn't find (any) money' with *ma: žbertš liflu:s* 'I didn't find the money'.

53. An apparent exception to this schema is that in Cairene *muš* is used to negate p-stem verbs that have the *ha-* future proclitic. However, this *ha-* is a decayed form of the active participle *ra:yih* 'going', which is formally a nominal. The same thing applies in Damascene, where *mu* occurs as an optional alternative to *ma:* before the future proclitic *rah/lah*, both historically derived from the same source as Cairene *ha-*.

54. In relaxed speech, some speakers regularly omit the *ma:* element of the *ma: . . . š* discontinuous negative, so that, for example, either *ma: biddi* or *ma: biddi:š*, or simply *biddi:š*, can all be used for 'I don't want'.

55. These are Damascene examples (Cowell 1964, 387) but parallel functional contrasts exist in equivalent sentences in most Arabic dialects.

7

Syntax and Semantics II: Sentence Structure
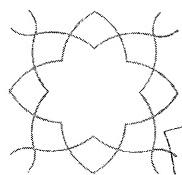
7.1 Word order

7.1.1 The functions of word order

Both CLA and MSA mark nouns for case, conventionally termed in western accounts of Arabic grammar "nominative" (Ar. *raf*), "accusative" (*naṣb*), and "genitive" (*jarr*).[1] Much is made in the indigenous grammatical tradition of the flexibility of word order that this case marking of nouns supposedly gave. In medieval treatises from Sībawayhi on, and in standard treatments of Arabic grammar even today, VSO, SVO, VOS, and OVS are all adduced as possible constituent orders for the Arabic equivalents of sentences such as "Dog bites man" (usually "Zayd hit 'Amr"), the case system supposedly obviating any ambiguity about whether it was the man or the dog who did the biting, whatever the order of the sentence elements. But like many of the sentence examples adduced by exponents of transformational generative grammar in the 1950s and 1960s, those given by the Arab grammarians have a flavor of artificiality about them: sentences constructed by schoolmen in order to prove a point whose truth had been assumed a priori without reference to the facts and contexts of actual usage, although of course no one would deny that they must have been based on the grammarians' native-speaker intuitions.

Although it is true that word order in CLA was, especially in poetry, flexible to a degree (although much less so than in a language such as Latin), it is doubtful whether this flexibil-ity was ever attributable to the disambiguating power of case marking, and in today's MSA, it certainly is not. As was pointed out in 4.5, case marking in most categories of noun involves only differences in the final short vowels, which do not appear in the script: only in (careful) reading aloud is the reader required to assign case endings. The lack of overt marking in the script does not lead to misunderstanding, precisely because there is a normal word order for MSA that is determined primarily by a principle of information organization: what is already "known" from the previous text or context (and is usually grammatically definite) precedes what is "new" (and is usually indefinite), regardless of whether what is known/definite is the grammatical subject or object.[2] In cases where both subject and object are known, or both are new, the subject, all things being equal, precedes the object. In MSA (and for that matter CLA), these basic principles account for word order in the vast majority of sentences in prose texts but may be overridden in cases of contrastive emphasis or topicalization (see below). But by no means does the fact of grammatical case marking in itself explain anything about the observable patterns of word order.

Before examining word order in more detail, we need an operational definition of the Arabic sentence. The European division of written texts into self-contained meaning groups ("sentences") that are divisible into constituents ("phrases" and "clauses"), marked off by the devices of punctuation, is absent from much MSA writing. Until perhaps the latter part of the nineteenth century, much Arabic writing contained no punctuation at all and no fully standardized system of punctuation exists even today. Usage of periods and commas can be highly variable and idiosyncratic, particularly in literary writing. This does not in fact matter: whether punctuation is used or not, it functions alongside the native system of textual chunking, which relies on coordinating and subordinating conjunctions that perform the dual role of signaling formally the beginnings and endings of sense groups and indicating the nature of the logical or functional relationships between them.

Grammatically speaking, a sentence in written Arabic consists of a subject and a predicate. The subject (= S) may be freestanding, that is, a noun/independent pronoun; or dependent, that is, consisting of one or more bound morphemes that form part of the verb (= V) (if there is one) and that indicate the person, number, and gender of the subject. The predicate may or may not contain a verb. If it does contain one, the subject may or may not be freestanding; if it does not, the sentence subject must be freestanding. The verb may or may not have a complement (= COMP). In the structurally simplest type of verbal sentence, the most statistically frequent[3] order is VSCOMP. In verbless sentences, it is SCOMP. Thus the following are all complete grammatical sentences in Arabic:

(a) freestanding subject:

V   S   COMP

yaskunu   ʔaxi: fi: ba:ri:s

3msg-lives brother-my in Paris

'My brother lives in Paris'

(b) dependent subject (indicated by verbal inflection):

V   COMP

ʔaskunu fi: landan

1sg-live in London

'I live in London'

(c) dependent subject, no complement

V

mariḍat

fell-ill-3fsg

'She fell ill'

(d) verbless sentence

S   COMP

ʔana ʔinkili:zi:yun

I   English

'I am English'

In VSCOMP sentences, there is no morphological restriction (p-stem or s-stem) or semantic limitation (stative or dynamic) on the type of verb that may fill the V slot. However, there is a strong tendency in narratives (in newspaper reporting par excellence) for VSCOMP to be reserved for what have been described[4] as "event-oriented" messages—where the focus is on who did what, what happened, or how it happened. The main alternative order, SVCOMP, is common in sentences that describe or provide incidental background or explanatory information about the already defined agents and patients of the text but do not themselves take the narrative forward. Entities that are textually "new" (and grammatically indefinite) cannot appear as S in SVCOMP structures, because the textual function of this ordering of elements is precisely to impart new information, in the predicate, about entities that are already known and are hence grammatically definite.

Expository writing, which often involves defining key terms or describing the structure of entities with component parts, tends to contain a high proportion of sentences with SVCOMP order. The verbs may be colorless statives such

as "consist of," "include," "comprise," or performatives such as "can be defined as": the text is "entity oriented" (whether the entities are defined generically or specifically—see 6.1.1) rather than "event oriented." Typically, p-stem verbs, which often describe current or timeless states and durative, nonpunctual actions, occur in such SVCOMP sentences; active participles invariably do. SVCOMP sentences in MSA are thus often functionally akin to verbless SCOMP sentences, which, *a fortiori* describe states, not actions, and in which the subject must also be definite.

A caveat must be entered here, however, that, in the case of "event-oriented" sentences, the preference for VSCOMP is to some degree a matter of personal choice. Some journalists occasionally use SVCOMP where it is clearly an "event" that is being described but for no contextually obvious reason. A contributory influence may be the writer's dialectal background: "urban" dialects, such as Cairene and Damascene, which are spoken by the majority of writers and journalists, have SVCOMP as the normal order for all types of message. Another possible influence may be the typically SVCOMP order of European languages, especially English, with which many Arab writers are familiar and which have a predominant position in news dissemination and in the cultural transfer of technical information of all kinds. We shall return to the topic of word order in media Arabic in more detail in chapter 8.

VCOMPS occurs relatively infrequently. This infrequency is unsurprising, given that the main circumstance in which it would occur, in line with the principle stated above, is where the subject is indefinite and the complement definite. This is a relatively unusual occurrence because what is unknown (and hence grammatically indefinite) is usually predicated of something already known (and hence usually grammatically definite). But such sentences do occur. For example:[5]

daxala   lqa:ʕa   rajul qaṣi:r

entered-3msg the-hall man short

'A short man entered the hall'

In this example, an indefinite subject, which brings a "new" element to the narrative, is shifted to the right of the complement, a textually "known" element. It can then become the focus, or theme, and occupy the normal subject position in the subsequent development of the story. The shifting of the subject away from its usual position has the concomitant effect of highlighting it: a fact that can be exploited rhetorically. In an open letter addressed to all Kuwaitis just after the Iraqi invasion (2 August 1990), the emir of Kuwait stated (my transla-

tion): "This small, peaceful, and peaceable country has been invaded by hundreds of tanks, overrun by tens of thousands of troops, and its clear skies have been rent asunder by squadrons of airplanes spreading terror and destruction."[6] In the Arabic text, the verbs "invaded," "overrun," and "rent apart" are in the active voice, and their subjects "tanks," "troops," and "airplanes" are all placed after the object "this small . . . peaceable country" (or its anaphoric pronoun) in a series of three structurally parallel VCOMPS sentences. The point of choosing to describe what happened in this way is to throw emphasis onto the displaced subjects that, cumulatively, have the effect of underlining the scale of the Iraqi attack. The first part of the Arabic sentence thus reads:

V |———————— COMP ———————|
ijta:ḥat   ha:ða: lbalada   ṣṣaġi:ra   l?a:mina
invaded-3fsg this   the-country the-small the-peaceful

———————|——————— S ——————————|
lmusa:lima   mi?a:tu   ddabba:ba:t
the-peaceable   hundreds (of) the-tanks
'Hundreds of tanks invaded this small, peaceful, peaceable country'

Coincidentally, these examples illustrate a second, somewhat less tangible principle involved in word order: "end weighting." In MSA, the "heavier" sentence constituent—that is, whichever lexical string occupying the S or COMP slot contains the most lexical items—is generally placed to the right of the "lighter" one. This is because the rhythmic break in an Arabic sentence, which occurs at a major constituent boundary, typically falls at, or slightly before, the halfway point in the sentence. Where, as in the first VCOMPS example cited above, the indefinite subject is "heavier" than the complement, this rhythmic principle requires that the "heavier" element, S, follow the lighter, COMP; if these elements were ordered in the normal way, viz. VSCOMP, the rhythmical break would come after VS and before COMP, which in this case would shift the break too far toward the end of the sentence. The effect of this auditory aesthetic is usually to reinforce the "definite first" principle, in that textually new elements, by virtue of so being, often contain more lexical items than ones that are already known and are thus shifted to the right for both information-structural and rhythmic reasons. However, where there is a conflict, as when an indefinite S is "lighter" than a definite COMP (and should according to the "rhythmic" principle precede it but according to the "informational" principle follow it), the information-structural principle usually wins. For example:

V   COMP   S
daxala   lqa:'a   lkabi:ra   lmuzaxrafa rajul
entered-3msg the-hall the-large the-ornate man
'A man entered the large ornate hall'

If COMP is an object pronoun, the VCOMPS order is obligatory if the verb construes its object without a preposition: the object enclitic, a bound morpheme, is attached directly to the verb. VCOMPS is virtually obligatory also if the verb construes its object with a preposition, whether the subject is definite or indefinite. This follows from both of the word order principles just stated, in that first, pronouns necessarily refer to textually "known" entities and therefore precede "unknown" nominal ones, and second, if the nominal subject is "known," it is virtually bound to be lexically "heavier" than a preposition + object enclitic form consisting of a single word and on this ground is dislocated to the right. Thus the normal order, whether S is definite or indefinite, is VCOMPS, as in:

V   COMP   S (def)
yuqbilu   'alayhimi   zza:?iru   lġari:b
3msg-approaches upon-them the-visitor the-foreign
'The foreign visitor approaches them'

        V   COMP   S (indef)
waði:fa qa:ma   biha: mas?u:lu:na ḥuku:mi:yu:n
duty   performed-3msg it   officials   governmental
'A duty that government officials performed'

However, the Arabic equivalent of sentences such as "The boy's father loved him," which one might expect, in the light of the above, would be turned 'loved-him the father of the boy', where 'him' and 'the boy' have the same referent, can in fact be rendered only, if a verb-first sentence structure is utilized, as 'loved the boy his-father'. This is because pronouns do not normally refer forward in Arabic, so 'him' in 'loved-him the father of the boy', although perfectly grammatical, could be understood only as referring not to 'the boy', but anaphorically to some other previously mentioned masculine referent.

The other possible combination of V, S, and COMP, in simple sentences where no elements other than these three occur, is COMPVS, where COMP is fronted to preverbal position. This order occurs in cases where emphasis, usually contrastive, is being put on COMP, as in:

```
COMP        V        S
ʔalḥaqi:qa lam yaqul  ʔaxu:ka
the-truth  not 3sg-tell brother-your
```
'*The truth*, your brother didn't tell' (sc. 'he told a lie')

In nominal sentences, as already noted, the normal order is SCOMP. The reverse order is possible and occurs in the same circumstances as COMPVS, viz. for emphasis:

```
COMP            S
rajul ġari:b      ha:ða:
man    strange this
```
'Strange man, this one'

Where there are two complement phrases, it is very common indeed in MSA writing to find an adverbial one in sentence-initial position, particularly where it is temporal or where it provides a logical link with the preceding section of texts, for example, "On the morning of the 2 August 1990 . . . ," "Despite these frequently voiced accusations, ..." As in English, this preposing of often very lengthy adverbial phrases provides a means of locating the main proposition in a contextual frame or of assuring the semantic coherence of the text by making explicit the nature of the linkage between sentences or paragraphs. Such sentence-initial adverbial phrases are then normally followed by the verb, with the order of S and COMP being determined by the principles already outlined. There seems to be no preferred order for any additional adverbial complements of location or time that follow VSCOMP or VCOMPS, save that the principle that the longer, "heavier" elements are normally further to the right than the lighter also applies and that the positioning of an adverbial element at the end is an indication that it may be taken up as the theme of the succeeding text. Hence following a sentence such as[7]

```
V            COMP (temp)     COMP (loc)
ʔihtada:      baʕda suʔa:l   ʔila: ša:riʕi lfusṭa:t
arrived-3sg  after asking    at    street (of) Fustat
```
'He found his way, after asking, to al-Fustat Street'

the natural expectation is that the locative complement will form the next topic of the discourse. The next sentence does indeed fulfill this expectation: "It was like Rashad Street in its size and tranquility, flanked by tall trees . . . ." Similarly, the expectation at the end of a sentence such as[8]

```
V        COMP (loc)   COMP (temp)
ʔaqa:ma fi: lqa:hira    ʔusbu:ʕayn    ʔaw ʔakθar min  ʔusbu:ʕayn . . .
spent-3msg in Cairo  two-weeks   or   more  than two-weeks
```
'He spent two weeks or more in Cairo'

is that the temporal adverbial will be taken up, and, in the event, the next sentence does describe what the protagonist did in the course of these two weeks.

There is one special type of adverbial complement that is almost always in final position: the so-called absolute accusative. In this structure, a verbal noun that is cognate with the main verb (or participle in a sentence without a finite main verb) becomes its object, with or without further adjectival modification. For example:

```
ḍarabahu       ḍarban       šadi:dan
hit-3msg-him   hitting-ACC strong-ACC
```
'He hit him hard'
```
ʔaʕrifuhu       tama:ma              lmaʕrifa
1sg-know-him completeness-ACC (of) the-knowing
```
'I know him very well'

The position of sentence elements, it should now be clear, is actually relatively fixed in MSA. Differences in order reflect mainly a division of sentences into "event-oriented" (V first) and "entity-oriented" (S first) messages, and in both types the norm is for contextually known nominal elements to precede new ones, regardless of their narrowly syntactic role. Right-located "new" elements in any given sentence often then form the thematized (and hence left-positioned) "known" element in a subsequent sentence. The principle of left positioning of thematic elements is also involved in the devices of extraposition and clefting, both of which move an already "known" element from the expected postverbal position to preverbal. In both cases a resumptive pronoun is used in the proposition that is predicated of the moved element to refer back to it. We will briefly examine the function of each.

Extraposition had a wide variety of discoursal functions in CLA[9] but perhaps the most commonly encountered in modern writing is to signal the continuing relevance of a topic to the development of an argument: the extraposed element tells the reader what the text continues to be "about." Extraposed nominals are sometimes, but not necessarily, marked syntactically by the particle ʔinna (which in CLA served the more general function of introducing noun-initial sentences).[10] Clefted elements may also very often be introduced by ʔinna, which

seems, in modern writing, to indicate emphasis. The functional difference be-
tween extraposition and clefting seems to be one of degree of emphasis: extra-
position merely highlights what the text is/continues to be "about," whereas
clefting extraposes *and* emphasizes at the same time. The following is an illus-
tration of how these two devices are used.

In an exposé of the shortcomings of the Kuwait government's demographic
policy,[11] the writer begins his chapter by pointing out that an understanding
of the concept of a society's 'human potential' (*ta:qa bašari:ya*)—placed inside
speech marks in the Arabic to mark it as a key technical term—is necessary. In
the next paragraph, he then develops this concept via a preposed (= thematized)
adverbial: "In order to understand the nature of a society's 'human potential,'
it is necessary . . . ," going on, in the lengthy main clause, to underline the im-
portance of first dividing society into distinct economic strata. In the next sen-
tence, the writer returns to his key term "human potential," extraposing it out
of a complex nominal to sentence-initial position in order to highlight it as the
superordinate topic and emphasizing one aspect of this topic by the use of a
cleft sentence introduced by *?inna*. The sentence, literally translated, begins

> EXTRAPOSED TOPIC   V
> 'The "human potential"   does not participate
> SUBJECT   COMP
> all of its components effectively in economic activity,
> CLEFT SENTENCE
> but *?inna* a part of them only, that is what participates . . . '.

The point of extraposing 'human potential' in the first of these coordinated
sentences is to reaffirm, after the preceding lengthy sentence, that this concept
continues to be the focus of the attention. In fact, the whole paragraph is con-
cerned with arriving at a definition of what it includes and what is excluded
from it. The second of the coordinated sentences ('but a part of them . . . ') be-
gins this process. If these nominals had not been moved to sentence-initial po-
sition, the slant of the text would have been subtly different, akin to the dif-
ference between the "neutral" nonclefted and "committed" clefted English
structures: "all of the components of the (total) human potential do not partic-
ipate . . . " versus "it's not all of the components, but only a part of those that
make up (total) human potential that participate . . . ," where the Arabic extra-
posed and clefted structures share the more emphatic tone of the second of the
two English sentences. Change of text topic—as opposed to emphasizing the
durability of the same topic—may also occasion the fronting of nouns, pro-

nouns, and even whole clauses. This type of fronting is signalled by the parti-
cle *?amma:* 'as for', which precedes the fronted element, and *fa*, which precedes
what is said about it.

We turn now to word order in spoken Arabic. It is often claimed, even in
works devoted to word order, that SVCOMP is the basic order in the dialects
just as VSCOMP is the basic order in MSA.[12] Both claims oversimplify reality.
As in MSA, word order in the dialects is determined by syntactic and discourse-
structure factors. Furthermore, there are differences in word order preference
from one dialect type to another, despite the leveling effects of MSA and in-
terdialectal contact. Broadly speaking, the Bedouin-type dialects of the eastern
Arab world exhibit a word order system that is more akin to that of CLA: in the
simplest form of verbal sentence containing V, a freestanding S and COMP,
VSCOMP is commonly employed where subjects have a dynamic or "event-
stating" predicate, whether S is definite or indefinite; SVCOMP, however, tends
to be used where the S is definite and the predicate is descriptive of a state or
circumstance. This pattern of usage is particularly true of uneducated speakers
and in all types of context, although par excellence in narratives.

In the urban dialects of the Levant, by contrast (and elsewhere in the east-
ern Arab world, to the extent that educated speakers throughout the region are
exposed through the media and dialect contact to these dialects), word order
depends on a combination of syntactic, rhythmic, and semantic factors. In Dam-
ascene, for example, which is not untypical of the Levant region as a whole, def-
initeness of S, "weight" of S (in the sense defined earlier), and the nature of ver-
bal predication—completed versus noncompleted aspect, and dynamic versus
stative predicate—all play a part in determining word order.[13] These factors, in
descending order of strength, seem to be as follows.

(a) Definiteness

VSCOMP is the norm for all sentences in which S is grammatically indefi-
nite, for example, *ša:fha rižža:l* 'a man saw her' (lit. 'saw-her a man'). This ap-
plies equally to MSA and Bedouin-type dialects, in which an indefinite subject
cannot occur in SVCOMP.[14]

(b) Predicate type, transitivity, and aspect

If S is a definite freestanding nominal, VSCOMP is more likely than SV-
COMP if V is "event stating"; that is, if it describes an action (rather than a state)
and one that actually occurred and was completed in the past (i.e., an s-stem).
The major categories of verb involved here are motion verbs such as *ra:ħ* 'went',
aja 'came', *nizil* 'descended', *wiṣil* 'arrived', *daxal* 'entered', and punctual transi-
tive verbs whose object is a bound pronominal enclitic, as in *ša:fu:ha rriža:l* 'the
men saw her' (lit. 'saw-her the men'), *ža:buh ?aḥmad* 'Ahmed brought it' (lit.

'brought-it Ahmed', *ʔaxadha lwalad* 'the boy took her' (lit. 'took-her the boy').
But if such a V has a freestanding object, the likelihood of SVCOMP is greater,
especially in cases where, contextually, an ambiguity might arise as to which
noun is S and which COMP.[15] Finally, if V describes a state or a habitual or con-
tinuing rather than a punctual completed action (i.e., cases of p-stem with as-
pectual prefix), SVCOMP is more likely.

(c) Weight of S

"Heavy" subjects, whether definite or not, may override the above constraints
and precede V, whether they are definite or not and whatever the type of pred-
icate. "Heavy" is of course a relative term: SVCOMP may be preferred even if
S has only two lexical items in the string, as long as its object has only one.
For example:[16]

lazʔa   ba:rde bitxaffef   lwažaʕ
compress cold   *b*-3fsg-relieve the-pain
'A cold compress will relieve the pain'

In many cases, the preference for SVCOMP with "heavy" subjects seems to be
consequent on the need to avoid a lengthy dislocation between an initial verb
and its complement(s). For example:[17]

ħayalla ra:dyo može ʔaṣi:re biži:b   lqa:hira bi   kill ishu:le
any   radio wave short *b*-3msg-bring Cairo   with all   ease
'Any short-wave radio will pick up Cairo quite easily'

To the above syntactic and semantic constraints on word order must be added
discourse-related factors. Word order in spoken Arabic is more flexible than in
written Arabic, in response to the unpredictable requirements of conversation.
Predicates of all kinds may, for many reasons (e.g., the expression of astonish-
ment, contradiction, or contrast), be placed before their subjects, and nominals
may be extraposed for the same reasons as in MSA. Spoken narratives are par-
ticularly interesting in these respects, and some illustrative examples of the nar-
rative uses of word order are given below. The elements filling the S, V, and,
where appropriate, EXTRA slots are capitalized in both the Arabic and English
translation for the sake of clarity (an ellipted item is in parentheses).

(a) Damascene

The following examples are taken from an oral narrative ("The Jew and the
Merchant's Two Sons") delivered apparently impromptu by an unlettered Dam-
ascene at the turn of the twentieth century.[18] In this story, VS(COMP) is reserved

virtually exclusively for "event-stating" (s-stem) predicates, whereas SV(COMP),
where V is p-stem or active participle, tends to be used for describing the cir-
cumstances that surround the main events but that do not advance the action:

VS(COMP) sentences (= "event-stating" predicates):

    V   S
KA:N   TA:ŽIR   u   luh   tne:n u:la:d
'There WAS once A MERCHANT and he had two sons'
    V   S
KAMAŠUH ELYAHU:DI wi ʔašar wiššuh
'THE JEW GRABBED HIM and flayed his face . . . '
    V   S
ILTAFAT IṢṢABI ʔa:l lha
'THE BOY TURNED AROUND (and) said to her . . . '
    V   S
ŠA:FTUH UMM ISHA:ʔ min eššuba:k . . .
'UMM ISHAQ SAW HIM from the window . . . '
    V   S   V   S
WIʔ Iʕ LWALAD min ʕala i:di wi INKASARET RAʔBTUH . . .
'THE BOY FELL from my grasp and HIS NECK BROKE . . . '
    V   S
AŽET UMM ISHA:ʔ wabu isha:ʔ . . .
'UMM ISHAQ CAME with Abu Ishaq . . . '

SV(COMP) sentences (= "circumstantial" predicates)

    S   V
HUWA ʔA:ʕID yo:m min eliya:m, aža wa:ħid yahu:di
'(while) HE WAS SITTING one day, a Jew came'
    S   V
we HINNE MAŠYI:N bi dderb . . .
'while THEY WERE WALKING down the path . . . '

(b) Eastern Arabian (rural) dialects

The following examples are from an unprepared narrative spoken by an il-
literate Bahraini Shiʕi woman.[19] The (true) story is a dramatic one of a fire in
the speaker's village. Here, there is an alternation between VSCOMP order for
sentences that carry the story forward or summarize the overall situation at a
given point, and SVCOMP or EXTRAVS sentences (in either case, sentences

with an initial, topicalized nominal) that spell out the result of that event as experienced by individual participants, or otherwise fill out the detail of the picture but do not advance the action (compare the use of SVCOMP in "circumstantial clauses" in the Damascene example):

(1)
(EVENT – overall description)

                      V   S

ma: nisma' illa wag'at iḍḍaw u YIGU:MU:N INNA:S fara:ra

'We heard the commotion of a fire, and EVERYONE STARTED rushing around'

↓

(DETAIL—deictic/specific focus)

  S   V   S   V

HA:DA YIFIRR YIŠI:L margad waladah, u HA:DI TIFIRR ibbitha . . .

'THIS ONE SNATCHES UP his son's bed, and THAT ONE her daughter . . .'

(2)
(OUTCOME—overall description)

  V   S

IFTAKKAT IL-'A:LAM min bala:ha:

'THE WORLD WAS RELIEVED of this calamity' (i.e., the village of the fire)

↓

(DETAIL—deictic/specific focus)

      EXTRA   V   S   EXTRA (V)   S

la:kin WAḤDA ḌA:'AT fiya:bha:, u WAḤDA   mirgadha: . . .

'but ONE WOMAN LOST her clothes, ONE (lost) her bed . . .'

(3)
(OUTCOME—overall description)

V   S   S   S

TAMMAT HAWA:YIJHUM, GDU:RHUM, KILLAH 'ala ssi:f

'THEIR THINGS, THEIR POTS AND PANS, ALL OF IT WAS on the beach'

↓

(DETAIL—deictic/specific focus)

  S   V   V   S   (V)

wi NA:S TIĠRIF min il-baḥar u TIKUTT 'ale:ha wi NA:S . . .

'and SOME PEOPLE WOULD SCOOP from the sea and POUR it (on the fire) and SOME OTHER PEOPLE WOULD . . .'

7.1.2 Word order and agreement

One oddity of MSA word order is that number agreement between a free-

standing S and its V depends on their position relative to each other in the sentence. Whenever V precedes a freestanding S, V agrees with S in gender but not number, so if V comes before S (as it most frequently does), V is singular, whatever the number of S. But if S comes first, V agrees in both number and gender. However, since all nonhuman plural nouns systematically take feminine singular agreement (see 6.1.2), this positionally determined difference in S–V agreement is neutralized where such nonhuman plural nouns are sentence subjects and becomes apparent only where S has a human referent that is plural or dual. One possible explanation for the anomaly in the agreement system lies in the different grammatical statuses of verb-first and noun-first sentences, a structural difference recognized by the native Arab grammarians. Since "event-oriented" (verb-first) sentences are the norm in sentences that have verbs, the absence of number marking could be seen as a consequence of the "event" rather than the "entity" being the focus of attention. The verb need only be in a morphologically simple form because nouns are in any case obligatorily marked as such if they are dual or plural, and the principles of word order already stated obviate any potential ambiguity concerning which of the following nouns is subject. As a matter of fact—and vestiges of this tendency can still be observed in some MSA literary styles today—in early CLA it was not even necessary for verbs in verb-initial sentences to agree with their subjects in gender: the morphologically simplest form of all (masculine singular) could serve for all subjects, whatever their number or gender. Gender and number marking was redundant, and optional, where V came first.

If VS can be considered the unmarked order in verbal sentences, SV becomes the marked one. The shift of a "known" entity to preverbal position is a means of focusing on or topicalizing it: of making the sentence "entity oriented." Thus, although a verbal sentence such as

(1) *ma:ta   rrija:lu*
    died-3msg the-men
    'The men died'

is "event oriented" and does not require number agreement, its marked equivalent

(2) *(?inna) rrija:la   ma:tu:*
    (TOP) the-men died-3pl
    'The men, they died'

does. Here, however, *rrija:la* can be considered a topic (optionally "presented"

by *?inna*) but not the grammatical subject of the verb: *ma:tu:* is after all a complete sentence, whose subject, 'they', is implicit in the plural enclitic *-u:*. That this is a plausible grammatical analysis is suggested by other topicalized structures such as

(3) (*?inna*) *rrija:la   ma:ta   θna:ni   minhum*
    (TOP)   the-men died-3msg two   of-them-pl
    'The men, two of them died'

where 'the men' is, perhaps more obviously than in (2), an extraposed topic, and not the grammatical subject of the verb. *ma:ta θna:ni minhum* 'two of them died' is a complete event-oriented verb-initial sentence in which the (singular) verb, as initial element, does not agree, as in sentence (1) with its (here dual) subject. At all events, MSA makes a consistent distinction between "event-oriented" and "entity-oriented" sentence types in terms of how VS number agreement works.

Unlike in MSA, agreement between V and S in the modern dialects is not dependent on word order: an initial V agrees with a following S in the same way as it does when S is initial.[20] This is not to say, of course, that in the dialects there is always "strict" agreement between S and V in the sense defined in 6.1.2, because, as we have seen, the "strict" or "deflected" agreement of a plural noun with a verb, adjective, or pronoun is dependent primarily on semantic factors: whether the noun is referring to human or nonhuman entities and whether the individual referents that comprise the plural noun are being viewed as acting *en bloc* or as individual agents. But the point is, relative position per se of V and S does not seem to affect agreement. Recent work on agreement variation in Cairene, however, suggests that *distance* (in phonological words) between a head and the locus of its agreement has a secondary effect on type of agreement, with "deflected" agreement being preferred where the two are close together, and "strict" (= "natural") agreement when they are far apart.[21]

7.2 Clause order and type

In 7.1 we were concerned with the syntax of simple sentences, viz. those containing only a main verb, a subject, and a complement. Here we will examine the structure of coordinated and complex sentences. In this exposition, a *coordinated* sentence is defined simply as one consisting of more than one simple sentence (= Sen), conjoined by one of a closed set of conjunctive particles. Thus the Arabic equivalents of "Her hair is black and her eyes are blue" and "I arrived and she smiled" are conjoined sentences (Sen1 *wa* ('and') Sen2) as they are in

English, although in Arabic the first would contain no verbs, whereas the second would contain only verbs, with no freestanding subjects. A *complex* sentence is one in which one or more of the nonverbal slots in the simple sentence that forms its substructure is filled by a string that could itself otherwise form a freestanding simple sentence, as defined in 7.1. We term a string occupying such a slot a dependent clause. The syntactic functions of dependent clauses correspond to those of the simple sentence slots—subject, object, adverb, adjective—which they fill. Dependent clauses are conjoined to main clauses by a variety of particles or function-specific phrases, in contrast to the more limited options for conjoining simple sentences into coordinated ones. Some types of dependent clause are morphologically marked as different from main clauses in both MSA and the dialects, although, as we noted in 6.2.2, by different systems: in MSA by differences in the final inflections on p-stem verbs and, in (most) dialects, by the absence of aspectual/modal prefixes.

It is noticeable that, compared with the older style of prose writing, in which a loose, syndetic coordination of short sentences was common, MSA has shown a tendency toward the use of complex sentences. Such sentences may contain several dependent clauses, linked by one or other of a large class of lexical phrases ("regardless of the fact that," "by virtue of the fact that," "in accordance with," etc.). In the written Arabic of the period before about 1850, the logicosemantic link between coordinated simple sentences was often less lexically specified than it was subsequently, heavy use being made of a few general-purpose particles that were capable of a wide variety of interpretations. The reader was often left to deduce the correct intersentential connection from the general context. This is now, in the twenty-first century, much less true of descriptive, ratiocinative, and nonnarrative writing generally, no doubt because of the desire to avoid potential ambiguity—at least on the surface of syntax—in expository writing of all kinds.

A further development in MSA sentence syntax, compared to older writing, is the greater flexibility in clause order. Certain types of clause—noun clauses functioning as subjects, for example—had been restricted positionally, so that the English order in, for example, "That the Iraqi attack on Kuwait was a crime is undeniable" was felt to be ungrammatical in Arabic, or at least very poor style (and still is by traditionalists). It still apparently feels more natural to most Arab writers to say, literally, "It is (lit. of that which is) undeniable that . . . ." There are both information-structural and "rhythmic" reasons for this preference (7.1): "that which is undeniable" is, to the Arab way of thinking, the "known" (and hence defined) entity and should therefore come first in the sentence, followed by the clause that describes what is informationally "new" (in that it is being

added to the already established category of "undeniable things"). Also, as we noted earlier, placing the lengthy dependent clause first offends against the rhythmic principle that requires that the main break in the sentence comes no later than about halfway through and, hence, that the "heavier" elements come later in the sentence. Notwithstanding the objections of the traditionalists, Arabic sentences structured like the English one are now frequently encountered in media writing and some modern literature, both of which have been influenced by foreign textual models.

Clauses that in CLA described the circumstances in which the agent or patient found themselves at the time of the action described in the main verb, such as that translated by "smiling at those present" in "He left the room, smiling at those present" were also restricted in the position they could occupy, following, never preceding, the main clause. Thus one wrote, literally translated, "He left the room *and* he is smiling at those present" but not "Smiling at those present, he left the room." This rule of clause order is now sometimes relaxed in MSA, for a combination of reasons. First, some dialects (Cairene, Damascene, Baghdadi, and Palestinian)[22] impose no restriction on the position of circumstantial clauses, nor any on the type of verb, stative or dynamic, that can occur in them. In other words, in the dialects, the syntax and semantics of circumstantial clauses and other types of dependent clauses, which are formally and functionally distinct from each other in prescriptive MSA, have begun to resemble each other. Second, CLA/MSA itself has always permitted the fronting of *dependent* clauses of time ("when . . . ," "after . . . ," "before . . . "). The fronting of a syntactically independent but logically subordinate sentence, especially if it contains a dynamic verb (e.g., "*And* he was crossing the road, he felt a sharp pain") can again be seen as a generalization based on the functional and semantic similarities of dependent clauses of time and independent sentences of contingent circumstance. Third, in the European languages with which Arabic is in most contact (English and French), clauses equivalent in function to the Arabic circumstantial clause have no positional restrictions. Thus, as in other areas of MSA, several factors—dialectal syntax, the structure of languages with which Arabic is in contact, and the process of internal analogical simplification—seem to have conspired to produce the same effect.

7.2.1 Coordinated sentences

The normal means of coordinating sentences (and other elements) in Arabic is syndetic—that is, by the use of conjunctions. So, for example, in a narrative description of a sequence of actions, Arabic typically links each sentence by a conjunctive particle. This is unlike the pattern in English, in which a combina-

tion of syndetic and asyndetic linkage is stylistically—even perhaps grammatically—required. Thus in English one says, and writes, "I put the cat out, locked the door, put the lights out, and went to bed," whereas in Arabic one normally says and writes "I put the cat out *and* I locked the door *and* I put the lights out *and* I went to bed." Asyndetic linkage in descriptions of sequences of events is alien to Arabic, although in the descriptions of people and entities and in general where there is no temporal sequence, asyndetic linkage between descriptive phrases is becoming more common.[23] Where, however, actions occur simultaneously, as in "He walked away, carrying the suitcase," asyndetic (as in the English sentence) or syndetic linkage is possible. In either case there is an aspectual contrast: s-stem for the main, event-stating verb, and p-stem, participle, or no verb at all for the second, which gives additional detail or describes the circumstances in which this main action occurred.

The conjunctive particles in Arabic are *wa, fa, θumma, ʔaw, ʔam,* and *bal.* These express very general relationships of conjunction ('and'), result ('so'), sequence ('then'), disjunction ('or'), and adversative emphasis ('(not X), rather Y') between units of language that vary in size from single words to whole paragraphs. Some particles, notably *wa* and *fa,* may express more than one of these relationships, as we shall exemplify. Other apparently conjunctive elements, such as the adversative *laːkin* 'but' do not, strictly speaking, form part of this set, because they require one of the properly conjunctive particles—*wa*—to precede them.

7.2.1.1 wa

wa is the primitive conjunctive particle: it is the most commonly encountered sentence connective and has the widest variety of uses, analogous in these aspects to English 'and'. Unlike English 'and', however, *wa* regularly functions as a textual connective and a sentence connective. In simple narratives, such as press reports containing twenty or more paragraphs, it is quite normal for *wa* to begin every paragraph except the first. The function of *wa* in such texts is simply to mark the beginning of the next episode in the report. Like *fa,* but unlike the other conjunctives, *wa* is not a freestanding word but a clitic that is prefixed to the first word of the coordinated sentence. *wa* basically *adds* (in descriptions of static scenes and objects) and *sequences* (in descriptions of actions) but can also indicate simultaneity of actions or a contrastive or adversative relationship between the clauses it conjoins. Below are some illustrative examples, from both modern written and modern oral sources. The examples have been literally translated, and in the literary examples, the punctuation of the original Arabic is preserved.

(a) static description (X and Y)

'. . . *wa* there were a few women, some of them revealing dainty arms that carried handbags resembling shoe or jewel boxes. *wa* there was not a single peasant woman among them.' (MSA written example).[24]

These two sentences from a modern novel form the second half of a descriptive paragraph and follow a period. It is typical that *wa*, the indigenous device for sentence concatenation, continues to be used alongside the period, which here is performing the same function of marking the end of one sentence and the beginning of another.

(b) narrative sequence (X then Y)

'They brought out the pot *wa* took the mashed dates *wa* threw them into the middle of the pot *wa* mashed them . . . mashed them until they were completely ready *wa* sieved away the water *wa* put them on to cook . . . ' (spoken example)[25]

Here, *wa* links successive episodes in a narrative.

(c) simultaneous action (X at the same time as Y)

'I water the crops *wa* eat' (spoken example)[26]

The speaker, a farm laborer, had just been asked at what time he stopped work for a bite to eat. His reply means that as a rule he did his watering and *at the same time* ate a snack, not that he watered and *then* ate, as the questioner had assumed. Compare this use of *wa* with the similar structure in the following:

(d) circumstance (X in circumstance Y)

'He abandoned them *wa* they were small' (spoken example)[27]

The difference between (c) and (d) is that (d) indicates the surrounding circumstances in which the main action 'abandoned' occurred, while in (c), neither the action of 'watering' nor 'eating' is given particular topical prominence.

Whether Sen2 in a Sen1–*wa*–Sen2 structure is to be given a sequential, simultaneous, or circumstantial interpretation depends not just on the speech context, but also on the semantics of the verbs in the conjoined sentences and on the syntactic structure of Sen2. Thus *qa:ma wa qahqaha* (lit. 'he got up and laughed') would normally be understood as meaning 'He stood up, laughing', that is, a simultaneous interpretation. A sequential interpretation of the same two verbs would be indicated by conjoining them with the more specifically sequential particle *θumma* 'and then': *qa:ma θumma qahqaha* 'He got up, and burst out laughing'. In cases such as *qa:la wa ṣa:ḥa* (lit. 'He said and shouted'), where the conjoined verbs refer to different degrees on a single scale (here of "vehemence"), the verb in Sen2 is understood as merely amplifying the verb in Sen1: 'He said, shouting, . . . '.

Syntactic structure also points to how Sen2 is to be interpreted relative to Sen1 in sentences conjoined by *wa*. If there is an aspectual contrast—s-stem verb in Sen1 + p-stem/participle/no verb in Sen2, or s-stem in Sen1 and *qad* + s-stem in Sen2—a circumstantial rather than a sequential interpretation is indicated. Compare

(a) *ʔinṣarafa rrajul   wa rakiba   ṣadi:quhu ddarra:ja*
    left-3msg the-man and mounted-3msg friend-his the-bicycle
    'The man left and his friend got on the bicycle'

where both verbs are s-stems (sequential or simultaneous interpretation), with

(b) *ʔinṣarafa rrajul   wa ṣadi:quhu yarkabu   ddarra:ja*
    left-3msg the man and friend-his 3msg-ride the-bicycle
    'The man left while his friend was riding (or 'getting on') the bicycle' (circumstantial interpretation)

in which the verb in Sen2 is p-stem and its subject comes before it. As we have already noted, the p-stem is typically used for background description rather than the stating of events, and the placing of S before V is also associated with a descriptive, rather than an event-stating, predicate. Furthermore, if the subjects of Sen1 and Sen2 are coreferential, the *wa* can be omitted (asyndetic linkage), and the p-stem verb *yarkabu* can be replaced by the active participle *ra:kiban* in the accusative case (that is, with adverbial function), viz.:

ʔinṣarafa rrajul   yarkabu/   ra:kiban   ddarra:ja
left-3msg the-man 3msg-ride/ ride-A.PART the-bicycle
'The man left, riding the bicycle'

These syntactic factors are indicative of the subordinate status of Sen2 vis à vis Sen1, both formally and functionally.[28]

Compare also the following (translated) pair of coordinated sentences, in which all the verbs are s-stem and that are identical apart from the particle *qad* preposed to the verb in Sen2:

(c) 'She fell silent *wa qad* closed her eyes'
(d) 'She fell silent *wa* closed her eyes'

Sentence (c) is equivalent to the English 'She fell silent, having closed her eyes', whereas (d) means 'She fell silent and (then) closed her eyes'. Sen2 in (c) is again circumstantial, and *qad* here indicates the "anterior circumstance" in which the action in Sen1 occurred, whereas in (d) the action in Sen2 would be construed as *following* that of Sen1.

We commented earlier that occasional examples are encountered in recent MSA writing of circumstantial clauses occurring before main clauses. Whereas in CLA the circumstantial clause was merely an adjunct that described "the state or condition either of the subject or one of the complements of that first clause or else a new subject,"[29] in the dialects it broadened its functions to include temporal, causative, or even conditional meanings. Like the dependent clauses that express these same meanings, it can precede as well as follow the main clause. For example:[30]

w   ana ra:žiʿ,   žtamaʿt bi   ʔabu nabi:l (Damascene)
and I return-A.PART, met-1sg with Abu Nabil
'When I got back, I met Abu Nabil'
w   iḥna xargi:n   ʔabilna   šwayyit šubba:n (Cairene)
and we go out-A.PART met-1pl little   youth
'As we were going out, we met a few youths'

These two dialectal examples illustrate the temporal function of preposed circumstantial clauses in the modern dialects. Two further recent examples, this time from MSA,[31] illustrate the temporal and the causative functions:

CIRCUMSTANTIAL CLAUSE
wa   naḥnu ʿala: maša:rif   niha:yati lqarni   l'išri:n
and we   on   approaches (of) end (of) the-century the-twentieth
MAIN CLAUSE
tataḥaqqaqu   ha:ðihi lmaqu:la   ʔakθar wa ʔakθar
3fsg-is-realized this   the-dictum more   and more
'As we approach the end of the twentieth century, this dictum is becoming more and more true'

CIRCUMSTANTIAL CLAUSE
wa   hadi:b wa šaʿla:n yaʿu:da:ni   bi ha:ða: ššakli   lmufa:jiʔ
and Hadīb and Shaʿlān 3md-return in this   the-fashion the-sudden
MAIN CLAUSE
ʔaθa:ra: mina lfaraḥi bi   miqda:ri ma: ʔaθa:ra: mina lmufa:jaʔa
caused-3md of   the-joy with amount what caused-3md of   the-surprise
'Hadib and Shaʿlan returning so suddenly caused as much joy as surprise'

The more "correct" way of phrasing both these examples would have placed the main clause first. The reason that the second example in particular is not so structured is clear: although, formally, 'Hadib and Shaʿlan returning so suddenly' is a circumstantial clause (introduced by conjunctive *wa* and with p-stem verb in contrast to s-stem main verb), functionally, its relationship to the main clause is not that of describing the circumstance surrounding the action 'caused joy'. Like a dependent clause (which could easily replace it), it explains the cause of the action expressed by the main verb, and the English translation reflects this.

(e) adversative relation (X but Y)

Without any adverbial support, *wa* may link two sentences that are overtly or implicitly mutually inconsistent or where the second implies a restriction or concession of some kind on the first, for example: ' . . . as if she is in the city *wa* she is not of it' 'I hear your tormented voice *wa* I don't see you' (MSA written examples).[32] Here *wa* signifies 'but at the same time'.

7.2.1.2 *fa*

The basic difference between *wa* and *fa* sentence connection in MSA, and in the dialects, is that *fa* usually betokens a relationship between Sen1 and Sen2 or between the paragraphs of a text, such that Sen2 describes a state or an action that occurs as a consequence of Sen1. Texts (1–5) illustrate this typical function of *fa*:

(1) 'I discovered from the first puff that smoke was escaping from lots of holes *fa* I stubbed it out in the ashtray' (MSA)[33]
(2) 'One day I heard a boy selling books who kept calling out "Diary of a Tough-Guy" *fa* I called him over and bought a copy' (written colloquial Cairene)[34]
(3) 'I think *fa* I am' (= Cogito ergo sum) (MSA)[35]
(4) 'The plane got lost *fa* crashed in the desert' (Damascene)[36]
(5) 'The sheikh entered *fa* we rose' (Gulf)[37]

Text (6) below is a good illustration of how *fa* and *wa* are used to signal different informational relationships between different parts of a text. *fa* marks the boundary between the assertions made in the first two sentences and the pieces of evidence that are adduced to support those assertions in the rest of it. *fa* further indicates that a logical relationship exists between these two parts of the text, whereas *wa* acts as a neutral link between the three pieces of evidence adduced, which are, by virtue of being so linked, not placed in any particular logical or other hierarchy:

(6) 'Men believe that they have the right to dominate women, and they treat them in accordance with this belief. The evidence for this is considerable.

fa It is not considered good manners within many families for the wife not to kiss the hand of the husband when greeting him, *wa* not for women to sit with the men, *wa* not the done thing for them to eat together.'[38] (MSA)

7.2.1.3 *θumma*

Like *wa, θumma* indicates sequenced action. The difference between the two, in both CLA and MSA, is that *θumma* marks a new development, event, or change of direction in the action described in the narrative. (*θumma* is used in the spoken dialects but only in relatively elevated style.) Compare the following two narrative literary texts, premodern and modern. In both, *θumma, wa,* and *fa* each have distinct functions: *θumma* acts as a superordinate staging marker for the narrative as a whole, *wa* adds information within each of the narrative frames thus created without taking the narrative forward, and *fa* introduces sentences that describe outcomes or results:

Text 1

θumma we put our trust in God, may He be praised and glorified *wa* we entrusted ourselves to Him *wa* we departed from Jurjānīya on Monday, 3 Dhū l-Qaʿda in A.H. 309 *fa* we stopped in a frontier post called Zamjān *wa* it is the entrance to the land of the Turks *θumma* we departed the next day *fa* we arrived at a place called Jit *wa* it snowed until it was up to the camels' knees *fa* we stayed in this place two days *θumma* we entered the country of the Turks.[39] [CLA]

There are thus three stages in this narrative, each introduced by *θumma*:
(i) *θumma* the journey from Jurjānīya to Zamjān
(ii) *θumma* the journey from Zamjān to Jit
(iii) *θumma* the crossing of the border into the land of the Turks

Within stage (i), the first three sentences, two metaphorical paraphrases, and one prosaic description all mean the same: 'we left'. All are linked by "additive" *wa* because they are merely in apposition to each other. The next sentence, introduced by *fa*, records the outcome of 'departure': 'arrival'. Next, a sentence introduced by *wa* simply—as it were, parenthetically—tells the reader what Zamjān is but does not advance the narrative: end of stage (i). In stage (ii) there is a similar pattern. *θumma* indicates a new stage in the action, 'departure', and *fa* again records the outcome, 'arrival'. The next sentence, introduced by *wa*, does not advance the narrative but describes something that happened coincidentally—a snowfall. The next sentence, introduced by *fa*, describes something that happened as a result of this snowfall. Finally, *θumma* introduces stage (iii), the resumption of the journey.

Text 2

I bought the book at a high price *θumma* I retreated to a corner of the coffee shop *wa* I started to turn the pages *fa* it turned out to be no more than phrases and conversations in English and their equivalents in Arabic *fa* I regretted the money I had spent on it *wa* asked myself what I should do with it.[40] [MSA]

θumma marks off the first event, the buying of the book, from the rest of the text = the retreat to the coffee shop and what happened there. *wa* links two events, the retreat to the coffee shop and starting to turn the pages, which follow temporally closely on each other. The first *fa* introduces a sentence that records the nature of the book—the result of 'turning the pages'—and the second sentence introduced by *fa* the closely connected results of this discovery = regret and (*wa*) self-questioning.

Whereas in narrative style there is relatively little difference in the patterns of sentence coordination and linkage between Classical and modern literature—apart, of course, from the now ubiquitous, if sometimes idiosyncratic, use of punctuation—the same cannot be said of expository style. In modern expository writing, three developments are noticeable: a greater variety of sentence connectives, a preference for dependent clauses over coordinated sentences, and a heavy use of adverbial phrases that indicate lexically the type of logical connection that the reader is to make between textual constituents. This is a far cry from the loose concatenation of unpunctuated sentences coordinated by the multifunctional connectives *wa* and *fa*. The following extracts illustrate this difference. Text 3 is a literal translation (the writer's) of a section of the *Muqaddima*, a fourteenth-century treatise on the philosophy of history, social structure, and economics. The passage quoted deals in abstract terms with the organization of a society's economy. The sentence connectives have been left untranslated. Text 4 is the standard published translation of the same passage for the English reader. In order to facilitate comparison, the texts have been divided into the same numbered sections, corresponding to stages in the argument, and the connectives and linking phrases in Text 4, the idiomatic translation, have been capitalized. Text 5, a modern text on housing economics, is provided for comparison.

Text 3

(1) *wa* the reason for this is that it is known *wa* it is established that one individual cannot satisfy his needs independently of others (2) *wa* that they all cooperate in their civilization to do this (3) *wa* what is produced through the cooperation of

a group of them is sufficient to satisfy the need of their number several times over (4) *fa* sustenance in wheat, for example, no one individual by himself can produce his ration (5) *wa* if six or ten are delegated to produce it including a blacksmith, and carpenter to make the tools, and an oxman, and a plowman, and a harvester, and the rest, *wa* they are distributed each to his task. . . .[41]

Text 4

(1) The reason for this is that, as is known and well established, the individual human being cannot by himself obtain all the necessities of life. (2) All human beings must cooperate to that end in their civilization. (3) BUT what is obtained through the cooperation of a group of human beings satisfies the need of a number many times greater [than themselves]. (4) FOR INSTANCE, no one, by himself, can obtain the share of wheat he needs for food. (5) BUT when six or ten persons, including a smith and a carpenter to make the tools, and others who are in charge of the oxen, the plowing of the soil, the harvesting of the ripe grain, and all the other agricultural activities, undertake to obtain food and work toward that purpose either separately or collectively. . . .[42]

One notices that in Text 4 the connectives *wa* and *fa* in Text 3 have had to be variously interpreted by the reader/translator in order to bring out clearly the structure of the argument. In Text 4, the *wa* at the beginning of the text has been omitted, its role being merely to signal that this forms part of the text that precedes it. In (2) *wa* has been replaced by a period: sentence (2) effectively makes the same point as (1), and the linking *wa* betokens no development of the argument. Sentence (3) proposes that the inference the reader might have made from (1) and (2), viz. that individuals when they act as a group can provide for themselves, is false: they can provide for more than themselves, hence the English translation 'but' for the *wa* that links (3) to (1)/(2) and that here must be interpreted as having an "adversative" sense. Sentence (4) then explains the truth of (3) by exemplifying and is linked to it by *fa*, translated by 'for instance'. Sentence (5), introduced by adversative *wa*, 'but', continues this example. Sentence (5) thus echoes in concrete terms the general statement delivered in (3). In other words, (5) is to (4) in concrete terms as (3) is to (1)/(2) in abstract terms. In both cases, the idea that there is a linear relationship between amount of labor and output is rejected.

This example illustrates the point that understanding CLA expository text can rely quite heavily on the reader's ability to interpret the relationship of juxtaposed sentences, because the conjunctive particles that link them have a range

of sometimes conflicting uses. Compare this now with a contemporary expository text—Text 5. Here, as in Text 3, the connective particles have been left untranslated, and additional conjunctive and other phrases that appear in the Arabic text and that help the reader to interpret the relationship between the sentences have been capitalized.

Text 5

wa BEYOND THIS, *fa* the difficulties of storing ready-made building materials, compared to other manufactured units, can be considered unique. *wa* ON ACCOUNT OF the large amount of space they take up *fa* storage costs are huge, WHILE warehouses cannot cope, EVEN THOUGH they have a large amount of space available. *wa* ACCORDINGLY *fa* sound production policies are those that are based on the selling of the products as they are produced SO THAT the problem of tying up invested capital is avoided. . . .[43]

Here, all six occurrences of connective particles (three *wa* and three *fa*) are actually redundant: punctuation, or the capitalized phrases, between them perform the identical functions of chunking the text and making explicit the logical relationships between the chunks. Although it does not seem to be true, as has been claimed,[44] that *wa* no longer functions as a punctuation device in narrative text, it does appear that expository writing in Arabic may be at a stage of transition from the older form of sentence coordination by conjunctive particle to one that relies mainly on punctuation and linking phrases.

7.2.1.4 *ʔaw* and *ʔam*

These two particles are used for disjunctive coordination and function like English 'or'. *ʔaw* is used in affirmative or interrogative sentences but not in negatives and may coordinate elements at any level: verbs, nouns, or whole sentences. It is used for synonymous or near-synonymous choice. For example:

'If the ship rises *ʔaw* falls *ʔaw* sways . . .'[45] [MSA]
'Have any letters come for her *ʔaw* has anybody asked about her?'[46] [MSA]

ʔam is used in interrogatives only, whether direct or in reported speech. In contrast to *ʔaw*, *ʔam* normally presents possibilities that are actually, or presented as being, mutually exclusive:

'Were you born a Christian *ʔam* did you embrace the religion as an adult?'[47] [MSA]

Where the coordinated sentences are negative, *wa* is used in combination with the negative particle *la:* (= 'and not'), equivalent to English 'nor':

'For an Arab is not (*la:*) better than a Persian because he is an Arab, *wa la:* the Persian better than an Arab because he is a Persian.'[48] [MSA]

In some dialects, there is a similar distinction between synonymous and non-synonymous choice, whether at the level of nouns, verbs, or clauses, although the distinction is carried by different particles in different dialect areas. In Damascus, for example, as in MSA, *?aw* is used for synonymous alternation, and *yamma* or *willa* (< CLA *wa ?illa:* = lit. 'and if not') is used for real alternatives. For example:

'The city is divided into a number of sections *?aw* quarters.'[49] [Damascene]
'Shall I put hair tonic on, *willa* just water?'[50] [Damascene]

willa (var. *walla*) as exemplified in the last sentence has very wide currency (e.g., Cairene, Tunis, and Gulf) and is also normally used for presenting a choice between a command and the consequences of not obeying it. For example:

'Shut up *willa* I'll hit you.'[51] [Damascene]
'Tell me the truth *willa* get out!'[52] [Gulf]
'Drive to the airport *walla* I'll kill you'[53] [Cairene]
'Shut up *walla* go home.'[54] [Cairene]

In all the dialects, negative alternatives are formed exactly as in CLA/MSA with *wa la:* 'nor' or its dialectal reflex. For example:

"He would not see her *wi la:* would she see him.' (Gulf)[55]

Polysyndetic coordination ("either . . . or . . . ") is syntactically similar in MSA and the dialects. In CLA/MSA, *?imma: . . . wa?imma:* or *?imma: . . . ?aw* are used, whereas in the dialects there are various possibilities, all involving *?imma/ ?amma/ ?umma* (< CLA *?imma:*) and either *ya* or *(?)aw: ya . . . ya* (Cairene and Levantine), *yimma . . . ?aw* and *?imma . . . ?aw* (Damascene), *yimma . . . yimma* (Cairo), or *?umma . . . aw* (Gulf). For example:

"The Koran was *?imma:* recorded on such surfaces *?aw* retained in people's memory'[56] [MSA]

'*?imma* you get out of here *?aw* I'll shoot you'[57] [Damascene]
'*yimma* you return the book *yimma* you pay the fine'[58] [Cairene]
'*?umma* we sit at home reading *aw* we go to the market . . . '[59] [Bahrain]

7.2.1.5 bal

The function of the MSA coordinating particle *bal* is simultaneously affirmatory and adversative: it affirms the truth of the proposition in Sen1 but at the same time introduces another in Sen2 that gives a more accurate, appropriate, or detailed description of the same state of affairs. It has various contextual equivalents ranging from '(nay) rather' to 'moreover' or *a fortiori*:

"An historian is not simply a translator *bal* he also examines the translations in order to ascertain the truth contained therein.'[60] [MSA]
'What I think, *bal* what I'm sure of, is that this devil loves her.'[61] [MSA]

Unlike most of the other CLA/MSA coordinating conjunctions, which have survived in the dialects with little change in their form or function, *bal*, like *θumma*, does not seem to have survived historically as a true dialectal particle. When it does occur in spoken dialectal Arabic as an "importation" from MSA, the content is usually abstract, technical, or literary and the context formal. For example:

'The spot does not disperse *bal* builds up.'[62] [science] [Cairene]
'This problem was not confined to Damascus alone, *bal* was widespread throughout the Arab world.'[63] [sociology] [Jordanian]

It is worth noting here that there are a number of coordinating conjunctions in the dialects that either do not occur in CLA/MSA or have a different status in written Arabic. *la:kin* 'but' occurs in all varieties of Arabic from Morocco to Oman but in CLA/MSA is normally preceded by *wa* (and hence is not strictly speaking a coordinator). In the dialects *la:kin* is not necessarily preceded by *wa*. Just as widely used as *la:kin* in the whole of the eastern Arab world is the coordinating particle *bass* 'but, except that . . . ', which has no MSA counterpart.

From this brief review of the syntax of simple sentence coordination in modern Arabic, it should be clear that the formal differences between the MSA and the dialectal systems are small: *wa, fa,* and *?aw* are found in most styles of Arabic from the plainest vernacular to the most high-flown MSA and perform very much the same functions in all.[64] The same is not true of some types of complex sentences, the matter to which we now turn. Sociolinguistically signifi-

cant differences can be observed in the extent to which subordination and co-ordination, as alternative means of clause linkage, are used between, on the one hand, written MSA and careful educated speech, and, on the other, relaxed or uneducated speech. Uneducated/relaxed speech generally contains relatively more coordinated and relatively fewer subordinated sentences compared with educated speech or MSA.

7.2.2 Complex sentences

In terms of the function they perform within the complex sentence, four types of dependent clauses can be distinguished: nominal, adjectival (or relative), adverbial, and conditional. However, these are merely convenient descriptive labels that indicate broad functional divisions; from a structural point of view, they obscure many similarities, for example, between some types of nominal and adverbial, and some types of adverbial and conditional clauses.

7.2.2.1 Nominal clauses

In CLA/MSA, nominal clauses are mostly syndetic, that is, they are linked to the main clause by a complementizer, although there are some specific verb categories (e.g., inchoatives) that govern a following clause directly, for example, *ʔaxaða yaṣiːḥu* 'he began to shout' (lit. 'he took he shouts'). The complementizers *ʔanna* or *ʔan*, equivalent in function to English 'that', are used for noun clauses in the form of statements, *ʔin, ʔiða:*, or *law* 'if, whether' for reported questions. If the governing element in the main clause—usually a verb—governs its complement with a preposition, then this is simply preposed to *ʔan* or *ʔanna*, for example, *raġiba fiː ʔan* 'to want that' (lit. 'to want in that'), *ʔaṣarra ʕalaː ʔan* 'to insist that' (lit. 'to insist on that'). If the following clause is a reported question, *maː* 'what' is attached to the governing preposition: *saʔalahu ʕammaː ʔiða: . . .* 'he asked him if . . . ' (lit. 'asked him about what if . . . '). A nominal clause may occur in either the subject or complement slot of the simple sentence, although there are restrictions and stylistic preferences in either case.[65]

In MSA, *ʔanna* is followed by an S(V)COMP and *ʔan* by VSCOMP or VCOMPS. *ʔanna* requires that the S of the governed clause be in the accusative case if it is a noun and that it be suffixed directly to it if it is a pronoun. The distinction between the two complementizers is that the governed clauses they introduce differ from the point of view of their factuality and aspect. If the action/state described is in the realm of the factual, whether complete (i.e., perfective) or incomplete (e.g., habitual or ongoing), *ʔanna* + S(V)COMP is the norm, although *ʔan* + VSCOMP is possible, provided that V is an s-stem (i.e.,

denotes a factual, completed action). If, however, the action/state described is in the future relative to the main verb or is nonfactual (and by definition incomplete), it is expressed by *ʔan* + VS(COMP) where V is p-stem a-set, that is, subjunctive. In the dialects, parallel distinctions of aspect and factuality are reflected in the form of the verb in the governed clause, although different morphosyntactic means are used. Further, in the dialects, the distinction between *ʔanna* and *ʔan* has been lost: a single complementizer, whose form differs slightly from dialect to dialect, serves to introduce both types of noun clause. (See the numerous MSA and Cairene examples of verbs in governed clauses in 6.2.2 above.)

As we saw in 7.1, the commonest word orders for simple verbal sentences in MSA are VS(COMP) (narrative) and SV(COMP) (descriptive/expository), with VCOMPS occurring as a minor variant of VSCOMP in certain well-defined circumstances. Verbless sentences normally have the structure SCOMP. In the main VS(COMP) and SV(COMP) sentence types, noun clauses introduced by *ʔan/ʔanna* clauses occur more frequently in the COMP slot than they do in the S. Where they do occur as S, they seem to be much more frequently encountered in V(COMP)S or VS(COMP) than in SV(COMP) sentences. The typical case, in both MSA and the dialects, is where S is a noun clause and V is an impersonal verb, for example, a modal verb with *ʔan*: 'It is possible for me to go' = lit. 'is-possible-me that I go' and an impersonal verb with *ʔanna*: 'It became clear that the company was bankrupt' = 'became-clear that the bank bankrupt.' Similarly, in sentences without a main verb, COMPS, not SCOMP, is the normal order if S is a noun clause, for example 'It is clear that he is wrong'= lit. 'of the-clear that he wrong'). Given the information-structure principle that what is known comes first (in this case the category of "clear things") and the rhythmic principle, already noted, that the optimal break in the Arabic sentence should occur no later than halfway through its length and hence that the heaviest element comes last, it is not difficult to see why there should be this distributional imbalance: noun clauses occurring as S in SVCOMP or SCOMP structures (assuming that COMP is not also a clause) would tend to produce sentence structures that offend against both principles. Nonetheless, SVCOMP sentences with S as a noun clause do occasionally occur (in MSA) but only where the information contained is contrastively foregrounded. Such foregrounded clauses are marked as such by *ʔamma* 'as for'. For example:

'*ʔamma:* that (= *ʔan*) its literature should have its own stamp, and its art its own personality, is the law of nature.'[66]

Here, the writer had been claiming that Egypt was an indivisible part of the Arab world in all respects; in the sentence quoted, he counterbalances this claim with a statement about the distinctive "Egyptianness" of the content of its art and literature. But in the normal case where a noun clause occurs as S and there is no special emphasis on it, there is an overwhelming preference for constructing the sentence so that the noun clause occurs late in the sentence. So, for example, the English 'For him to do this is important' (like the example cited, a sentence in Arabic without a main verb) becomes, literally:

COMP   S
of the-important that he does this

Noun clauses can be converted into verbal nouns. Thus for 'we were surprised at (the fact) that he went to Paris' one can equally well say 'we were surprised at his going to Paris' where 'his going to Paris' is expressed by a verbal noun functioning as the S of an active main verb (lit. 'surprised-us going-his to Paris') or as the COMP of a passive or mediopassive construction (lit. 'we were-surprised at going-his'). Alternatively, in either case, a noun clause introduced by ʔan or ʔanna may be used and, of course, allows factuality and aspect distinctions to be made that are obscured by the verbal noun construction: ðaha:buhu 'his going' can variously imply, as it can in English, 'he went', 'he will go', 'he may go', or 'his habitual going'. The choice in Arabic between nominal and verbal constructions in such cases is motivated often not so much by the desire to avoid ambiguity but by a feeling of what is stylistically appropriate. Expository texts in MSA in particular show a high incidence of complementation with verbal nouns rather than nominal clauses compared to, for example, narrative texts (see 8.3.3). In relaxed spoken Arabic, however, the opposite is the case, with noun clauses containing verbs being commoner. Consequently, the frequent use of the verbal noun in speech is heard by native speakers as an attempt to sound "formal."[67]

The syntax of nominal clauses when in COMP position is another feature that distinguishes the language of writing from speech: CLA/MSA favors syndetic linkage in many types of clause where the dialects favor asyndetic. The semantics of the main verb holds the key. All varieties of Arabic, spoken and written, favor syndetic linkage if the main verb denotes some form of cognition or perception, but this is not true where the main verb has a modal or desiderative meaning. In the latter case, the dialects favor asyndetic "verb strings" whereas CLA/MSA requires syndetic linkage. The following literally translated examples have been selected from short stretches of unscripted educated con-

versation[68] to illustrate this difference. In MSA, all of these examples would require a complementizer. The dialectal complementizer is represented by (that), its absence by (φ):

syndetic linkage (cognitive, perceptual main verb):
(S1) 'You *observe* (that) there is interference . . . ' [Palestinian]
(S2) 'I *understand* (that) you are a curriculum director' [Cairene]
(S3) 'She *saw* (that) God was angry with her' [Bahraini]
(S4) 'I *heard* (that) people were thrown overboard' [Bahraini]
(S5) 'I can *make out* (that) this radio station is Tunis' [Damascene]
(S6) 'I *hypothesized* (that) this . . . ' [Damascene]
(S7) 'I *think* (that) they want to discuss the details' [Jordanian]
(S8) 'You *know* (that) public parks are for everyone' [Jordanian]

asyndetic linkage (noncognitive, desiderative/modal main verb):
(A1) 'I *could* not (φ) I distinguish it' [Damascene]
(A2) 'God doesn't *order* you (φ) you go' [Bahraini]
(A3) 'I *wanted* (φ) I scream' [Bahraini]
(A4) 'They *like* (φ) they let them unfurnished' [Damascene]
(A5) 'What are you *intending* (φ) you do?' [Jordanian]
(A6) 'They *are forced* (φ) they talk to the child in dialect' [Palestinian]
(A7) 'You don't *have the right* (φ) you go' [Bahraini]

On inspection, it might be thought that the distinction between the "S" (syndetic) and "A" (asyndetic) sentences is not so much a semantic as a syntactic one of clause–subject coreferentiality: syndetic (S1–8) sentences lack it, whereas asyndetic (A1–7) have it. This turns out not to be true. Syndetic sentences with coreferential subjects such as

'I *felt* (that) I was at home with it' [Jordanian]

also occur in the same conversational transcripts, as do asyndetic sentences with different subjects, such as

'Are you *content* (φ) I record you?' [Bahraini]

The determining factor is thus not subject coreferentiality. Furthermore, spoken examples can occasionally also be found of syndetic linkage with modal/desiderative main verbs. For example:

'I am able (*ʔastaṭiʕ*) (that) I understand' [Jordanian]
'I *like* (that) I use my time better' [Jordanian]

and of the converse, asyndetic coordination with a main verb of cognition. For example:

'I think [same verb as in (S7)] (ϕ) there is a shared Classical language' [Jordanian]

In fact, the patterning of the data suggests that Arabic speakers—at least those from the eastern Arab world—make a dialectal distinction in the syntax of verb complementation when speaking informally that is semantically motivated but that is subject to variation as speech style becomes more elevated. If the level of speech formality rises, and the speaker begins to choose lexical elements normally associated with writing, as in the quoted example containing the stylistically elevated *ʔastaṭiʕ* 'I am able', in preference to dialectal *ʔaʔdar*, there may be a concomitant switch to the MSA system of syndetic linkage in association with all verbs that have that potentiality, including those modal verbs (and other modal expressions) that are normally asyndetic in the dialect (see 9.4. on co-occurrence phenomena).

7.2.2.2 Adjectival clauses

Adjectival, or relative, clauses, are placed directly after the noun they refer to, like attributive adjectives. They can be restrictive or nonrestrictive, as in English. In MSA, a relative adjective that agrees with its antecedent noun in number, gender, and case, is required in most cases[69] if the noun is grammatically definite (see 6.1.1); if it is indefinite, there is no relative pronoun and the adjectival clause follows the noun asyndetically. For example:

ʔalqiṣṣatu llati: la: tantahi:	*qiṣṣa la: tantahi:*
the-story which-f not 3fsg-ends	story not 3fsg-ends
'The story that never ends'	'A story that never ends'

The formal similarity between definite noun phrases of the type *ʔal*-X *ʔal*-Y where Y is an adjective and the structure of adjectival clauses that modify a definite antecedent noun is not fortuitous. Formally, the CLA relative adjective seems originally to have been a demonstrative, composed of the definite article *ʔal-* plus a *t-* or *ð-* deictic morpheme (see 5.2.1). Functionally, there is also a parallel with the defined noun phrase: just as *ʔalbaytu ljami:lu* 'the beautiful house' can be construed as 'the house, the beautiful (one)', so the example above is 'the

story, the (one) that never ends'. Adjective clauses are freely convertible into adjectival phrases (see below). As we shall see later in this section, in some dialects the syntax of the definite noun phrase and that of adjectival clauses has developed to a point of virtual merger.

Adjectival clauses are considered by Arab grammarians to be complete sentences in themselves, and the relative adjective is in fact merely a "link" (Ar. *ṣila*) between it and an antecedent noun in the main clause. That this is so is indicated by the fact that if the antecedent of the adjectival clause is the object of its verb, a so-called resumptive pronoun refers back to the antecedent, agreeing with it in number and gender:

ʔalqiṣṣatu llati: qaraʔaha:	*qiṣṣa qaraʔaha:*
the-story that-f read-3msg-it-f	story read-3msg-it-f
'The story that he read'	'A story that he read'

Restrictive adjectival clauses containing a verb may be converted into participial phrases without any syntactic constraint. Thus, the following both mean 'the countries that are beginning to develop':

ʔadduwalu   llati: taʔxuðu   fi: nnumu:
the-countries that-f 3fsg-begin in the-development
ʔadduwalu   lʔa:xiða   fi: nnumu:
the-countries the-begin-A.PART in the-development

In cases where the subject of the adjectival clause is not coreferential with its antecedent, the use of conversion structures, unlike in English, is quite routine, so that an antecedent noun + adjective clause structure such as:

li l'umma:li   llaði:na tanxafiḍu rawa:tibuhum
to the-workers who-mpl 3fsg-fall wages-their
'. . . to the workers whose wages are falling'

can be converted to, as it were, 'to the falling-waged workers':

li l'umma:li   lmunxafiḍati   rawa:tibuhum
to the-workers-GEN the-fall-A.PART-f-GEN wages-NOM-their

In this typical conversion structure, the active participle *munxafiḍa*, which is actually the predicate of *rawa:tib* 'wages', is attracted to the antecedent masculine

plural noun *'umma:l* 'workers', with which it is made to agree in definiteness and case, but retains its feminine marking to agree with its underlying subject, the noun *rawa:tib* 'wages'. If the antecedent is the object of the verb in the adjectival clause, a similar conversion structure is possible with the passive participle, which follows the same pattern of "mixed" agreement. Both of the following, for example, mean 'the woman who was sentenced to death':

?almar?atu llati: ḥukima   *'alayha: bi l?i'da:m*
the-woman who-f sentenced-PASS-3msg on-her   the-death

?almar?atu lmaḥku:m   *'alayha: bi l?i'da:m*
the-woman the-sentence-P.PART-m on-her   the-death

Clause conversion of this kind is especially common in MSA expository text (e.g., economics, sociology, and law), where it can become syntactically quite complex. For example:

?almu'addala:tu lmu'tarafu   *biha: wa lmuttafaqu*
the-rates   the-acknowledge-P.PART to-them and the-agree-P.PART

'alayha: musabbaqan
on-them previously
'The previously acknowledged and agreed-on rates'

Adjectival clauses in the dialects work in a similar way to the way they do in MSA. The main formal difference is that the relative adjective (in truth an adjective/pronoun) has everywhere become a syntactically invariant, uninflected form that is the same for all genders and numbers: *illi* (Cairo and Baghdad), *halli* or *yilli* (Damascus), and *illadi* (some Gulf dialects). Given that it is dialectally so widespread (and that the morphosyntax of its MSA analog is so complex), *illi* and its variants, like the expression of number, is one of those nonstandard features that tends to maintain its ground even in relatively formal educated spoken Arabic (see 9.4). It is also used as a nonattributive pronominal of general or vague reference, equivalent to English 'what(ever)', 'who(ever), 'that which', '(the one) who', etc. In this substantival sense, MSA sometimes employs the relative adjective/pronoun but tends to use *ma:* 'what' and *man* 'who' in cases where the reference is indefinite. For example:[70]

attributive:
ilmada:fi' halli biddill 'ala niha:yt iṣṣala: [Damascene]

'. . . the cannons **that** signal the end of prayer'
(. . . *?allati:* . . . in MSA)

we:n iṣṣandu:g illi ča:n 'a lme:z [Baghdadi]
'Where is the box **that** was on the table?'
(. . . *?allaði:*) in MSA)

inna:s illi jaw [Bahraini]
'. . . the people **who** came'
(. . . *?allaði:na* in MSA)

nonattributive:
bta'rif halli ḍarabak [Damascene]
'Do you know **the one who** hit you?'
(. . . *man* or *?alladi: ḍarabaka* in MSA)

illi ma: y'arf iṣṣagar yišwi:h [Bahraini]
'**Whoever** doesn't know what a falcon is roasts it' (proverb)
(*man la: ya'rifu* . . . in MSA[71])

illi yigu:la yijri [Baghdadi]
'**What** he says goes'
(*ma: yaqu:luhu* . . . in MSA)

There are two widespread dialectal innovations that are worthy of note, because they show that simplifying reductional tendencies are at work in dialectal syntax just as they are in dialectal morphology (see chapters 3 and 4). First, in many parts of Lebanon, Syria, and Palestine,[72] the attributive relative adjective may be used whether the antecedent noun is definite or indefinite. Second, in the same dialects (and in Baghdadi also[73]), the *il-* of a definite antecedent may be dropped, and the relative adjective reduced from *illi* to *il-* (Lebanon *el-*). The result is a surface merger of two structures that in CLA (and MSA) were separate but that serve similar functions: the definite construct phrase (6.1.3) and the adjectival clause. For example:

di:na:r iṣṣirafta . . . [Baghdadi]
dinar the-spent-1sg-it
'The dinar (that) **I've spent** . . . '

t'a:l li 'indi   *sa:'t elbetrid* [Lebanon]
come to with-me hour the-2msg-want
'Come to my place at the time (that) **you want**'

Thus, in these two examples, the noun and its adjectival clause *di:na:r iṣṣirafta . . .* and *sa:'t elbetrid* are surface-structurally identical to a definite construct phrase in the same dialects (noun + def. article + noun) but with the verb of the clause occupying the place of the second, "annexing" (semantically "specifying") noun. Three points of convergence between the two structures can be identified: (i) the definite article has been dropped from the (as it were) "annexed" nouns *di:na:r* and *sa:'a*; (ii) the *l* of the reduced relative adjective *il-* (as it were, the "definite article" of the annexing element) in the first example has assimilated to the *ṣ* of the "noun substitute" exactly as if it were a noun; (iii) the *t* at the end of the feminine noun *sa:'a* 'hour' is sounded—this is normal in the dialects only where such nouns occur as the first element of a construct phrase and is absent when followed by a true adjectival clause introduced by *illi/halli*. Furthermore, in Baghdadi Arabic, this constructlike adjectival clause varies freely with another "specifying" construction, the definite noun–adjective phrase. Here, a verb replaces the adjective. For example:[74]

ššaxṣ	*ilxa:bar*	*ma: ga:l*	*isma*
the-person	the-called-3msg	not said-3msg	name-his

'The person who called didn't leave his name'

What all these diverse constructions—the construct phrase, the noun–adjective phrase, and the adjectival clause—have in common is that they specify and limit the reference of the head noun in the construction. It is this similarity of function that seems to be leading to a formal convergence of these structures in many (eastern) dialects.[75]

7.2.2.3 Adverbial clauses

7.2.2.3.1 Time clauses

The primitive (in the sense of oldest) conjunctive particles of time are *lamma:*, *ʔi*ð*, and *ʔiða:* 'when', *munðu* 'since', and *ḥatta:* 'until'. In CLA, the difference between the first two of the conjunctions meaning 'when' is that a clause introduced by *lamma:* describes the state that has resulted from an anterior completed event and provides the context for the (foregrounded) event in the main clause: hence *lamma:* clauses invariably precede the main clause. *ʔi*ð, however, refers to the inception (often sudden) of an event, which is not anterior to the action in the main clause but simultaneous with it: it normally follows the main clause. Compare, for example, the CLA examples 'When (*lamma:*) we had struck (s-stem) him with our swords, Abdallah drove (s-stem) his sword into his stomach',[76] and

'You did them an injustice when (*ʔi*ð) you imprisoned them'.[77] *lamma:* invariably, and *ʔi*ð normally, were followed in CLA by an s-stem verb. In contrast to the narrative function of *lamma:* and *ʔi*ð, *ʔiða:* had a sense in CLA of 'if-and-when', 'whensoever'—that is, it envisaged an eventuality about which, because of habit, experience, or logic, there can be no doubt and on which the action in the main clause is dependent, for example 'When(ever) he enters a house he prays where he likes'.[78] Here, as in conditional clauses, the action in the adverbial clause is envisaged as completed, hence the use of the s-stem.

The CLA use of these conjunctions and their associated verb forms has been preserved to some degree in MSA, but there have been stylistic refinements and a blurring of distinctions with adjacent semantic areas. *lamma:* in contemporary writing has a somewhat "literary" flavor and is more likely to be encountered in a novel than in a newspaper report, where compound conjunctions that are felt to have more precision (*ba'dama:* 'after', *'indama:* 'at the time when') are commoner. *ʔi*ð has taken on an explicative sense of 'for, since', as can be seen in examples such as[79] 'But Heaven's help arrived, for (*ʔi*ð), at that moment, a man came in' where *ʔi*ð both explains the nature of 'Heaven's help' and marks temporal sequence. The latent conditional function of *ʔiða:* has become more pronounced in MSA, for example,[80] 'If (*ʔiða:*) we examine (s-stem) it comprehensively, we find (s-stem) . . .'; 'Go to her if (*ʔiða:*) you wish (s-stem)'. The nonconditional use of 'if' (= 'given that') in English also has an exact parallel: 'If (*ʔiða:*) Muslims consider Muhammad the great messenger of God. . . . so the Arabs consider him . . .'.[81]

In the dialects, reflexes of *lamma:* and *ʔiða:*, but not *ʔi*ð, have survived (an exception is Morocco, where *lamma:* as a dialect form disappeared), although with more sharply differentiated ranges of use. *ḥatta:* also survived, whereas *munðu* was replaced by an extension of the meaning of *min* 'from'. In the present-day dialects *lamma(n)* covers the temporal senses of all three of the CLA conjunctions—*lamma:*, *ʔi*ð, and *ʔiða:*—noted above and in many dialects (Cairene, Baghdadi, and Gulf) has acquired an additional sense, equivalent to English 'until' (possibly < *ʔila: ma: (ʔan)*). *lamma(n)* may be used to refer indiscriminately to completed or future actions and states, with the appropriate dialectal s-stem or p-stem verb form. The shift that occurred in MSA (compared to CLA) in the scope of *ʔiða:* has gone even further in the dialects, its dialectal reflex *iza* having completely lost its temporal sense and become one of the two main means of introducing conditional clauses.

In addition to the basic set of morphologically unanalyzable conjunctions described above, there is a set of time conjunctions in all varieties of Arabic formed from compounds of prepositions plus *ʔan* or *ma:* 'that'. These prepositions are themselves originally mostly nouns denoting some relational aspect

of time or place, and in MSA they are marked with the final -*a* of the accusative (adverbial) case, for example, *qabla ʔan* 'before' < *qabla* 'before, in front of', *baʿda ʔan* 'after' < *baʿda* 'after', *ḥi:nama:* '(at the moment) when' < *ḥi:n* '(point of) time', *ʿindama:* 'when' < *ʿinda* 'with, at', *ʔawwala ma:* 'as soon as' < *ʔawwal* 'first', *baynama:* 'whereas' < *bayn* 'interval', etc. The -*ma:* compounds have been particularly productive in the dialects, for example (Bahraini) *xalfma:* 'after' < *xalf* 'latter part', *ʿugubma:* 'after' < *ʿugub* 'end part', *yo:mma:* 'when' < *yo:m* 'day' (and in nontemporal clauses *šikilma:* 'like, just as' < *šikil* 'shape', *gaddma:* 'as much as' < *gadd* 'amount').

With the exception of *qabla ʔan*, which is followed by the p-stem a-set (subjunctive) because it always refers to actions/states that are in the future relative to whatever action is described in the main clause, all these conjunctions may be followed by verbs denoting completed (s-stem) or nonrealized/notional actions (p-stem). In contrast to *lamma:*, clauses introduced by . . . *ʔan* or . . . *ma:* conjunctions are flexible as to position, but there is a noticeable preference in MSA for placing them before the main clause. Although, as has been pointed out, *qabla ʔan* . . . and *baʿda ʔan* . . . adverbial clauses are structurally nothing more than prepositions followed by noun clauses, their greater flexibility of sentence position compared to noun clauses that are the subject or object of verbs (see 7.2.2.1 above) suggests that in MSA they are treated as if they have a different syntactic status from noun clauses.

In the dialects, the compound conjunctions used are usually those formed with -*ma:*. The verb forms that follow them may be s-stem, p-stem, or *b(i)*- prefixed p-stem, depending on aspectual/factual considerations. Thus *baʿadma: taʿašše:na* (s-stem) 'after we dined' implies that 'our dining' is a completed (past) action, as in 'after we dined, we went to bed'; *baʿadma: nitʿašša* 'after we dine' (p-stem) looks forward to a situation not yet realized, as in 'after we dine, we'll play cards', whereas *baʿadma: bnitʿašša* (*b(i)*- prefix) would be used to describe, for example, habitual actions, as in 'after we dine, we go for a walk'.[82] These distinctions apply, with minor local variations, to all of the eastern dialects, including the Bedouin ones (although here the p-stem distinctions do not exist in all cases, because in many "Bedouin" dialects there is no comparable system of prefixes to mark mood/aspect).

7.2.2.3.2 Purpose, result, and reason clauses

A variety of conjunctive particles are employed more or less interchangeably to introduce purpose clauses, the commonest in MSA being *li, liʔan* (neg. *liʔalla:*), and *ḥatta:*, all meaning 'so that, in order that'. *(li)kay* and its negative counterpart *(li)kayla:* are less common and have a somewhat "high-flown" literary fla-

vor. All of these conjunctions require their verbs to be p-stem a-set (the subjunctive), for example, 'The emir was burning with a desire for (*liʔan*) the man to finish (p-stem a-set) talking as soon as possible, so that (*likay*) they could again be free (p-stem a-set) to operate this miraculous machine'.[83] Clauses that express the desired but unrealized result of an action are sometimes introduced by *fa* (although the "purpose" conjunctions are also used for this purpose, as in the above example). In contrast to its use as a coordinating conjunction, where it expresses a relationship of actual, realized consequence and normally requires the s-stem, *fa* as a subordinator is followed by the p-stem a-set, for example, 'But his preoccupation was how it was able to ring by itself, and whether Muslims would be able to change it *fa* it would say (p-stem a-set) "God is most great" instead of making that noise'.[84] Clauses of reason are introduced by *liʔanna*, a combination of *li* 'to, for' (see immediately above) and the nominal complementizer *ʔanna* 'that'. Like *ʔanna*, *liʔanna* is followed by SVCOMP word order and a "factual" verb form: s-stem or p-stem u-set.

Dialectal equivalents of these conjunctions vary, for example, *la, ta*, and *laḥatta* (Damascus), *lijil, ḥagg*, and *ḥatta* (Gulf). Perhaps the commonest pandialectal "purpose" conjunction is *ʿalaša:n/ʿaša:n*, which seems originally to have been Cairene but now has gained wide dialectal currency outside Egypt, especially in relaxed cross-dialectal speech contexts.[85] In Cairo the same conjunction is often also used, somewhat confusingly, in clauses of reason to mean "because." Elsewhere, reason clauses are most often introduced by a local reflex of *liʔanna*. As in MSA, purpose clauses, whose content is by definition nonrealized, require the "nonfactual" verb form (the prefixless p-stem), whereas reason clauses are followed by the s-stem or prefixed p-stem.

One important difference between the MSA use of purpose and reason clauses and the CLA use is the greater flexibility in their position in MSA. In CLA and early modern writing, it is rare to find such clauses preceding the main clause, whereas nowadays this is not uncommon, at least in fictional prose. Because the preposing of these particular types of adverbial clause is also unusual in the dialects, it may well be that its modern use reflects a European linguistic influence on educated writers. The following example from a recent novel, in which two conjoined purpose clauses, one of reason and one of purpose, are preposed, is typical: 'And because (*liʔanna*) she didn't want her sacrifice to be for nothing, and so that (*likay*) she could make it easy for him and obviate any difficulties for herself, she . . . '.[86] In expository and documentary written texts, lengthy preposed adverbial modifications of the main clause are also common but usually appear in the form of adverbial phrases, rather than clauses, probably because of the impersonal style that this type of exposition requires and that

the use of phrases, but not finite verbs, allows (see 8.3.3). Thus, for example, in a recent tract on economic and social development in Kuwait, one reads: 'In order (*li*) to grasp (verbal noun) what "human potential" is, it is necessary to analyze (verbal noun) . . . '.[87] In fact, the use of logical/temporal connectors consisting of adverbial phrases derived from clause conversion is a very noticeable feature of the style of this kind of writing: later on in the same text, the writer states that developing a society's human potential is important *bi ḥukmi kawniha: lmuḥarrika l?asa:si:y* . . . [88] 'by virtue of its being the basic engine of . . . ', rather than simply saying *li?annaha: lmuḥarriku l?asa:si:y* 'because it is the basic engine of . . . '.

7.2.2.3.3 Concessive/exceptive clauses

MSA concessive clauses are introduced by one of a number of prepositions followed by *?anna: ma'a ?anna, bi/'ala: rraġmi (min) ?anna, 'ala: ?anna, ?illa: ?anna, bayda ?anna,* and *ġayra ?anna,* all of which have the sense of 'although, except that, despite the fact that'. As in other types of clause introduced by *?anna,* SV-COMP order is required, and because concessive clauses necessarily refer to completed actions or existent states, they require s-stem or p-stem u-set (i.e., indicative) verbs. A functional distinction in written MSA can be drawn between clauses introduced by *ma'a ?anna* and *bi rraġmi min ?anna* versus those introduced by the others. Clauses introduced by the first two conjunctions occur before or after (but most often before) their main clause, whose content they modify rather specifically. In the case of *ma'a ?anna* 'although', this close linkage is reflected by a resumptive *fa* at the beginning of the main clause, for example 'Although (*ma'a ?anna*) Thuwaiba only nursed him a few days, (*fa*) he always retained great affection for her'.[89] In the case of *bi rraġmi min ?anna,* 'despite the fact', the main clause may also be introduced by *fa* or by the seemingly tautologous *?illa: ?anna* 'nonetheless', for example, 'Despite the fact that (*bi rraġmi ?anna*) these developing societies enjoy the health advantages that scientific progress offers to all, resulting in a fall in the death rate, (*?illa: ?anna*) there has been no concomitant development in patterns of thought and custom'.[90] A common exceptive locution in MSA involves the phrase *?illa: ?iða:* 'only when, unless', as in, for example, 'The trade embargo will not be lifted unless (*?illa: ?iða:*) the military offensive ceases'. The other concessive conjunctions, especially in media texts and expository academic writing, also occur in sentence- or even paragraph-initial position, but the difference is that they introduce a proposition that qualifies or gainsays a *preceding* proposition (or a whole argument) contained in a unit of text that is usually larger than a single clause and often as large as a whole paragraph. Thus, unlike clauses introduced by *ma'a ?anna* and *bi rraġmi min ?anna,*

they float completely detached from the main clause they modify and at first sight seem like incomplete parts of sentences. For example, having enumerated at length the reasons for the indispensability of applying the principles of economic analyis to the study of housing provision, a writer begins his next paragraph: 'Except that (*ġayra ?anna*) the efforts made by the economists only became obvious in the context of the postponing, at the beginning of the 1970s, of the very issues that this new science addressed'.[91] As can be seen from this example, the functional similarity of Arabic "subordinating" conjunctions to English ones has limits. Rather than subordinators, they function like English coordinating conjunctions or adverbs and, in the illustrative example *ġayra ?anna,* would be more acceptably rendered in English by 'Yet, . . . ' or 'However, . . . '.

The dialects generally employ reflexes of *ma'a ?anna* and/or *bi/'ala: rraġmi (min) ?anna* to introduce concessive/exceptive clauses. For example:

*raḥ aku:n ihni:k **ma' ?annu** wa?ti ḍayyi?* [Damascene][92]
'I'll be there, **although** my time is limited.'

*baḥibb issiga:yir **raġm inni** hiyya ?aġla min išši:ša* [Cairene][93]
'I like cigarettes **although** they are more expensive than the water pipe.'

Other functionals that introduce concessive clauses in MSA are either not used at all or are used differently in the clause syntax of the plain vernaculars. *?illa* 'except' and *ġe:r* 'other than' (MSA *ġayra*) are widely used and do occur as conjunctions but usually in combinations such as *?illa ?iza* (most dialects) 'except if', and *ġe:r lamma* (Cairene) 'unless', so that they are really part of the dialectal conditional system (see 7.2.2.4) below). However, educated speakers when speaking formally do use slightly dialectalized forms of MSA concessives, for example, 'The fact is that the housing situation in Damascus up until two years ago was normal, and rents were in line with other prices, except that (*?illa ?anno*) in the last two years it appears that inflation has spread to housing . . . '.[94] Less educated speakers, or the same educated speaker in more relaxed circumstances, would be likely to use plain dialectal *la:kin* 'but' rather than *?illa ?anno* in a sentence such as that exemplified.

At this point, it is perhaps worth making a few remarks on the differences between the dialects, MSA, and CLA concerning the order of dependent and main clauses and the more general question of "message organization." We have noted that in spoken dialectal Arabic, clauses of time occur much more freely before the main clause than those that express some types of logical relationship—purpose, reason, and concession—that tend to follow. In the latter cases,

speakers usually arrange their messages syntactically in a way that mirrors the real-world order of the events these messages describe or that presents what they see as the essential point of their message first—that is, by an initial main clause that describes a scene, recounts an event, or communicates a proposition, which is then modified in some way by the following clause. However, the preposing of clauses expressing logical relationships has become a salient feature of written MSA that contrasts with both CLA and contemporary speech; so has the preference for clause subordination, passivization, and nominalization over clause coordination and the use of active finite verbs (see chapter 8). It seems in fact that the syntactic preferences for mapping meanings onto sentences in modern writing are different from those that underlay older writing and that still underlie speech. In part, the calqueing of European syntactic models may lie behind the syntactically convoluted style of much MSA expository writing on abstract subjects. But equally, a phenomenon such as dependent clause preposing (as well as long sentences and multiple embedding) can be seen as one outcome of the particular type of syntactic planning and principles of message presentation that are possible precisely because of the non–real-time conditions in which a text is written (revised and re-revised) and read (and reread and come back to). In particular, clause preposing in written Arabic provides a means of highlighting visually aspects of a message that may be expressed in spoken Arabic by linguistic means (stress and intonation, among others) that are not available to the writer. Written MSA and the spoken dialects are not, in other words, simply alternative codes for "saying the same thing" appropriately in different social contexts: each is associated with distinctively different ways of organizing messages because of the physical differences in the typical contexts in which the message sender sends and the message receiver receives. Although we can have no access to how Arabs in the premodern era spoke, the (to modern sensibilities) loose-knit, "orate" organization of CLA prose writing, even where the subject matter was abstract and the tone expository, illustrated by the passage from the fourteenth-century polymath Ibn Khaldūn quoted in 7.2.1, suggests that the gap between Arabic writing and speech may have been much narrower in the premodern period than it is now.

7.2.2.4 Conditional clauses

Conditional sentences consist of two structurally independent clauses that contain propositions, the validity of one of which is dependent on the validity of the other. Only if the proposition in the conditional clause is realized can the proposition in the answering clause be deemed valid. This relationship is entailed whether the conditional sentence describes events in a possible future

world, for example, 'If he agrees, he will/is going to regret it'; in a less probable one, for example, 'If he agreed/were to agree, he would regret it'; or in an impossible one, when a no longer fulfillable condition is imagined to have been realized, for example, 'If he had agreed, he would have regretted it'. Sometimes, as in sentences such as 'If he agreed, he (may/will have) regretted it', the temporal/modal value of the verb in the conditional clause can be correctly construed only in the light of the value of the verb in the answering clause: here, 'if he agreed' refers to a past world about which the speaker is uncertain, rather than to some future but unlikely world. These types of dependency relations between verbs in English main and conditional clauses find an echo in Arabic, although, as we shall see, the choice of conditional particle plays a role in Arabic that is largely absent in English.

The salient features of conditional sentences in MSA are the sequence of verb forms used and the particles used to introduce the conditional clause (the protasis) and, in some types of sentence, the answering clause (the apodosis). As in the English conditionals above, there is a typical sequence of verb forms in MSA conditional sentences that sets them off from other types of complex sentence. But, unlike English, the type of condition—real, possible, or unreal—is signaled chiefly by the particle used to introduce the conditional clause, rather than verb form sequence per se. And again unlike English, the meaning of some complete conditional sentences (i.e., the type of "world" they refer to) is often ambiguous out of context.

The protasis, which may occur before or after the apodosis, is introduced by one of several particles equivalent to English 'if': ʔin, ʔiða:, law. In all cases, a verb-initial clause normally follows. If the proposition contained in the clause would not normally be expressed by a verb, then the verb ka:na is used: ʔin ka:na zayd huna: 'If Zayd is here'. ʔin presents the proposition that follows it as something that one cannot be sure will happen but that is in the realm of the possible. ʔiða:, however, usually implies that the proposition in its clause will definitely occur—it is only the timing that is uncertain. The classic illustration of the difference between the two in CLA is ʔin ma:ta zayd versus ʔiða: ma:ta zayd. The first clause means 'If Zayd dies/should die' (e.g., of his present illness), and the second 'When(ever) Zayd dies' (i.e., like all men, he eventually will). In MSA, as we noted in 7.2.2.3.1 above, ʔiða: has to some extent encroached on the territory of CLA ʔin, whereas in spoken Arabic its various local dialectal reflexes are now easily the most common particle used in "open" conditional clauses. law is used to introduce conditions that are viewed as improbable or impossible of fulfillment: law ma:ta zayd can mean 'If Zayd died/were to die' or 'If Zayd had died'—only context determines which.

In all of the example clauses above, the form of the verb is the same—an s-stem—regardless of whether Zayd's death is viewed as a more likely or less likely, or is implicitly denied ('if Zayd had died'). The s-stem here has the function of simply projecting Zayd's death, forward or backward in time, as a fact from which consequences have flowed/might flow/or would have flowed. It has no preterital meaning—its most common tense coloring outside conditional clauses[95]—except where the context so determines. Thus different shades of probability of a conditional clause being fulfilled are signaled in written Arabic by the choice of particle and not, as in English, by the form of the verb.

Nonetheless, in CLA, there are preferences (one cannot really say rules) for verb sequences in conditional sentences. Space does not permit us to deal exhaustively with them here; but if, for example, as is most common in "open" affirmative conditions ('If you go, I'll go too'), an s-stem is used in the protasis after *ʔin*, an s-stem is also usually found in the apodosis. Similarly, a p-stem base set (jussive) form can be used in the protasis after *ʔin* but, again, normally requires the same form in the apodosis. It is (according to CLA grammar books) also possible, but in practice unusual, for s-stem and p-stem base set verbs to combine in conditional sentences of this type and to do so in either order. In negative conditionals, *ʔin* and *law* negate the protasis verbs with *lam* + p-stem base set; following a negative in the protasis introduced by *ʔin*, a negative in the apodosis can be *laː*, *lan*, or *lam* but must be *maː* + s-stem after one introduced by *law*. Where formal parallelism is not maintained after *ʔin* or *ʔiðaː*, as in the negative combinations described in the previous sentence, or when the apodosis is an imperative ('If you see her, tell her . . . '), *fa* is inserted before it. If the temporal coloring of the protasis verb needs to be made unambiguous because context by itself does not make it so, the s-stem auxiliary verb *kaːna* is used, for example, *ʔiðaː maːta zayd* 'If Zayd dies/should die' but *ʔiðaː kaːna zayd maːta . . .* 'If Zayd died' (= 'If it is the case that Zayd died'). The same applies to hypothetical conditions: *law maːta zayd* means 'If Zayd were to die/had died', but if it is desired to indicate unambiguously the second, then *law kaːna zayd maːta* is possible.

If this partial sketch of the CLA rules seems complicated—and to almost every "rule" it seems possible to find exceptions—it is probably because the description of CLA we have was an attempt by early medieval Arab grammarians at rationalizing a mass of sometimes conflicting raw language data (chiefly from the Koran, Hadith, and the ancient poetry) that did not originate in one area of Arabia but was culled from a variety of ancient dialects and spanned several centuries. If one is to judge by present-day grammar books, the "rules" for the

sequence of verbs in CLA conditional sentences are still supposed to apply to MSA. But in nonliterary genres (and in the writing of some modern novelists too) in which models of "correct" usage, inevitably based on literary or religious CLA, are not at the forefront of the writer's concerns, the reality is different. In order to illustrate this, we will present some examples of one modern writer's use of conditional sentences in a single text: ʿAbdallah al-Nafīsī's short polemic (seventy pages) on the political, social, and economic policies of the states that form the Gulf Cooperation Council, published in 1982.

What strikes one most forcefully is the functional particularization of both particle use and verb sequence and the fact that so many of the theoretical combinatorial possibilities in CLA grammar books are not used. *ʔin* and *ʔiðaː* are clearly distinguished from each other, but the distinction is connected to text function in a rhetorical sense, rather than expressive of different condition types, as it is in CLA. *ʔin* is found exclusively in parenthetic conditional clauses or "anticonditional"[96] clauses introduced by *wa ʔin* 'even if' (i.e., the main proposition will still be valid whether or not the condition is fulfilled.) These clauses do not form part of the thread of the main argument but are rather evaluative comments on it, or on the language used to describe it. In such clauses al-Nafīsī invariably uses *ʔin* followed by the s-stem or, in the negative, *lam* + p-stem base set, whatever its time reference. For example:

> ' . . . it is justified in creating a similar alliance—even if (*wa ʔin*) it differs (s-stem) in name—with an Arab state, which is Libya' (p. 17)

> 'this economic factor, even if (*wa ʔin*) it has (s-stem) a great importance, . . . ' (p. 59)

> 'because it was a good initiative—even if (*wa ʔin*) it was (s-stem) off-the-cuff and unplanned' (p. 30)

> ' . . . their "policy"—if (*ʔin*) the expression is permissible (s-stem) - . . . ' (p. 60)

> 'the moral obligation—if (*ʔin*) we don't call it (*lam* + p-stem base set) the geopolitical one— . . . ' (p. 19)

The same pattern of verb use was found in formulaic types of particleless conditional phrase such as *ʔaradna: ʔaw lam nurid* 'whether we like it or not' and *šaːʔat ʔam ʔabat* 'willy-nilly'. For example:

> 'Nations of the GCC—whether they like it (s-stem) or not (s-stem)—are affected by this war . . . ' (p. 19)

In the rest (the bulk) of the conditional sentences—that is, the nonformulaic ones that form part of the main argument—it is striking that *ʔin* did not occur at all: *ʔiða:* is used in all examples of "open" conditions. *law* is used for contrafactives. Furthermore, a function-based subdivision can be made within the "open" group between "organizing" conditionals and "textual" conditionals. First, the "organizing" conditionals: here, where the writer intervenes to help the reader follow his argument (in rather the same way as some of the parenthetical asides exemplified above), a CLA-like structural parallelism of s-stems is consistently used. This is the only type of conditional sentence in the whole text where it is, although paradoxically such sentences are "conditional" only in the formal, surface syntactic sense; the conditional frame is really a handy stylistic variant of 'We will (now) do X in order to show (you, the reader) Y.' For example:

'If (*ʔiða:*) we move (s-stem) to Saudi Arabia, we observe (s-stem) that . . . ' (p. 47)

'If (*ʔiða:*) we exclude (s-stem) children, old people, and the disabled, the rest of the population represent (s-stem) the work force.' (p. 59)

Second, "textual" conditionals—the bulk of the data: here, the conditional sentences form part of the substance of the text, a description of the GCC's current political and economic problems. The pattern of verb sequences is, again, quite distinctive. The data fall into two distinct categories, and six example sentences, three from each category, have been selected as illustrations:

Type A: protasis = s-stem
(1) 'If (*ʔiða:*) there exists (s-stem) a will in the GCC states to treat the security problem seriously . . . defining the concept of political security becomes (p-stem u-set) necessary.' (p. 27)
(2) 'If (*ʔiða:*) the political leaders in the GCC continue (s-stem) with this approach, *fa* there is no doubt that they will remain (*sawfa* + p-stem u-set) in a vicious circle . . . .' (p. 61)
(3) 'The report, if (*ʔiða:*) it proves (s-stem) correct, throws (p-stem u-set) more light on western allegations . . . .' (p. 11)

Type B: protasis = *ka:na* + p-stem u-set
(4) 'If (*ʔiða:*) the neutral policy which the GCC states are pursuing vis à vis the combatants seems (*ka:na* + p-stem u-set) justified, *fa* this neutrality does not mean . . . ' (p-stem u-set) (p. 18)

(5) 'If (*ʔiða:*) the GCC states do not appear (*ka:na* + p-stem u-set) unified in their oil policy . . . (*fa*) how is it possible . . . .' (p-stem u-set) (p. 18)
(6) 'If (*ʔiða:*) the brotherly Arabs in the peninsula and the Gulf consider (*ka:na* + p-stem u-set) Yemen a non-oil state . . . it is/would be better (s-stem) for them to call the GCC the GOSCC (Gulf Oil States Cooperation Council).' (p. 15)

Without exception in "textual" conditionals, the verb after *ʔiða:* is either an unadorned s-stem or *ka:na* (itself an s-stem) + p-stem but is followed by a variety of forms in the apodosis. The difference in the protasis verb marks another functional differentiation. Where an s-stem is used (Examples (1)–(3)), the condition is a real one in the sense that the writer leaves open the possibility of the condition not being fulfilled: there may or may not be "a serious will" (1); the leaders may or may not "continue with their current approach" (2); and the report may or may not "prove to be correct" (3). But in the second set of examples ((4)–(6)), where *ka:na* + p-stem is used, the protasis does not state a real condition at all. As is made clear elsewhere in the preceding text, the writer is sure that the GCC states' policy *is* "justified" (4); that the GCC states *are not* "unified" (5); and that they *do* "consider Yemen a non-oil state" (6). So *ʔiða:* + *ka:na* + p-stem, far from betokening a condition, in fact means "given that it is the case that . . . ." Furthermore, it is noteworthy that the verb forms in the apodosis, even in properly conditional sentences such as 1–3, do not show the structural parallelism typical of CLA and still common in literary MSA: they simply reflect the time/factuality of the proposition in the same way as they would in an unconditioned main clause. This is not to say that structural asymmetry of the kind in sentences (1)–(3) was totally absent from CLA conditional sentences—far from it—but it was not routine in the way it has become in much MSA writing. One reason for this, I would suggest, is that the range of semantic relationships expressed in the form of a "conditional" sentence has become wider: in many cases a conditional relationship is not entailed between the constituent clauses at all. The end result is a looser, more flexible clause structure: verb forms in the protasis retain the original CLA "conditional" feature (the s-stem) but not only have the functions of *ʔin* and *ʔiða:* changed significantly, the form of the verb in the apodosis corresponds more closely to the aspectual values it has outside conditional clauses. One contributory cause of this has probably been the influence of the dialects, the topic to which we now turn.

Some important aspects of the structure of conditional clauses in the dialectal Arabic of ordinary speech bear a close similarity to the variety of nonliterary MSA just described. We may cite the following points:

(a) Of the three conditional particles, *ʔin* has everywhere receded to the pe-

riphery of the dialectal system, if not completely disappeared from it, being commonly heard in relaxed speech only in fixed expressions such as *ʔinša:llah, ʔin ra:d allah* 'if God wills', *ʔin allah sahhal* 'if God permits', and in proverbs, in which cases it invariably precedes an s-stem verb. Compare this with the fact that in al-Nafīsī's text (unlike in CLA and more literary MSA) *ʔin* also occurred only in a narrow range of syntactic frames and ready-made phrases. As in non-literary MSA, *ʔiða:* (or its local equivalents *ida, iza,* and *ila*) has become the usual means of introducing all "open" conditions, although in some dialects (Cairo and the Gulf) reflexes of *law* are also common in this type of conditional clause. Where the condition is contrafactive ('If you had . . . , you would have . . . '), *law* or a reflex of it is almost invariably used.

(b) The sequence of verb forms also shows a number of parallels with non-literary MSA. Although there is some interdialectal variation, the basic system in all types of "open" conditional sentence is s-stem in the protasis, p-stem (with aspectual prefixes in dialects that have them) in the apodosis (compare MSA examples (1)–(3) above, which have this structure). The modal/temporal value that is carried by the protasis s-stem verb is determined by the temporal/aspectual coloring of the apodosis verb and any adverbial complements it, or the apodosis, may have. As in MSA, there is still some possibility of ambiguity out of context, as the translations of the examples below make clear. In Damascene, for example, the protasis *ʔiza sa:far* can be equivalent to English 'if he leaves', 'if he left', or 'if he were to leave' ((1)–(3) below). It can also mean 'if he had left' ((3)–(4)), although here the contrafactive nature of the proposition may be emphasized by the insertion of *ka:n* in the apodosis and/or the use of *law* instead of *iza* and the use of *ka:n* + s-stem in the apodosis. For example:[97]

(1) *ʔiza sa:far hallaʔ byu:ṣal ilyo:m*
(a) 'If he leaves now he'll arrive today' (implied: I don't know if he will leave)
   or, depending on context:
(b) 'If he left now, he would arrive today' (implied: I doubt that he will leave)

(2) *ʔiza sa:far mba:rḥa byu:ṣal ilyo:m*
   'If he left yesterday he'll arrive today' (implied: I don't know if he did leave)

(3) *ʔiza sa:far mba:rḥa biku:n wiṣil ilyo:m*
(a) 'If he'd left yesterday, he would have arrived today' (implied: he didn't leave, but imagine that he did).
   or, depending on context:

(b) 'If he left yesterday, he'll have arrived today' (implied: same as in (2), to which this sentence is an alternative)

(4) *law (ka:n) sa:far mba:rḥa ka:n wiṣil ilyo:m*
   'If he had left yesterday he would have arrived today' (implied: same as in the (a) translation of (3), to which this is an alternative)

(c) There are cases where the protasis does not really contain a condition at all, in the sense that the validity of the apodosis does not depend on its realization. In such clauses, *iza* is followed by a p-stem verb, by a participle, or by no verb at all, and the auxiliary s-stem *ka:n* may be optionally inserted before any of these elements. This combination of forms has an exact parallel in MSA, where it realizes the same kind of nonconditional meaning. As in MSA, the insertion of *ka:n* does not change the temporal (present) reference of the p-stem but rather makes the "conditional" clause virtually an assertion of fact. Compare the dialectal examples (5)–(7) below[98] with the MSA examples ((4)–(6)) above containing the parallel structure *ʔiða:* + *ka:na* + p-stem:

(5) *ʔiza (ka:nat) ittaqri:re:n byitna:qaḍu, la tṣaddiʔ la: ha:d ula: ha:d*
   'If the reports conflict (*ka:n* + *b* + p-stem), (well then) don't believe (imperative) either one' (implied: 'given that the reports *do* conflict . . . ')

(6) *ʔiza (kunna) naḥnu ḥari:ṣi:n ʕa lluġa l'arabi:ya . . . fa yajib attarki:z . . .* 'If we are really eager (*ka:na* + a.part) to have (Classical) Arabic . . . (well then) we must (p-stem) concentrate . . . ' (implied: 'given that we are eager . . . ')

(7) *iza (ka:n) ʕale:k ši la titʔaxxar minša:ni*
   'If there's something you have (*ka:n* + prep. phrase) to do, (well then) don't delay (imperative) on my account' (implied: 'given that there is something you have to do . . . ')

Dialectal and nonliterary MSA conditional sentences thus appear in many respects to be two sides of one coin: even though on the surface the morphological detail is different, at a deeper level there is a similar relationship between corresponding forms and textual functions. It would appear that the more formulaic types of conditional remain a "special case," in terms of particle and verb form, in both nonliterary MSA and the dialects (i.e., in both they are like the CLA system) and that, although other types of condition are basically also expressed in the same way, this is probably as a result of long-term dialectal influence on MSA, rather than the other way round.

Notes

1. The usefulness of these commonly used Latinate terms is doubtful, because the syntactic uses of the Arabic cases to which they are applied correspond only very roughly to those that carry the same label in Latin or other European case languages such as German.

2. El-Yasin 1985, 112.

3. Agius 1991, 42–43 reports three studies of word order in verbal sentences in Arabic based on samples of modern writing, which show that VSCOMP accounts for about 70 percent of sentences containing a verb.

4. Khan, describing what was shown by corpus analysis to be typical of CLA prose texts, comments (1988, 30–31): "One consideration which has a bearing on this choice [sc. between SV and VS] is text genre: SV independent main clauses are attested far more frequently in expository or descriptive texts than in narratives. The operative factor behind this phenomenon is the aspect of the verb. SV is more common in independent main clauses with imperfect verbs [sc. p-stem], which are characteristic of expository and descriptive texts, than those with perfect verbs [sc. s-stem], which express events in a narrative." This observation seems to be equally true of MSA prose.

5. Cantarino 1974–75, 2:517.

6. Diyab 1991, 10.

7. Mahfouz, 1966, 55.

8. Husayn 1966, 3.

9. See Khan 1988, 29–62 for a detailed exposé.

10. The origin and function of *ʔinna*, which in CLA was usually an emphatic particle used to introduce a proposition (= 'Verily . . . ') has been the subject of much speculation (see, e.g., Bloch 1986, 102–36). In some dialects, there are sentence-initial particles that function somewhat like *ʔinna* does in CLA. These particles seem to be decayed forms of dialect (and ultimately CLA) verbs that mean 'to see' and are analogous to the French presentative 'voici!', for example, Moroccan/Algerian *ra-* (? < *raʔa:* 'to see'), south Yemeni *ša'-* (? <*ša'a:* 'to look') (Fischer 1959, 186ff, 199–200), Omani *ṣa:-* (Reinhardt 1894, 124).

11. Al-Nafīsī 1978, 17.

12. El-Yasin 1985, 121 makes this claim for northern Jordanian "rural dialects," Britti 1980, 126 for Levantine dialects in general.

13. See Cowell 1964, 407–10.

14. The apparent exception to this is where, in MSA, as in the example given earlier, *ʔinna* is preposed to an extraposed indefinite nominal for reasons of emphasis. But in such cases, the indefinite noun is actually part of an already defined entity, for example, 'a part of them' (where 'them' = 'the components') in the example given.

15. Syrian, Lebanese, northern Galilean, and Iraqi dialects mark direct objects with a semantically empty particle *la*, which, given its geographical distribution, is likely to be an Aramaic substrate influence. See Levin 1987 for Galilean data and references.

16. Cowell 1964, 410.

17. Ibid., 409.

18. Oestrup 1897, 42–49

19. Holes 1984, 35–39.

20. This simplification seems to have been a relatively early consequence of the learning of Arabic by non-Arabs after the Islamic conquests (see Blau 1966–67, 275 ff.).

21. See Belnap 1993, 102–4.

22. Rosenhouse 1978; Woidich 1991.

23. See Beeston 1973. In modern Arabic literature, one encounters descriptive passages that use a combination of punctuation and asyndetic linkage reminiscent of English, for example, Ibrahim 1992, 12 (literal translation, preserving original punctuation and lack of conjunctions): " . . . he was good-looking, smart, armed with the golden accessories: cigarette case, lighter (Ronson), ring, Old Spice perfume, narrow pointed shoes, a knowledge of different types of food and how they should be served, a constant complaint about state policies which favoured the public sector and manufacturing . . . ." The passage continues for several more lines in the same vein, piling up attributes asyndetically in a way that would seem excessive even in a language such as English that favors this type of linkage. In some dialects, asyndetic repetition of the same verb occurs but has aspectual force, for example, (Bahrain) *fakkar, fakkar, fakkar, fakkar, fakkar . . . ga:l ta'a:l ya:ka:tib, ta'a:l, ʔiktib!* 'He thought, he thought, he thought, he thought, he thought . . . he said "Come here, writer come here and write!"' Here the five-times repeated *fakkar*, without intervening conjunctions, indicates 'he thought long and hard', that is, it has intensive force. Henkin (1996, 182) gives some similar dialectal examples from Palestinian Bedouin narrative.

24. Ibrahim 1988, 8.

25. Holes 1984, 56–57.

26. Example from the author's unpublished Bahraini data.

27. Example from the author's unpublished Bahraini data.

28. cf. Brockelmann 1908–13, 2:501: "Es sind asyndetisch oder syndetisch angereihte Sätze, die nicht einen Fortschritt im Bericht bringen, sondern Nebenumstände der Handlung darstellen." ["There are asyndetically or syndetically arranged sentences that do not advance the narrative, but present the circumstances surrounding the action."]

29. Wright 1896–98, 2:330.

30. Rosenhouse 1978, 229 for the Damascene example, Woidich 1991, 70 for the Cairene.

31. First (journalistic) example from *al-ʿArabī* magazine (pub. Kuwait), May 1993, 12; second (literary) example Munif 1988, 41.

32. Cantarino 1974–75, 3:19.

33. Ibrahim 1988, 8.

34. Nuṣṣ 1930, 2.

35. Cantarino 1974–75, 3:23.

36. Cowell 1964, 397.

37. Holes 1990, 63.

38. Ziadeh 1964, 34.

39. Lyons 1962, 59.

40. Ziadeh 1964, 1.

41. Ibn Khaldūn (n.d.), 270.

42. Ibn Khaldūn (tr. Rosenthal) 1986, 2:271.

43. Al-Shaykh Durra 1988, 65.

44. Stetkevych 1970, 93–94.
45. Cantarino 1974–75, 3:50.
46. Ibid., 49.
47. Ibid., 50.
48. Ibid., 1:115.
49. Cowell 1964, 394.
50. Ibid., 394.
51. Ibid., 396.
52. Holes 1990, 64.
53. Wise 1975, 154.
54. Ibid., 154.
55. Holes 1990, 64.
56. Cantarino 1974–75, 2:501.
57. Cowell 1964, 396.
58. Olmsted Gary and Gamal-Eldin 1982, 34.
59. Holes 2001, 18 sub '-M-M-A.
60. Cantarino 1974–75, 3:46.
61. Ibid., 47.
62. Hinds and Badawi 1986, 95.
63. Example from Leeds University Corpus of Educated Spoken Arabic, tape 11, side 2, lines 12–13 of the transcription.
64. Some exceptions are the dialect of the Jews in Tunisia and Libyan and Moroccan dialects generally, in which *fa* does not occur; and in the Arabic dialects in which it does occur, *fa* may be a recent import from MSA.
65. See Persson 2002 for a data-based study of sentential complements in object position in MSA.
66. Ziadeh 1964, 115.
67. Various extralinguistic factors may cause this formalization—topic, interlocutor relationship, context, etc. A good example was provided by a semiliterate Bahraini farmer whom I asked in the course of what up to that point had been a relaxed conversation conducted in the local dialect, to explain to me the object of making the pilgrimage to Mecca. His reply was *ilqaṣd min iddaḥa:b ʔila lḥajj kaffa:ra ʔila ddunu:b* 'The point of going on the pilgrimage is to do penance for (one's) sins'. *daḥa:b* 'going' and *kaffa:ra* 'penance' are both verbal nouns that have the effect of depersonalizing and generalizing the proposition, and both, lexically speaking, are purely CLA/MSA items. Later on he explained what he meant using the equivalent dialectal verbs *ra:ḥ* 'to go' and *ta:b* 'to repent', and in clause format 'If you go to Mecca . . . '.
68. The examples are taken from the Leeds University Corpus of Educated Spoken Arabic, and recordings I made in Bahrain in 1977–78 and Jordan in 1989. As well as a difference in the means of verb complementation, there is a consistent difference in the mood marking of the verb in the dependent nominal clause. In the urban dialects of Cairo and the Levant, which have a fully developed system of aspectual/modal verb prefixes, cognitive verbs in the main clause govern p-stem verbs in the nominal clause that carry a *b*-prefix (i.e., the 'thought', 'belief', 'observation' is presented as a fact), whereas desider-

ative main verbs do not (i.e., they present their objects as nonfactual, notional states of affairs).
69. After the "generic" definite article, when the clause is nonrestrictive, the relative adjective is often omitted. For example, in a passage of Ahmad Amin (quoted by Ziadeh 1964, 196), the Cairene lower classes are described as 'living most of their days on falafil, beans, and fish bought ready-fried from a shop'. Here, the nouns 'falafil', 'beans', and 'fish' are all of generic reference, which in Arabic requires the definite article, but the relative clause that is attached to the last of these grammatically definite nouns does not have a relative adjective:

as-samak	*yustara:*	*maqliyan*	*mina*	*ddukka:n*
the-fish	3msg-be bought	fried	from	the-shop

' . . . fish bought (already) fried from a shop'.

70. Damascene examples from Cowell 1964, 494–95; Baghdadi examples from Erwin 1963, 380, 386; Bahraini examples from the author's unpublished data.
71. Examples do occur in MSA of the masculine form *ʔallaðiː* being used where the reference is indefinite but generic, perhaps as a result of dialect influence, for example, 'Anyone who (*ʔallaðiː*) reads the constitutive charter of the Council will find that . . . ' (Al-Nafīsī 1982, 16).
72. See Haddad and Kenstowicz 1980. For the examples quoted see, for Lebanon, Feghali 1928, 311ff.; for Syria, Cowell 1964, 499; for Palestine, Bauer 1926, 76 and Blau 1960, 256.
73. Erwin 1963, 381.
74. Ibid., 380.
75. This convergence does not seem to have occurred in Cairene or the Maghrebi dialects.
76. Reckendorf 1895, 658.
77. Ibid., 653.
78. Ibid., 642.
79. Cantarino 1974–75, 3:288.
80. Ibid., 303.
81. Ibid., 300.
82. Sometimes the prefixless p-stem is used in this kind of sentence, because the *b(i)*-marking of the p-stem verb in the main clause is sufficient to indicate that habitual action is meant.
83. Munif 1988, 408.
84. Ibid., 556.
85. This use of *ʿaša:n* by speakers for whom it is not a native dialectal feature is a good illustration of the "leveling" phenomenon described by Blanc (1960, 82). Rather than use a conjunction that is specific to their own dialect or use an MSA one such as *likay* that is not part of any dialect and sounds too "high-flown," speakers opt for a "dialectal standard" that is well known everywhere.
86. Ibrahim 1992, 92.
87. Al-Nafīsī 1978, 17.

88. Ibid., 26.
89. Cantarino 1974–75, 3:140.
90. *Hunā Landan* (*London Calling*) (BBC Arabic Service magazine), July 1987, p. 14.
91. Al-Shaykh Durra 1988, 12.
92. Cowell 1964, 530.
93. Hinds and Badawi 1986, 342.
94. From unpublished material from the Leeds University Corpus of ESA, tape 11, side 2, lines 4–5 of the transcription.
95. Bravmann 1977, 563–71 argues that s-stem use in conditional sentences arose via a semantic shift in the meaning of sentences in which this s-stem originally had an incident-specific preterital meaning. It is well known that the origin of the conditional particle *ʔiða:* is the now obsolete noun *ʔið* 'time' (Wright 1896–98, 1:292). According to Bravmann, *ʔin* has a similar provenance, from Proto-Semitic/Old Arabic *yawma* 'day' (cf. Hebrew *ʔim* from the same source). Both were originally, as Gulf dialectal *yo:m < yawm* still is today, conjunctions of time. Gradually, a "general" meaning replaced the original "concrete" meaning of these sentences, without a change occurring in the form of the verb. Thus 'When you did X, you did Y' (specific reference to one actual event) came to mean 'Whenever (on every occasion when) you did X, you did Y' (less specific) and then (most generally) 'Whenever (= if) you do X, you do Y'.
96. This term is borrowed from Beeston 1968, 86.
97. Cowell 1964, 333.
98. Examples (5) and (7) taken from Cowell 1964, 332; example (6) from the Leeds University Corpus of ESA, tape 23, side 1, lines 83–84 of the transcription.

8

Lexical and Stylistic Developments
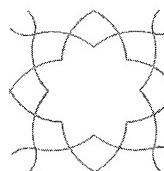
8.1 Foreign borrowing into Arabic

There are plentiful examples of the borrowing of foreign lexical elements into Arabic, even in the earliest Classical texts. The word *ṣira:ṭ* 'way, path' from Latin (via Greek and Aramaic) *strata* appears in the *fa:tiḥa*, the opening *su:ra* of the Koran, and in other *su:ras* we find Greek borrowings, such as *qisṭa:s* 'balance, instrument for weighing', and a large number of Persian ones such as *zanjabi:l* 'ginger', *ʔistabraq* 'brocade', *firdaws* 'paradise', and *tannu:r* 'furnace, hell'. The extent of foreign borrowing into Arabic, which has occurred at all stages of the language's development,[1] is obscured by the degree to which foreign elements have tended to be assimilated to the root-pattern system of the morphology and also by the organizing principles of Arabic dictionaries, whereby assimilated borrowings are listed under a theoretical "root." A typical early example—one of many hundreds that could be cited—is the medieval borrowing *di:ba:j* 'silk brocade' < Pers. *di:ba:*, which eventually gave rise to a Pattern II verb *dabbaja* 'to embellish', and subsequently to secondary derivations such as *mudabbaja:t* 'figures of speech'. All these words are arranged in modern lexica under a notional triliteral etymological root *DBJ*. Words borrowed from other Semitic languages, with which Arabic shares many roots and the same morphological principles, are less easily detectable. For example, the early Syriac borrowing *malaku:t* 'realm, kingdom' is ultimately derived from the common Se-

mitic *MLK* 'rule, own', a root that produced many regular Arabic formations such as *malik* 'king', *milk* 'property'. *malaku:t* is listed under *MLK* in Arabic dictionaries, although it is formed on a pattern (*CaCaCu:t*) that Arabic originally lacked.

In the centuries before the Islamic conquests, we know that there was trade between the Arabs living in the Red Sea ports and non-Arabs who traveled to them, and that the Arab tribes living on the northern fringes of Arabia had been in contact with speakers of other languages, Semitic and non-Semitic, for several centuries. It seems likely that parts of what is now Syria had been partially arabized for at least 200 years before the middle of the seventh century, when the conquests began.[2] But without doubt it was, initially, the processes of postconquest urbanization and long-term assimilation of the Arab invaders into indigenous non-Arab populations in the Levant, Egypt, and Mesopotamia that provided the conditions for the absorption into Arabic of many foreign terms, as well as the evolution of Arabic syntax. All the historical evidence suggests that, in the domain of ordinary speech, there must have been a long period of bilingualism over vast tracts of what is now monolingual Arabic-speaking territory, when various dialects of Aramaic, Syriac, Persian, Coptic, and Greek were being spoken alongside Arabic. The dialects of Arabic now spoken where these languages were formerly dominant still contain ancient lexical vestiges of them. In the Arabic of Mesopotamia, for example, many basic items of vocabulary such as *bazzu:n* 'cat' and *zabi:l* 'basket' are of Aramaic (and in the case of *zabi:l* ultimately Akkadian origin); *falla:š* 'to destroy' is a Syriac element, whereas *ča:ra* 'remedy', *čarix* 'wheel', *sangi:n* 'strong', and a host of current agricultural and food terms such as *šuwandar* 'beetroot' and *turši* 'pickles', all derived from Persian, are also of ancient origin.

Later on, with the flowering of Islamic civilization between the ninth and thirteenth centuries, lexical borrowing into literary Arabic reflected the changing nature of cultural contacts with the surrounding areas. At different periods over this long span of time, Greek philosophical and scientific works were translated into Arabic, and the literary and material culture of Persia also left its mark on the lexicon of medieval CLA. Among the many words borrowed directly into Arabic from these sources, one may cite Greek *faylasu:f* 'philosopher' < philosophos, *na:mu:s* 'law, principle' < nomos, *qa:nu:n* 'law' < kanon, *qa:mu:s* 'dictionary' < okeanos, *biṭa:qa* 'piece of paper, ticket' < pittakion, and *?iqli:m* 'region' < klima. Persian provided *sirwa:l* 'long trousers', *jawhar* 'jewel', *ward* 'rose', *?ibri:sam* 'silk', *?usta:ð* 'teacher', *sijill* 'register', *namu:ðaj* 'model, exemplar', *fihrist* 'index', and hundreds of others. Such borrowings were accepted into the lexicon of literary Arabic and fully assimilated morphologically, forming verbs (e.g., Pattern II *sajjala* 'to register' < *sijill*, Pattern V *tawarrada* 'to become rose colored' < *ward*) and broken plurals (e.g., *fala:sifa* 'philosophers', *jawa:hir* 'jewels', *wuru:d* 'roses').

The prolonged Ottoman Turkish political domination of the Middle East and North Africa, roughly from the mid-fifteenth to the late nineteenth century, had, comparatively speaking, little impact on the lexicon of literary Arabic. In fact, Arabic, as the language of religion and law, and Persian, as the language of a high courtly and literary culture, were regarded by the Ottomans as superior to their own and provided a reservoir of literary phraseology and legal and other technical terminology on which they drew freely. The main impact of Turkish on Arabic was on the spoken idiom of the cities and the material culture of everyday life. Few of these Ottoman-Turkish borrowings have found a permanent home in literary Arabic, but some are still current in the dialects, into which they have been fully assimilated morphophonologically, for example, Cairene: *duġri* 'straight ahead', *kubri* 'bridge', *oḍa* 'room', *šaku:š* 'hammer', *baški:r* 'towel'; Baghdadi: *belki* 'perhaps', *hič* 'nothing', *ham* 'also', *ča:ku:č* 'hammer', *parčam* 'rivet', *buġča* 'bundle', and *pa:lto* 'overcoat'.

As Turkish political influence over the Arabs waned, so that of other imperial powers, and with it that of their languages, grew. In the Gulf, effectively under the political and commercial control of British India from the late nineteenth century, and in North Africa, dominated and colonized by France from 1830, another wave of foreign vocabulary was absorbed into the local dialects. Again, as in the case of Turkish, words associated with the material culture and way of life of the imperial power poured in, in stark contrast to the imperviousness of literary Arabic to such a direct foreign assault. Typical of the influence of Anglo-Indian and English on the Gulf dialects during this preoil, preliterate period are borrowings such as *ga:ri* 'donkey-cart', *kra:ni* 'clerk', *kankari* 'gravel, concrete', *kirfa:ya* 'bedstead', *hafi:z* 'office', *le:t* 'light, street lamp', *sbe:tar* 'hospital', *dre:wil* 'driver', *bančar* 'puncture', and *še:wil* 'shovel'.[3]

Although, in the nineteenth century, literary Arabic did not suffer the same invasion by foreign terms of material culture as the dialects did—because literary Arabic did not serve as a medium for everyday communication—it was certainly not immune from outside pressure. The influence of European languages became increasingly strong but was more variously and indirectly mediated, chiefly through loan translations. Following the first modern contacts with Europeans after the Napoleonic invasion of Egypt in 1798, the development of the vocabulary of literary Arabic in the early nineteenth century was the result of piecemeal, individual effort by pioneering individuals such as Rifā'a Rāfi' al-Ṭahṭāwī (1801–1873), who began to promote cultural and technical transfer via translation from European languages (especially French). Al-Ṭahṭāwī it was who, in 1826, compiled the first list of translated Arabic technical terms and coined the Arabic word for electricity: *kahraba:?*. Later, as the economic importance of

modernizing the language came to be realized, the movement started by al-Ṭahṭāwī gained strength, and government schools of translation were set up. However, the influence of foreign languages on the vocabulary of Arabic through most of the nineteenth century continued to be essentially through the medium of translation projects carried out by individual Arabs translating into Arabic and was not part of any systematic policy.

In the later part of the nineteenth and the early twentieth century, as western political, economic, and scientific ideas proliferated and ramified through the Arab world, transliterated foreign words, especially in the sciences, and uncontrolled and sometimes inaccurate loan translations began to pour into written Arabic. Coincidentally, in the literary field, there was much experimentation going on with the western genres of the novel and drama (the latter in particular raising the thorny problem, when transferred to the Arabic context, of a highbrow literary genre in which a "speakable" as opposed to a "readable" idiom was required) and the associated apparatus of western literary-critical thought, for which Arabic lacked any distinct terminology.[4] If to all this is added a debate on language planning policy that had simmered on since the 1880s, fueled by the publicly expressed views of some colonial administrators in Egypt that the "dead" idiom and script of literary Arabic should be abandoned as a vehicle of modernization and replaced by a Romanized written form of the dialect, it becomes possible to understand the reaction of the literary–educational establishment. Horrified at the disfigurement of the literary language by intrusive foreign elements and by the potentially corrupting influence of the dialect, these self-appointed custodians of linguistic purity committed themselves to modernization in a more organized and consistent manner than had been attempted in the previous century, by setting up language academies that restricted themselves to coining words, wherever possible, by analogy with preexisting Classical patterns.

From the 1860s, the establishment of American missionary schools in Syria and Lebanon (the Protestant College, later to become the American University of Beirut, was founded in 1866) had led to a linguistic and literary revival of Arabic. Dictionaries and textbooks in Arabic began to appear. From around 1909, however, following the Young Turks' taking of power in Istanbul, the Ottoman government began an aggressive policy of Turkification in its non-Turkish-speaking provinces, with the aim of making power the preserve of a Turkish elite and countering embryonic secessionist tendencies among the subject peoples, particularly the Arabs.[5] This had the effect of transforming what had hitherto been a slow cultural revolution into a political one. The Arabic language became a symbol of Arab unity, political as well as cultural. The promotion of its use and development came to be seen as a cornerstone of the task of nation building, and there was never much doubt that the only kind of Arabic that fit-

ted the bill was that which was (theoretically at least) shared by all: Classical Arabic, albeit in a modernized and simplified form.

8.2 The language academies

We referred briefly to the activities of the language academies in chapter 1. The oldest of them, the Arab Academy of Damascus, was founded by Muḥammad Kurd ʿAlī (1876–1953) and held its first meeting on 3 July 1919. Thirteen years later, in 1932, the Royal Academy of Cairo (still in existence but now called the Academy of the Arabic Language) came into being. Each has published since its foundation the minutes of its meetings and a journal that, among many other things, records its decisions on the principles of word coinage and provides lists of approved neologisms. More recently, language academies have also been set up in Baghdad and Amman, and the Rabat-based Bureau of Arabization, affiliated to the Arab League, was set up in 1961 to fulfill a pan-Arab coordinating role. Although their policies have differed somewhat, the general aims of all these organizations can be summed up in the stated objectives of the first of them, the Arab Academy of Damascus, which was modeled on the Académie Française:

(a) The protection of the Arabic language against dialectal influence (including the proposal to Romanize, which was part of the agenda of some of those who wished to see one or other of the dialects adopted as national language(s));

(b) The maintenance of linguistic purity by ridding the language of intrusive foreign lexical elements brought in via the press and the radio, and some writers; and

(c) The adaptation of the language to modern needs, particularly in the area of science and technology.[6]

The language academies aimed to bring together not just panels of language experts, but academics and others working in technical fields in which foreign influence was great and in which the rapidity of the development of knowledge was likely to result either in the mass adoption by Arabic of foreign terms or in a chaotic, ad hoc coining of terms that might produce terminological inconsistencies from one part of the Arab world to another. In fact, as noted in chapter 1, the very proliferation of the language academies themselves, a reflection of the political fragmentation of the Arab world, has itself been a factor in ensuring that the terminological inconsistencies that the academies sought to avoid have come about and continue to plague communication in Arabic in the sciences.[7] But this has not been the only difficulty. The sheer volume of tech-

nical terminology for which Arabic equivalents needed to be coined, as well as the long lead time taken to coin new terms by committee and to disseminate them, has meant that "leading edge" science is conducted in the Arab world, as elsewhere, in English, with the Arabic neologisms finding their way only much later, if at all, into institutionally produced undergraduate textbooks.[8] The prima facie plausible objectives of the academies conceal a number of internal contradictions and historical inaccuracies, which we will now briefly discuss.

8.2.1 Protecting the "purity" of the language

The academicians' desire to protect Arabic from dialectal and foreign influence was grounded in the belief that the dialects were debased forms of a language that had been corrupted by contact with foreigners but that at some (undefined) point in the past had been "pure." But Arabic was never, in recorded history, a "pure" or even a completely homogeneous language in the normal senses of these words: as we have noted, the earliest texts, including the Koran, contain evidence of morphological and syntactic variability, probably geographical in origin (western versus eastern Arabia), and there are undeniable cases of foreign lexical borrowing in the earliest extant material. The drawing, in the early medieval period, of a line that distinguished "Classical" from "non-Classical" language was chiefly motivated by religiopolitical considerations and was bound to be, linguistically speaking, an arbitrary freezing of the language at a particular point in time. But although the medieval grammarians' detailed prescriptions, in the form of a list of grammatical do's and don't's, are regarded as a kind of linguistic "gold standard" that has been dinned into countless generations of learners up to the present day, it is obvious that the less easily prescribable, more amorphous area of phrase and sentence structure has remained open (although perhaps imperceptibly so over any observing individual's lifespan) to foreign influence.[9] Foreign vocabulary, too, either via direct borrowing or loan translation, continually found its way into Arabic, particularly at the zenith of Abbasid power (ninth/tenth centuries). And over the centuries, the gap between how literate people spoke and how they wrote narrowed as society became more complex and the demarcation between "speaking" and "writing" contexts became blurred, facilitating the reciprocal transfer of elements between written and spoken Arabic.[10] The academicians' concept of the "purity" of literary Arabic was, in a literal historical sense, a myth. In reality the language had always been evolving, but after the long stagnation of the Ottoman period the pace of change accelerated much more obviously, occasioning their reaction. It is a debatable point whether these defenders of linguistic purity would have taken the conservative view they did had the Arab world in the late nineteenth/early twentieth century not been in

a position of political subservience to the West and had the West not therefore been perceived as such a cultural threat. In earlier periods of history where contact with foreign cultures had resulted in linguistic borrowing, but during which the Arabs had not been politically dominated, there had been no such reaction.

8.2.2 Adapting Arabic to the needs of the modern world

If, in the view of the academicians, the flow of foreign words and expressions into Arabic had to be stopped to preserve its purity, an alternative native lexicon had to be created. More than this, an apparatus for word coinage had to be put in place that was productive enough to meet a continuous demand for new terminology and flexible enough to render into Arabic the delicate differences of the source language terms. This is a conundrum that the prescriptive policies of the academies have found it impossible to solve satisfactorily in at least one major area—science and technology.

From the outset, the basis of the academies' policy was to sanction word coinage only according to the principles of the medieval grammarians. This meant that the simple transliteration of foreign words into Arabic (*taʕriːb*) was to be avoided at all costs, and new words were to be derived according to the principle of analogical derivation, using the root and pattern system—*ʔaliʃtiqaːq bi lqiyaːs* 'derivation by analogy'—or by extending the semantic range of archaisms reclaimed for modern use. This meant that the academies' options were limited by the stock of attested roots (whether original or "notional" in the sense introduced in 8.1 above) and the relatively small set of morphological templates. In practice, and particularly in the last forty to fifty years, this self-imposed limitation has proved to be unworkable, and in practice ignored, even by the academies themselves. In the early days, an excessively narrow interpretation of the doctrine that the purity of the language must be preserved led to some odd, even comical results. Archaisms such as *ʔirziːz* 'sound of rain, tremor', *ʔatiː* 'rivulet', and *jammaːz* 'swift-footed camel' were proposed by the Cairo Academy for relaunch as words for, respectively, 'telephone' (to replace the borrowing *tilifuːn*), 'siphon' (to replace the colloquial *ʃaffaːta*), and 'tram' (colloquial *traːm*).[11] None of these words gained any acceptance, because they were outside the stock of roots in active use in the language and were semantically opaque to the nonacademic user. On the positive side, however, the use of analogical derivation has been successful where Arabic had ready-made patterns appropriate for systematic and wide-ranging application in both everyday and scientific contexts. For example, the various patterns for nouns of instrument, *miCCaC, miCCaCa,* and *miCCaːC* have been heavily used in coining words such as *misʕad* 'elevator' (< *saʕada* 'to climb, go up'), *mijhar* 'microscope' (< *jahara* 'to become perceptible, vi-

sually or acoustically'), *mitya:f* 'spectroscope' (< *ṭayf* 'spectrum' (itself originally 'phantasm, vision')), *mis'ar* 'calorimeter' (< *sa'ara* 'to start a fire'), or extending the meaning of words such as *mirwaḥa* 'propeller', which originally meant 'fan'. In these cases the meaning of the new term, or the semantic extension of the old one, was more transparent and hence had a better chance of being accepted by potential users.

However, the basic lexical problem confronting Arabic in building a standardized and internally coherent technical and scientific vocabulary remains: it is poorly equipped with affixes. The Greco-Latin 'un-', 'in-', 'ultra-', 'infra-', 'sub-', 'supra-', 'proto-', 'para-', 'hypo-', '-itis', '-pathy', '-graph', '-scope', '-ide', '-ite', etc., which English and other European languages adopted as convenient building blocks to meet the needs of science, have no morphological analogs in Arabic, and it was felt that a policy of forming composite words (called in Arabic *naḥt*) by joining free-standing prepositions that have similar meanings to these affixes (e.g., *taḥta* 'under' for 'infra-', *fawqa* 'above' for 'ultra-') with content words went against the historical grain of the language. The academicians' preferred solution was to resort to paraphrase, so that, for example, '-logy' is rendered by a noun phrase consisting of *'ilm* 'knowledge, science' followed by the relevant noun and '-itis' by one consisting of *ʔiltiha:b* 'inflammation' followed by the name of a part of the body. Thus to replace the borrowings *biyu:lu:jiya:* 'biology' and *saykulu:jiya:* 'psychology', the construct noun phrases *'ilmu nnafs* (lit. 'science of the soul') and *'ilmu lʔaḥya:ʔ* (lit. 'science of living things') were coined, and as an equivalent of 'arthritis' *ʔiltiha:bu lmafa:ṣil* (lit. 'inflammation of the joints'). But this solution leads to other problems. One is a certain syntactic clumsiness. The word 'indivisibility', and hundreds of Latinate formations like it, have to be rendered into Arabic by rearranging the component morphemes of the source word into lengthy noun phrases:

> *'adamu lqa:bili:ya   li ttajziʔa*
> lack (of) the-susceptibility to the-division

A further, related, difficulty is that relational adjectives equivalent to 'biological' and 'psychological' cannot be derived simply by the addition of the relational suffix *-i:y*, because structures such as *'ilmu lʔaḥya:ʔ* are construct noun phrases, which do not allow suffixation of this type. The solution is either to go back to the source loan word and add the relational suffix, for example, *biyu:lu:jiya:* → *biyu:lu:ji:y*, which is unacceptable to the purists, or to form a relational adjective from a plural noun, *ʔaḥya:ʔi:y* ('living things-pertaining to'), which is, strictly speaking, an offense against the rules of Arabic word derivation (al-

though one honored more in the breach than in the observance). The approved method, which takes the unaugmented root consonants of a word and forms a *CaCaC-i:y* or *CvCC-i:y* relational adjective, for example, *kani:sa* 'church', *kanasi:y* 'ecclesiastical', causes ambiguity in many such cases because the relational adjective "slot" is already occupied in many roots by a word with a different, established meaning. In the cases cited, for example, the relational adjective *ḥayawi:y* (root *ḤYY*, basic meaning 'life') already has many of the literal and metaphorical senses of the English 'vital'; *nafsi:y* (root *NFS*, basic meaning 'breath, spirit') has the long-established basic meaning 'spiritual'. In order to avoid potential ambiguities of this kind, some scientific writers have tended to avoid jettisoning the loan word formations completely.

The desire to preserve the "spirit" of the language by using only derivational patterns based on the root-pattern system has thus proved difficult to reconcile with the consistency, clarity, and concision demanded by modern science. The sheer demand for new coinages is too great, and the root-pattern system by itself has not been able to generate large enough families of morphosemantically related terms analogous to those found in western scientific discourse (e.g., groups of words such as English 'rheumatology', 'rheumatism', 'rheumatic', 'rheumatoid', etc., in which semantically stable suffix morphemes can be freely added to a fixed stem).

Compared with the philologists, Arab scientists have been more pragmatic in their approach, employing a variety of *naḥt*, or 'composite', constructions to create Arabic equivalents of Greco-Latin formations.[12] The simplest type of composite construction involves the compounding, without substantial abbreviation, of the contributing words, for example, *barr* 'land' + *ma:ʔ* 'water' + *-i:y* 'relational suffix' gives *barma:ʔi:y* 'amphibious', *ʔarbaʕ* 'four' + *rijl* 'leg' gives *ʔarbaʕrijl* 'quadruped'. Such formations, while attested, are not very common in CLA. Other more extreme types of portmanteau composites have also been coined, such as *kahraḍawʔi:y* 'photoelectric' < *kahraba:ʔi:y* 'electric' + *ḍawʔi:y* 'pertaining to light', and *šibġarawi:y* 'colloidal' < *šibh* 'likeness' + *ġarawi:y* 'glutinous'. The principle involved in the last example, in which an all-purpose noun or preposition in slightly abbreviated but still recognizable form is prefixed to a content word, has now gained general scientific acceptance and is productive, despite purists' objections, for example, *taḥbaḥri:y* 'submarine' < *taḥta* 'under' + *baḥri:y* 'marine', *ġibjali:di:y* 'postglacial' < *ġibba* 'after' + *jali:di:y* 'glacial'. But more ad hoc composites that involve reduction of the components further than this, such as *nazwarah* 'defoliation' < *nazʕ* 'removal' + *waraq* 'leaves' and *zahraja* 'dehydrogenate' < *ʔaza:la* 'remove' *hi:dru:ji:n* 'hydrogen' have not generally found favor because of the difficulty of relating the reduced components to the

source words and hence of divining the meaning. In cases like this, loan translation of the source word into a multiword Arabic paraphrase is the preferred solution of purist philologists and practical scientists alike.

Modern Arabic scientific vocabulary is thus the result of the efforts of the philologists to ensure that the complexion of its morphology is not at variance with that of the nonspecialized vocabulary of the language, tempered by the less puristic, more pragmatic approach of scientists themselves, for whom clear communication has been the overriding aim in word coinage and usage. This desire for clarity expresses itself in other ways: for example, it is not uncommon in Arabic technical writing to find writers inserting bracketed English translations of Arabic technical terms, which may also be put within quotation marks, where they feel there might be a possibility of misunderstanding if an Arabic phrase or neologism (especially if it is a translation) were to be used *tout court*. This seems to be a way of telling the technical reader " . . . ." is the Arabic expression I am using as the equivalent of the (known) foreign term "(. . .)", for example, in an economics text:

> . . . and creating an accepted formula of measurement, or "*ta?ši:r*" (indexation) which permits the prices of exported raw materials to be linked to those of industrial or manufactured goods imported into developing countries.

8.3 The language of the media

"Newspaper Arabic" began to take shape as a separate stylistic genre early this century, with many syntactic and phraseological innovations (the butt of criticism from purists),[13] as journalism gradually divorced itself from the rather flowery style of nineteenth-century Arabic prose. Today, more than ever, it is in the language of the press, television, and radio that external influences on Arabic are most obvious, and constant exposure to this "media MSA" seems to be having far-reaching effects on the vocabulary, grammar, and phraseology of the Arabic used by educated Arabs in many other contexts, written or spoken. Influence in the reverse direction, from spoken Arabic to newspaper MSA, also occurs, although much of this is indirect, consisting of the avoidance of structures and vocabulary items that, while perfectly correct and acceptable as MSA, have the "taint" of colloquialism because they also occur in the dialects. Some examples of this tendency are given later in this chapter. Only in texts that are seen as marginal to the concerns of serious journalism—such as captions on cartoons—is a written form of the colloquial used. In Egypt, where the dialect has a status somewhat different from those of other areas (see chapter 9), it is oc-

casionally used in newspaper articles but only here in nonserious material, and especially if aimed at women, such as in the fashion pages.[14]

Much of the news reporting in the Arab media is in the form of rapidly produced and often rather literal translations of English or French language news agency reports. In this way, quantities of new phrases are coined ad hoc by journalists and thence find their way into everyday use without ever having received the endorsement of the academies. A few coinages may remain limited to one particular country, and some even to one particular newspaper; others—literal translations of English idioms, for example—may simply be too opaque and remote from the experience of the average reader unfamiliar with the western source languages to have any chance of survival.[15] Some randomly noted recent media examples of apparently successful literal loan translations of foreign phrases and idioms—successful in that they reoccur frequently and in different Arab countries—are:

al-diblu:ma:si:ya al-maku:ki:ya	'shuttle diplomacy' (*maku:k* = shuttle)
'amali:yat al-sala:m	'peace process'
?asliḥat al-dama:r al-ša:mil	'weapons of mass destruction'
al-sayyida al-?u:la:	'the first lady'
al-misma:r al-?axi:r fi: al-na'š	'the last nail in the coffin'
naṣi:b al-?asad	'the lion's share'
siya:sat al-'aṣa: wa al-jazara	'the policy of stick and carrot'
siya:sat farriq tasud	'the policy of divide and rule'
qimmat jabal al-θalj	'the tip of the iceberg'
zawba'a fi: finja:n	'storm in a tea-cup (lit. 'coffee-cup')'
faraḍa ta'ti:man ?i'la:mi:yan	'to impose a news black-out'
?a'ṭa: al-ḍaw? al-?axḍar	'to give the green light'
?iqtaṭa'a juz?an ?akbar min al-ka'ka	'to take (lit. 'cut') a bigger slice of the cake'
daqqa ?isfi:nan bayna	'to drive a wedge between' (e.g., two friends)
naqqa: al-?ajwa:?	'to clear the air'
sarraba al-xabar	'to leak the news'[16]

All of these examples involve the transfer of a foreign language multiword idiom directly into Arabic via literal translation, but the result is not so culturally alien to readers/listeners who do not know the source language from which they come that they cannot divine the meaning in context by themselves. In addition to this kind of phraseological calque, there is also a vast stock of essentially journalistic usages in which the metaphorical extended meaning of a for-

eign concrete item has also been transferred to the Arabic equivalent of that for-eign concrete term, for example, *tajmi:d* (lit. 'making solid, inert') cf. English "freezing" (of assets, parliament, and the like); *haš* (lit. 'crisp, easily broken (e.g., bread)') cf. English "brittle" (ceasefire, lull, truce); *juyu:b* (lit. 'pockets in a shirt') cf. English "pockets" (of resistance, Christian pockets in Beirut); and *ḍuba:bi:y* (lit. 'pertaining to fog, mist') cf. English "foggy" (idea, perception). In some MSA phrases that are by no means marginal or ephemeral, the calque suggests a mis-construal of the meaning of the source expression. Thus *'umla ṣa'ba* is univer-sally used as an equivalent of 'hard currency', even though "hard" in the origi-nal collocation clearly has nothing to do with "hard" in the sense of "difficult," which is what *ṣa'b* means. Similarly *iḥtiwa:ʔ*, the verbal noun of the verb 'to con-tain' is now routinely used in Arabic in collocation with words such as *muqa:wama* 'resistance', *ḥamla* 'attack', and *ʔa:θa:r* 'effects', apparently calqued on the English 'contain (= stop) the attack, etc.', although *iḥtiwa:ʔ* in non-journalistic Arabic only means 'contain' in the sense of 'to hold within itself'. And in a recent novel by an Egyptian writer, one reads *ʔawšaktu ʔan ʔa'u:da ʔadra:ji:....*[17] 'I was about to retrace my steps ...' where *ʔadra:j*, plural of *daraj* meaning 'steps' in the sense of "stairs," is used rather than a word meaning "foot-steps." This is a calque of some age.

Apart from in the news media, there has been a massive wave of new coinages in material culture. Here one can distinguish two concurrent processes. First, there is a growing tendency in all Arab countries for an older layer of borrow-ings from foreign languages to be replaced by MSA neologisms. In some cases, Arab governments actively promote this as part of a general policy of arabi-cization. In Saudi Arabia, for example *ba:ṣ* 'bus', which originally seems to have come into MSA via the dialects (in which it is still universally used), was replaced on street signs and other forms of public writing in the 1970s by the neologism *ḥa:fila*, as was, on telephone booths, *tilifu:n* 'telephone' by *ha:tif*.[18] A second, concurrent wave of vocabulary innovation, involving a variety of coinage prin-ciples and translation strategies, came with the more recently invented para-phernalia of modern domestic and office life. Examples here are *dabba:sa* 'sta-pler' (analogical derivation from *dabbu:s* 'pin' (originally 'mace') using the "instrumental" pattern *CaCCa:Ca*); *xalla:ṭ* (< *xalaṭa* 'to mix', again using the 'in-strumental' pattern *CaCCa:C*) 'food mixer'; *jiha:z taḥḍi:r al-qahwa* (paraphrase, lit. 'device for preparing coffee') 'coffee percolator'; *sa:'a raqmi:ya* 'digital watch' (di-rect loan translation); *mukabbir θuna:ʔi:y al-ittija:h* 'bidirectional speaker' (com-bination of analogical derivation (*mukabbir*) using active participle of Pattern II (*muCaCCiC*) and loan translation); *sayya:ra 'a:li:yat al-ʔada:* 'high-performance car'

(direct loan translation). The MSA phrases in such cases as these seem to have been devised not so much by the language academies as by advertising copy-writers.[19]

Media Arabic is also distinguished from other kinds of written Arabic—lit-erary prose writing, for example—by the frequent use it makes of certain per-iphrastic structures, phraseological calques, and syntactic variants. None of these are absent from other genres but, taken together, they constitute a definite "media style." The commonest of these are exemplified below.

8.3.1 Periphrastic passives

As we saw in chapter 3, the passive of a verb in MSA is regularly formed by a change in its internal voweling, for example, Pattern I *qatala* 'he killed', *qutila* 'he was killed', Pattern X *staqbala* 'he greeted, met', *stuqbila* 'he was greeted, met'. Pat-terns V and VII of some verbs are also used with passive, pseudopassive, or re-flexive meanings, often in particular extended or metaphorical usages, for example, *takawwana* 'to be formed out of, consist of', *ta'arraḍa* 'to be exposed to, expose one-self (e.g., to a danger)', *nfataḥa* 'to become or get opened, to open oneself (e.g., to outside influence)', and *nsaḥaba* 'to be pulled out, to withdraw (of troops)'. In jour-nalistic Arabic, we see not only a particularization of the passivizing resources of the language such that certain types of morphosyntactic structure tend to be used for particular kinds of meaning, but also heavy use of periphrastic structures.

The most striking, perhaps, is the widespread use of the *tamma* 'to come to pass, be completed' + verbal noun periphrasis. This structure is used instead of an internally voweled passive to report the completion of durative or iterative processes (rather than for the description of punctual events, cognitive activi-ties, and emotional states) where the focus is on the result of the process rather than on the process itself or on the (usually multiple rather than individual) agents who performed it. The following are typical, randomly noted in a rapid perusal of a couple of pages of one issue of the newspaper *al-Sharq al-'Awsaṭ* in mid 1992:

> *tamma*   *ntixa:bu John Major raʔi:san li lwuzara:ʔ* ...
> was-completed-3msg election John Major prime minister ...
> 'John Major was elected prime minister ...'

> *tamma*   *tazwi:du ljayš   bi kammi:ya kabi:ra min* ...
> was-completed-3msg supply the-army with quantity large of ...
> 'The army has been supplied with a large quantity of ...'

tamma   ktiša:fu   'ila:j   jadi:d li maraḍi ssaraṭa:n
was-completed-3msg discovery treatment new for disease the-cancer
'A new treatment for cancer has been discovered . . . '

kita:b yatimmu   tada:wuluhu   bi sirri:ya . . .
book 3msg-is-completed circulation-its in secret
'a book that is being circulated clandestinely . . . '

In all these cases, passives formed by internal voweling (ʔuntuxiba John Major,
etc.) would be perfectly grammatically acceptable alternative structures involv-
ing no discernible difference in meaning, but they are certainly less common
in journalistic Arabic when the action reported is "nonpunctual" (in the sense
of not beginning and ending at a single point in time) and are especially liable
to occur where the result of a durative or iterative process is being described.
Where, however, punctual actions are being reported, the internal passive form
is normally used. For example:

qutila   θala:θ atu ʔašxa:ṣ ʔams   ʔaθna:ʔa maʕraka . . .
killed-PASS-3msg three   persons yesterday during   battle . . .
'Three people were killed yesterday in a battle . . . '

Although not ungrammatical, the use of a tamma construction to express this
kind of action, viz.:

tamma   qatlu θala:θ ati ʔašxa:ṣ ʔams   ʔaθna:ʔa maʕraka
completed-3msg killing three   persons yesterday during   battle
'Three people were killed yesterday in a battle . . . '

would sound odd, because death in battle (or as a result of an accident and the
like) is by definition sudden and unpredictable, whereas the tamma construc-
tion expresses the outcome of a planned, purposeful, and usually durative ac-
tivity. So, by contrast, one does typically read:

tamma   ʔiʕda:mu θala:θ ati ʔašxa:ṣ ʔams . . .
completed-3msg execution three   persons yesterday . . .
'Three people were executed yesterday . . . '

The verb 'execute' here describes not just the fact of (presumably instantaneous)
death as in qutila, but implies that death was the end point of a composite
process (trial, sentence, and firing squad). Conversely, in the background to an
account of the lives of Muslims who live on the Silk Road (al-Sharq al-'Awsat,
color section, 18 July 1992), one reads:

wa qad(i) ktušifa   lḥari:r fi: ṣṣi:n fi: 'a:m 2640
and qad   discovered-PASS-3msg the-silk in China in year 2640
qabla lmi:la:d . . .
before the-birth
'Silk was discovered in China in 2640 B.C. . . . '

where an ordinary internal passive is used in contrast to the earlier example de-
scribing the discovery of a new cure for cancer. The implication of the tamma
construction in the cancer example is that the discovery was the result of (prob-
ably lengthy) purposeful activity (e.g., laboratory research), whereas the use of
the internal passive to describe the discovery of silk was a fortuitous "punc-
tual" event: in the following clause the discovery is attributed directly to the ob-
servation of an individual Chinese emperor.

Like "punctual" verbs, verbs expressing cognitive activities, speech, and emo-
tional states are also typically passivized using the morphological resources of
the language rather than with a periphrastic tamma construction. Such verbs are
used to express depersonalized, unattributable, or otherwise "off-the-record"
evaluative or background comment. Because they are not part of the event being
reported, they tend most often to be p-stem (i.e., nonnarrative) forms:

tuʕtabaru   ha:ðihi lmuškila   min ʔahammi   lmaša:kil
3fsg-consider-PASS this   the-problem of more-important the-problems
'This **is considered** to be one of the most important problems . . . '

yustabʕadu   ttixa:ðu miθli   ha:ðihi l?ijra:ʔa:t . . .
3msg-rule-out-PASS taking   likeness these   the-measures . . .
'The taking of such measures as these **is ruled out** . . . '

tuʕarrafu   ssiya:satu ssukka:ni:ya   bi?annaha:. . . .
3fsg-define-PASS the-policy the-demographic in-that-it . . .
'Demographic policy **can be defined** as . . . '

bi kammi:ya:t la: **tuʕaddu**   wa la: **tuḥṣa:**
in quantities not 3fsg-count-PASS and not 3fsg-number-PASS
' . . . in **countless** and **innumerable** quantities' (lit. 'are not counted'/'are not
numbered')

8.3.2 Agent passives

According to the prescriptions of purists, the agent of a passive verb should
not be overtly expressed in its predicate. However, this "rule" is now largely
ignored in journalistic Arabic—another result of the tendency for the syntax of

Arabic journalism to mirror that of European languages. The prepositional phrases *min qibal* or *min ṭaraf* ('from the side/party of') are used with both internally voweled and periphrastic passives to introduce the agent. For example:

> *tamma*   *listi:la:ʔ*   *'ala: lqaṣri*   *lmalaki:y min qibali*
> was-completed-3msg the-capture of   the-palace the-royal by
> *quwwa:ti linfiṣa:li:yi:n*
> forces (of) the-secessionists
> 'The royal palace was captured by secessionist forces'

> *ʔustuqbila*   *lwafdu*   *lmiṣri:y*   *fi: lmaṭa:r*   *min*
> met-PASS-3msg the-delegation the-Egyptian in the airport from
> *ṭarafi*   *wazi:ri*   *dda:xili:ya . . .*
> side (of) minister (of) the-interior . . .
> 'The Egyptian delegation was met at the airport by the interior minister'

It would, again, be perfectly possible to express the content of these latter two sentences by an active verb ('Secessionist forces captured . . .' etc.) without any change in the meaning or any change in the focus of the text. The tendency toward the passive-plus-agent in media writing seems to be a straightforward syntactic transfer from European languages.

8.3.3 Tendency to nominalization

Compared with premodern Arabic prose of all genres and with contemporary literary prose, the language of news analysis and more discursive Arabic writing on economics, and the social and political sciences generally, contains a higher proportion of noun phrases and a lower one of finite verb phrases. Moreover, the noun phrases are longer and more complex than in other kinds of writing.[20] As in the case of the passive structures already exemplified, nominalization is another syntactic ploy that allows the writer to give the required flavor of objectivity to his or her statements and claims and is hence preferable to the alternative of clauses containing finite verbs that have to be specified for person. Nominalization, like passivization, allows "unattributable" claims to be made, for example at its simplest, *huna:ka 'tiqa:d ʔanna . . .* 'there is a belief that...', as opposed to 'I'/'he'/'they' 'believe that...', cf. the passivized example in section (a) above—*yu'taqadu ʔanna . . .* 'it is believed that . . .'—with which the nominalized construction is functionally identical. But more complex examples abound. Typical is the following paragraph-initial sentence from a dissident polemical tract on the need for political change in Kuwait,[21] in which the nom-

inalized phrases in bold (*li lwuqu:fi 'ala: . . . ḥatta: yumkina tta'arruf 'ala: . . .*) are used fore and aft of the main subject-verb complex (itself a depersonalized nominalization: *yalzamu taḥli:l . . . wa taqsi:m*) in order to provide contexts of "depersonalized" purpose:

> *wa li lwuqu:fi*   *'ala: ṭṭa:qati lbašari:ya, yalzamu*
> and for the-knowing of the-power the-human 3msg-is necessary

> *taḥli:lu*   *lmujtama'i ssukka:ni:y*   *wa taqsi:muhu ʔila:*
> analysis (of) the-society the-populational and its-division into

> *fiʔa:t ða:t*   *majmu:'a:t taba'an li ddawri llaði: taqu:mu*
> groups owners (of) subgroups according to the-role that 3fsg-play

> *bihi kullu majmu:'a* ***ḥatta yumkina***   ***tta'arruf***   ***'ala:***
> it each group so that 3msg-is-possible the-knowing of

> *ḥajmi*   *quwwati l'amal wa xaṣa:ʔiṣiha: fi: lḥa:ḍir*
> size (of) force (of) the-work and features-its in the-present
> ***wa ttija:ha:tiha: fi: lmustaqbali lqari:b wa lba'i:d wa***
> and directions-its in the-future the-near and the-far and

> ***kaða:lika 'ala: ʔaḥja:mi lfiʔa:t***   ***lʔuxra: wa***
> like-that of sizes (of) the-groups the-other and

> ***ṭabi:'atiha:***
> nature-their

'In order to grasp (nominal) what the "human potential" is, it is necessary to analyze (nominal) and divide (nominal) the population of a society into groups and subgroups, according to the role each subgroup plays, so as to determine (nominal) the size of the workforce and its characteristics now and its tendencies in the short and long term, as well as the size of and nature of the other groups.'

Relative clause reduction for the sake of compactness and concision is another tendency that adds to the complexity of the nominals in this kind of writing, for example, from the same political tract:[22]

> *'awa:milu dda'mi*   *lmuštaqqa min* ***siya:sa sukka:ni:ya***
> factors (of) the-support the-derived from policy populational

> ***mu'lana***   ***min qibali ddawla wa***   ***mu'taraf biha:***
> announced by the-state and recognized

'support factors derived from **a population policy publicly enunciated by the government and acknowledged as such'**

Second, the impression of heavy nominalization is partly also because of the complex nature of the abstractions written about in much journalism and other texts that deal with politics and allied areas. A high proportion of these are referred to by conventional labels consisting of multiword noun phrases whose components indicate their nature or function. For example:

'*ajzu   mi:za:ni:yati lqiṭa:'i   l'a:mm*
deficit (of) budget (of) the-sector the-public
'Public Sector Budget Deficit'

munaϑϑamatu   dduwali   lmuṣaddira   li   nnafṭ
organization (of) the-states the-exporting of the-oil
'Organization of Petroleum-Exporting Countries' (OPEC)

ʔaljama:hi:ri:yatu l'arabi:yatu lli:bi:yatu   ša'bi:yatu   lištira:ki:ya
the-Jamahiriyya the-Arab   the-Libyan the-popular the-Socialist
'The Socialist People's Libyan Arab Jamahiriyya' (SPLAJ)

Many more ad hoc formulations of equal length and complexity regularly occur, on occasion (see (ii) below) straining the limits of syntactic acceptability. For example:

(i) lengthy construct phrases:
   mi:za:ni:ya:tu ʔakbari xamsi šarika:ti   ʔadwiyatin fi: l'a:lam
   budgets (of) largest five   companies (of) drugs   in the-world
   'the budgets of the world's five largest drug companies' (*al-Sharq al-'Awsaṭ*, 17 September 1991)
(ii) lengthy construct phrases with split accusatives of specification:
   wa:ḥid min ʔakϑari jtima:'a:ti   dduwali ssab'i   nsija:man
   one   of most   meetings (of) the-states the-seven harmony-ACC
   wa tawa:fuqan
   and mutual agreement-ACC
'one of the seven nations' most harmonious and amicable meetings' (lit. 'one of the greatest of the meetings of the seven nations as regards harmony and mutual agreement') (*al-Sharq al-'Awsaṭ*, 17 July 1989)

A third, perhaps more language-internal reason for the nominalizing tendency may be the connotations that the alternative verbal structure carries with it. In the dialects, when there is a choice, the complement of the main verb is much more likely to be a clause, whether linked syndetically or asyndetically, than it is to be a verbal noun. Nominalization may thus be an avoidance strategy: one of the means (the negative system is another) by which written style can be distanced and consciously marked as different from dialectal Arabic. Thus structures of the general type:

ʔaṣarra   'ala: lmaji:ʔ
insisted-3msg on   the-coming
'He insisted on coming'

have a distinctly formal flavor to them, even though, with appropriate morphological adjustments, they can occur in dialectal speech. However, in the dialects, there is a distinct preference for the clausal alternative with a finite verb. For example:

ʔaṣarr   ('ala) ʔinnuh yi:ji
insisted-3msg (on)   that-he 3msg-come

8.3.4 Negativization

As noted in 6.3, in written Arabic there are two ways of negativizing actions/states conceived of as completed or having come to an end: the s-stem form of the verb preceded by *ma:* or the base set of the p-stem preceded by *lam*, for example, *ma: kataba/lam yaktub*, both of which can be translated as 'he did not write' or 'he has not written' in the appropriate context. The original difference between these forms in CLA, it has been claimed,[23] lay not so much in their temporal or aspectual reference as in the degree of personal participation and "affect" in the action or state expressed by the negativized verb: the *ma:* construction was typically used for direct (especially first person) speech and the *lam* construction was more common in the description of third person action. In the dialects, the *lam* construction disappeared, and *ma:* came to be used as the negative particle with both s-stem and p-stem form verbs. This latter fact helps to explain the pattern of negative verb usage in MSA. Although a few literary figures of the older generation, deeply schooled in linguistic traditions of CLA, still cling to the original functional differentiation of the two constructions, in the nonliterary written usage of journalism and other expository text, the *lam*

construction has come to predominate overwhelmingly in all contexts where a past-time negative is required. The use of the *ma:* construction has come to be associated either with a consciously "literary" style or with dialectal usage; on both accounts it is avoided in nonliterary MSA writing.[24] For the same reason—the dialectal "taint" on *ma:* as a negative particle—the CLA distinction in the u-set p-stem verb expressing noncompleted action, between *ma: aktubu* 'I am not (at the moment) writing' and *la: aktubu* 'I will not/do not/cannot write' has also ceased to be observed in MSA, the "nondialectal" *la:* form now being normal for all shades of modal/aspectual meaning expressed by the u-set p-stem verb.

These developments are another example, like the case of nominalization already discussed, of the conscious distancing of written Arabic from anything that smacks of dialect. In passing, it is worth noting that many other examples of this tendency could be cited in the choice of lexical items in MSA texts. The most obvious is the avoidance of the verb *qadara* 'to be able'. Although common in CLA, it has become unusual to encounter the finite forms of this verb in many genres of MSA. The synonym *?istata:'a* is preferred because it has no dialectal analog form, as *qadara* has, and so is saliently "formal."

8.3.5 Word order, time reference, and text type

Normal word order in verbal sentences in CLA was typically VSCOMP regardless of the semantic characteristics or the time reference of the verb occupying the V slot. As we observed in 7.1, word order in MSA sentences containing these same elements is becoming more stylistically determined. Like the nominalized verb phrase, word order is another element of syntax that can be exploited to mark a text as an example of a particular genre or to mark the status of particular sections of a text. For example, according to Somekh, there is now a tendency among novelists who employ a stream of consciousness technique to distinguish "interior monologue" from "exterior occurences" by reserving the SVCOMP order for the former and VSCOMP for the latter.[25] In the field of journalism, subtle and consistent use of different word orders is also employed to mark different text types. Almost invariably, newspaper headlines that contain main-clause verbs have SVCOMP, whatever the morphological or semantic characteristics of the verb, or the time reference of the headline content. Some randomly selected examples (mid-1992 from Saudi and Jordanian newspapers) will illustrate this:

p-stem verbs:

	S	V	COMP
Baršalu:na		*tuḥaqqiqu*	*ḥulmaha:*

| | | Barcelona | 3fsg-realize | dream-its |

'Barcelona realizes its dream'
(written a week before the opening of the 1992 Olympics; article reports progress of preparations)

		S	V	COMP
wafd	*min*	*mu?ta*	*yazu:ru*	*mudi:ri:yata*	*l?amni*
delegation	from	Mu'ta	3msg-visit	directorate (of)	the-security

l'a:mm
the-public
'A delegation from Mu'ta visits the Directorate of Public Security'
(written the day after the visit; article reports the details)

	S	V	COMP
?aššurṭa	*l?isra:?i:li:ya*	*tastajwibu*	*'abd ašša:fi'*	*wa*	*'ašra:wi:*
the-police	the-Israeli	3fsg-interrogate	Abdul-Shafi' and 'Ashrawi

'The Israeli police interrogate Abdul-Shafi' and 'Ashrawi'
(written the day after the interrogation; article reports what happened)

s-stem verbs:

	S	V	COMP
?al'arabu	*la'ibu:*	*dawran*	*?asa:si:yan fi:*	*qimmati*	*l?arḍ*
the-Arabs	played-3pl	role	basic	in summit (of)	the-earth

'The Arabs played a key role at the Earth Summit' (written three weeks after the end of the summit; article summarizes Arab contribution)

	S	V	COMP
?almuslimu:na	*ta'arraḍu:*		*li θala:θ i maða:biḥ*	*xila:la*
the-Muslims	were-exposed-3pl	to three	massacres	during

mi?ati   sana fi: lbu:sana
hundred   year in   the-Bosnia
'The Muslims have been massacred three times in the last century in Bosnia'
(written during a period of Serbian attacks on Muslims in Bosnia; article is an account of their historical sufferings)

The predominance of the SVCOMP order in the headlines is all the more interesting when it is contrasted with the reports themselves. Where the text is a straightforward narrative, in which the news content of the headline is merely replicated and amplified (as occurred in the texts headlined by the first three examples above), the word order of the sentences is always VSCOMP, and the "historic present" (p-stem) of the headline is replaced by the "factual narrative"

(s-stem) of the text. Thus the first line of the report that appears under the second of the headlines quoted above begins:

V		S		
za:ra	wafd	min ṭalabati	ja:miʕati	muʔta
visited-3msg	delegation	of students (of)	university (of)	Muʼta

		COMP		
ʔams	mudi:ri:yata	lʔamni	lʕa:mm . . .
yesterday	directorate (of)	the-security	the-public

'A delegation of students from Muʼta University yesterday visited the Directorate of Public Security . . . '

The contrasting use of the p-stem verb in headlines and the s-stem in the text parallels European journalistic conventions (cf. "David Mellor *admits* affair" in the headline "David Mellor, the Heritage Secretary, yesterday *admitted* . . ." in the text). In cases where, on the other hand, an s-stem verb is used in the headline, it, generally speaking, has no news-reporting value but rather makes a general summarizing statement about some news event that is being discussed in depth in the text, rather than simply being reported on as "news." This is exemplified in the fourth and fifth examples, which are taken, significantly, not from the front-page news stories but from the "features" section. What of the word order difference between headline and text? Headlines are of course a very particular kind of text: they present the subject matter in as concise a way as possible to enable readers to decide instantly whether they wish to take the time to read the detailed report. In many cases, in Arabic as in English, this essential information is put baldly in a verbless noun phrase:

muṣa:dama:t	ʕani:fa	fi:	ʔanḥa:ʔi	ḍḍiffa	wa	lqiṭa:ʕ
clashes	violent	in	areas (of)	the-Bank	and	the-Strip

'Violent clashes in parts of the West Bank and Gaza Strip'

Verb phrases such as " . . . occurred yesterday" are hardly necessary in such headlines; but if they are added perhaps to enable further information such as the time and place of an occurrence to be included, it is more efficient for them to follow rather than precede the informational focus, if, as in Arabic, there are no syntactic constraints on this happening. It is this need to give priority to the information focus that explains the thematization of the subject in the SVCOMP order of the headlines.

Whereas the sentences in straightforward news reporting tend to have a VSCOMP structure (very often with the preposing of temporal, purposive, or other modifying phrase), the word order in more discursive writing is more variable and idiosyncratic. However, there is a discernible general tendency in texts that involve a mixture of narration and reflection or contextualizing background comment for the narrative sections that advance the action to have the VSCOMP order and for the nonnarrative to have SVCOMP. An example is a description of a journey into the Amazon rainforest (*al-Sharq al-'Awsaṭ*, color section, 18 July 1992). There is a main "narrative" framework describing the journey, into which a background commentary on the difficulties of the Brazilian government in balancing the preservation of the environment with the demands of the economy is inserted. There is a constant interplay of VSCOMP and SVCOMP sentence orders. The main clauses of the first few paragraphs (translated here), in which there was an open choice between the two orders (i.e., where there was no preposed phrase automatically triggering VSCOMP) divided up as follows:

VSCOMP
'The Brazilian Economics Minister *did not deviate* from the truth when he said . . . '
'All efforts to arrange a private visit to the heart of the jungle *failed* . . . '
'The journey *lasted* six hours from the time we set off . . . '
'The aim *was* to reach the point where the Rio Negro, that is "the black river," meets . . . '
SVCOMP
'Studies *show* that the slashing and burning of forests not only destroys the trees and wildlife, but deprives future generations . . . '
'This imbalance *spreads* in all areas and leads to . . . '
'But Brazil *is not* a poor country'
'Its current poverty *has not happened* overnight'
'Its poverty *results* from its economic policy'
'Truth *is* sometimes more impressive than imagination'
'The waters of the black river *carry* minerals, salts . . . '

It can be seen from this that the VSCOMP sentences form part of the narrative telling the story of the journey, whereas in the SVCOMP sentences the writer temporarily steps out of the narrative to provide a commentary on general environmental and economic issues, or in the last sentence to explain how the

Rio Negro acquired its name. It is no coincidence that, in texts like this, the verbs in VSCOMP sentences tend to be s-stem (completed actions), and those in SV-COMP p-stems (general truths, habitual actions, etc.).

8.3.6 Certain morphosemantic developments

We have already noted the problems involved in adapting the nonaggluti-native morphology of Arabic to the vocabulary needs of modern science. Similar problems confront those writing for a more general audience on politics and economics, although in terms of the range of categories of technical mean-ing to be covered, the problem is less acute. The method of choice in the coin-ing of neologisms in "media" Arabic is paraphrase, rather than the more out-landish *naḥt* composite words that some scientists, and Arab nationalists such as Sāṭiʿ al-Ḥuṣarī,[26] have favored. Certain constructions are of very high frequency:

(i) (English/French) 're-'

The analog of English/French verbs that have a flexible "re-" prefix with the meaning "again" is a verb phrase consisting of *ʔaʿaːda* 'to make (something) re-turn' + the relevant verbal noun. For example:

ʔaʿaːda	*tanðiːm* . . .	'reorganize'
make-return	organization (of) . . .	
ʔaʿaːda	*nnaðˁar fiː* . . .	'review'
make-return	the-looking in . . .	
ʔaʿaːda	*binaːʔ* . . .	'rebuild'
make-return	building (of) . . .	
ʔaʿaːda	*ntixaːb* . . .	'reelect'
make-return	election (of) . . .	
ʔaʿaːda	*ktišaːf* . . .	'rediscover'
make-return	discovery (of) . . .	
ʔaʿaːda	*kitaːbata taːriːx* . . .	'rewrite the history (of)'
make-return	writing (of) history (of) . . .	

(ii) "non-," "mis-," (French "non-," "mal-'")

Abstract nouns formed with Greco-Latin negative affixes generally form their Arabic equivalents by paraphrase. Straightforward negatives ("non-," "in-," etc.) typically make use of the noun *ʿadam* 'lack of' + verbal noun to form noun phrases, whereas phrases describing badly or wrongly performed actions ("mis-") use *suʔ* 'badness':

NOUN	NOUN	NOUN	
ḥarakatu	*ʿadami*	*linhiyaːz*	'nonaligned movement'
movement (of)	lack (of)	the-partiality	

NOUN	NOUN	ADJ	
ʿadamu	*listiqraːri*	*liqtiṣaːdiːy*	'economic instability'
lack (of)	the-stability	the-economic	

NOUN	NOUN	ADJ	
suːʔu	*lʔidaːra*	*liqtiṣaːdiːya*	'economic mismanagement'
badness (of)	the-running	the-economic	

NOUN	NOUN		
suːʔu	*ttaɣðiya*		'malnutrition'
badness (of)	the-nutrition		

(iii) English/French "multi-," "poly-"

English/French compounds in "multi-" (and "poly-") commonly form adjec-tive–noun phrases that consist of the active participle *mutaʿaddid* 'numerous' and a plural noun and function as a single adjective. This type of use of the con-struct phrase to form compound adjectives is extremely common in journalistic MSA. For example:

	ADJ PHRASE		
HEAD NOUN	ADJ	NOUN	
quwwaːt	*mutaʿaddidatu*	*ljinsiːyaːt*	'multinational forces'
forces	numerous (of)	the-nationalities	
mufaːwaḍaːt	*mutaʿaddidatu*	*lʔaṭraːf*	'multilateral negotiations'
negotiations	numerous (of)	the-sides	
mujtamaʿ	*mutaʿaddidu*	*lʔaθniːyaːt*	'multiethnic society'
society	numerous (of)	the-ethnicities	
qaːʿa	*mutaʿaddidatu*	*listiʿmaːlaːt*	'multipurpose hall'
hall	numerous (of)	the-uses	

The corresponding verbal noun *taʿaddud* 'multiplicity' or a secondary nominal extract *taʿaddudiːya* (see (iv) below) is used in nominalizations (e.g., *taʿaddudu zza-wjaːt* 'polygamy' (lit. 'multiplicity of wives')), or noun–adjective phrases (e.g., *taʿaddudiːya siyaːsiːya* 'political pluralism' (lit. 'political multiplicity')) .

(iv) English/French "anti-"

The active participle *mu'a:din* 'inimical, hostile' and its corresponding verbal noun *mu'a:da:* 'hostility' are used with the preposition *li* to form phrases equivalent to "anti-," e.g., *mu'a:din li ssa:mi:ya* 'anti-Semitic' (lit. 'hostile to Semitism'), *mu'a:da: li rra:sma:li:ya* 'anticapitalism'.

(v) "-ity," "-ism," "-ness," etc. (French "-ité," "-isme," etc.)

The vocabulary of English/French for denoting abstract qualities, movements, and tendencies is of course full of nouns ending in these suffixes. As noted in 4.3.2(i), MSA freely suffixes *-i:ya* to almost any noun, singular or plural, as well as to pronouns and even interrogatives, to form equivalent abstractions. This suffix is particularly productive in the social, political, and economic fields. For example:

wuṣu:l (v. noun I)	'arrival'	*wuṣu:li:ya*	'careerism'
?inhiza:m (v. noun VII)	'defeat'	*?inhiza:mi:ya*	'defeatism'
?ištira:k (v. noun VIII)	'sharing'	*?ištira:ki:ya*	'socialism'
ra?s ma:l (noun, composite)	'capital'	*ra?sma:li:ya*	'capitalism'
?uṣu:l (noun, pl.)	'origins'	*?uṣu:li:ya*	'fundamentalism'
ṭa:?ifa (noun, f.)	'sect'	*ṭa:?ifi:ya*	'sectarianism'
waṭan (noun, m.)	'homeland'	*waṭani:ya*	'patriotism'
'unṣur (noun, m.)	'race'	*'unṣuri:ya*	'racism'
na:ṣir (proper name)	'Nasser'	*na:ṣiri:ya*	'Nasserism'
miṣda:q (adj.)	'truthful'	*miṣda:qi:ya*	'credibility'
?imka:n (verbal n. IV)	'possibility'	*?imka:ni:ya*	'potentiality'
mawḍu:' (p. part.)	'topic'	*mawḍu:'i:ya*	'objectivity'
ða:t (noun, f.)	'self'	*ða:ti:ya*	'subjectivity'
jins (noun, m.)	'type'	*jinsi:ya*	'nationality'
huwa (pron.)	'he'	*huwi:ya*	'identity'
?ana (pron.)	'I'	*?ana:ni:ya*	'selfishness'
kayfa (interrog.)	'how?'	*kayfi:ya*	'quality'
kam (interrog.)	'how much?'	*kammi:ya*	'quantity'

In most cases, corresponding adjectives can also be freely formed by the suffixing of the relational *-i:y* to the noun stem: *wuṣu:li:y* 'careerist, arriviste', *waṭani:y* 'patriot(ic)', *'unṣuri:y* 'racial, racist', *na:ṣiri:y* 'Nasserite', *?ana:ni:y* 'selfish', etc. A recent and more radical analogical development is the formation of relational adjectives from noun phrases rather than nouns. For example, *?aššarqu l?awsaṭ*

'the Middle East', a definite noun phrase, has recently acquired a compound adjective *šarq ?awsaṭi:y* 'Middle Eastern' that is written in Arabic as two words even though it behaves syntactically as if it were a single-word adjective. In a similar vein, some journalists have recently started to use a compound adjective *'a:lam θa:liθi:y* 'Third World' derived from the calqued noun phrase *?al'a:lamu θθa:liθ* 'the Third World'. Such compounds differ from the now generally accepted portmanteau scientific neologisms such as *barma:?i:y* 'amphibious' described earlier in that their components are not abbreviated and joined according to the precepts of *naḥt* to form a composite single word that can be fitted into normal syntactic structures. They are considered a transgression against MSA grammatical rules, eschewed by the purists. Nonetheless, they are commonly used, and it is not difficult to see why. In the Arab media, a rapid and pressured transfer into Arabic of foreign-language material often has to be made, in a text-type, news reporting, in which the aesthetic sensibilities of the readers can largely be ignored. The emphasis is on rapid, economical, and clear communication. As we saw in an earlier example, these pressures can have direct results on the form of what is written: a foreign phrase such as "an artillery and rocket barrage" poses a problem because although there is an adjective *midfa'i:y* 'pertaining to guns, artillery' that can be used to modify *qasf* 'barrage', there is no generally accepted adjective corresponding to 'rocket' that could be used in an equivalent Arabic phrase with the structure:

NOUN	ADJ	+	ADJ
barrage	pertaining-to-artillery	and	pertaining-to-rocket
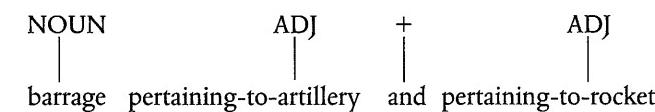
In the event, an adjective *ṣa:ru:xi:y* (not listed in any dictionary), derived from *ṣa:ru:x* 'rocket' seems to have been coined so as to fill a slot in the calqued phrase and provide the morphological parallel with *midfa'i:y*. Such phrasal parallelism is an important stylistic principle in written Arabic, as we shall elaborate in 8.4 below. The alternative would have been to sacrifice structural concision and lengthen the phrase by using parallel prepositional complements (again note the need for phrasal balance):

NOUN	PREP. PHRASE	+	PREP. PHRASE
qasf	*bi lmidfa'i:ya*	*wa*	*ṣṣawa:ri:x*
barrage	by the-artillery	and	the-rockets
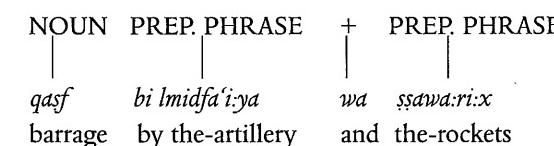
Formations such as *šarq ?awsaṭi:y*, typically used in phrases such as

NOUN   ADJ   +   ADJ
 | | | |
jawlatuhu   *l?urubbi:ya*   *wa*   *ššarq ?awsati:ya*
tour-his   the European   and   Middle-Eastern
'his European–Middle-Eastern tour' (*al-Sharq al-'Awsat*, 11 May 1992)

are simply a further step down the same road, whereby novel analogical derivations arise out of the twin pressures for internal phrasal balance and textual economy against the background of foreign-language source and parallel texts that are full of similar formations. But in contexts where notions of correctness are uppermost, such novelties are avoided: the phrase "Center for Middle Eastern Studies," for example, used in many universities, invariably translates on name plaques and letterheads as *markaz li dira:sa:ti ššarqi l?awsat* (lit. 'center for studies of the Middle East') rather than *markaz li ddira:sa:ti ššarq ?awsati:ya*.

8.4 Repetition and parallelism

So far in this chapter we have been considering features of MSA vocabulary, phraseology, and style typical of expository nonliterary prose. The language in such writing has been permeated by external influences at all levels of structure, from the compounding of words in science and technology to the calqueing of foreign phrases and syntactic patterns in journalism. The structure and idiom of the Arabic used in these spheres is the end result of a gradual adaptation of the resources of the language to what has now become a global scientific and media culture. But what of modern literature, which could perhaps be expected to reflect local values and patterns of language use not calqued on external models? Literary style in modern Arabic is a vast subject,[27] and to conclude, we shall briefly consider linguistic aspects of the traditional "literary" prose style, leaving the variable use of MSA and dialect in modern drama, prose, and poetry to the final chapter, which is devoted to the general question of "language level."

Until relatively recently, serious prose writing in Arabic was characterized by recurrent patterns of language that to the western eye (and ear) have a rather "poetical" feel to them. Assonance, rhyme, paranomasia (i.e., root-pattern echo and repetition) are intricately interwoven to produce balanced juxtapositions of sounds, words, phrases, and sentences. These surface structural effects are counterbalanced by semantic patternings of various degrees of synonymy and antonymy. On topics and in contexts that seem anything but literary in the western sense of that term, it is still not at all uncommon to find whole paragraphs

of carefully crafted parallelistic and repetitive phrase making, which, for western tastes, can often seem incongruously and tediously flowery. The general semantic effect has been described (for biblical Hebrew, although the same description could be applied to countless Arabic examples) as follows:

> 'The poetic texts consist of verses [here the conventional divisions of the biblical text, not anything to do with the prose–verse antithesis] formed from two— or more rarely three—stichoi combined, in which the stichoi or members are in some sense "parallel" to each other, in that they offer variations on the same idea. This may come about by the second member repeating the content of the first in different words (synonymous parallelism), or it may be that it sets it off sharply with contrasted thought (antithetic parallelism), or it may be that it simply takes the thought further and completes it (synthetic parallelism).[28]

The following example is the beginning of an article by the great Egyptian cleric and social reformer Muḥammad 'Abduh (1849–1905) that appeared in *al-Waqā'i' al-Miṣrīya*, the Egyptian government newspaper, on 19 October 1880 on the subject of charitable foundations.[29] It gives a flavor of the genre of "poetic prose" in general and of his style in particular. I have deliberately chosen a newspaper example because it shows the extent to which writers like 'Abduh, who had lived in Europe and knew European languages, seemingly still felt obliged to use an ornate Arabic prose style, even when they had no literary aim in view. The juxtaposed parallel and/or repetitious phrases in the original and in the gloss line are displayed one underneath the other with linking vertical arrows:

?inna ma: taθluju   bihi   ṣṣudu:r
TOP what 3fsg-be-cool at-it the-breasts
                          ↓
    wa   tarta:ḥu   lahu nnufu:s
    and 3fsg-be-relieved at-it the-spirits
                          ↓
    wa   yab'aθuna: 'ala: θθiqati bi ḥusni   mustaqbalina:
    and 3msg-make-us to   the-trust in goodness (of) future-our
ma: nara:ḥu   min
what 1pl-see-it of
?iqda:mi   ?abna:?i qutrina: 'ala: l?a'ma:li lxayri:ya
proceeding (of) sons (of) country-our to the-deeds the-charitable
↓   ↓
wa   jiddihim   wa naša:tihim   fi: ta?li:fi   lkalima

and assiduity-their and energy-their in reconciling the-word

$\downarrow$

wa ḍammi   ššamal
and joining-together the-union

$\downarrow$

wa ttiḥa:di   lmaqṣid li naja:ḥi   lbila:d   wa taqaddumiha:
and unity (of) purpose for success(of) the-country and progress-its

$\downarrow$

wa ʔaxðihim   bi lwasa:ʔili lḥaqi:qi:ya   . . .
and taking-their of the-means the-true

Literal translation:

> What soothes the breast, and relieves the spirit, and gives us confidence in our future is what we see of our countrymen performing charitable deeds, and acting seriously and energetically in bringing everyone together and everything together, and their unity of purpose for the success of the country and its progress, and their adoption of the true means to . . . .

The text begins with a cleft sentence in which three parallel noun clauses occupy the first part of the cleft. These three clauses show various degrees of morphological, syntactic, and semantic similarity. In the first two clauses, taθluju bihi ṣṣudu:r and tarta:ḥu lahu nnufu:s, we have near identity of surface syntactic structure (in word order, number of words, and prepositional verb complementation), as well as the rhyme and the isomorphism of the nouns in subject position ṣudu:r 'breasts' and nufu:s 'spirits' (plural pattern CuCu:C). Although the two phrases mean exactly the same thing, as well as being structurally the same, there is a contrast between the concreteness of reference of the first (coolness/breast) and the abstractness of the second (relief/spirit). The third phrase exhibits less structural parallelism but makes more specific the reason for the positive feelings—itself a signal that the "loop" of repetitious phrases is at an end and that the argument will now proceed. The second half of the cleft sentence, beginning "what we see of . . ." is itself an embedded cleft, paralleling the first half ("what soothes the breast . . . is what we see of . . . ") and its complement similarly involves a list of syntactically parallel phrases with the general structure:

VERBAL NOUN (of) N +   PREPOSITIONAL COMPLEMENT
ʔiqda:mi   ʔabna:ʔi quṭrina: ʕala: . . .
jiddi   -him ⎤
naša:ṭi   -him ⎦   fi: . . .

ttiḥa:di   lmaqṣidi   li . . .
ʔaxði   -him   bi . . .

Within this macrostructure there are microstructures that also involve balanced repetition: taʔli:fi lkalima and ḍammi ššamal are construct phrases consisting of verbal noun + object complement that mean the same thing; the two verbal nouns naja:ḥ 'success' and taqaddum 'progress' annexed to lbila:d are also virtual synonyms.

A line or two later 'Abduh reveals that the subject of his article is jam'i:ya:t xayri:ya 'charitable foundations', which he describes in phrasal couplets consisting of matched pairs of structurally identical or near-identical verb forms with opposite meanings (taxtalifu → tattaḥidu, tata'addadu → tattafiqu) predicated of broken-plural nouns with similar morphological shapes but antithetical meanings:

taxtalifu   ʔaška:luha:
$\diagup$3fsg-be-different   forms-their
(1) $\downarrow$ $\downarrow$
$\diagdown$wa tattaḥidu   maqa:ṣiduha:
and 3fsg-be-one   objectives-their

$\downarrow$ $\downarrow$

wa tata'addadu   ʔama:kinuha wa ṭuruqu sayriha:
$\diagup$and 3fsg-be-many   places-their and ways (of) running-their
(2) $\downarrow$ $\downarrow$
$\diagdown$wa tattafiqu   ġa:ya:tuha: wa fawa:ʔiduha:
and 3fsg-agree   aims-their and benefits-their
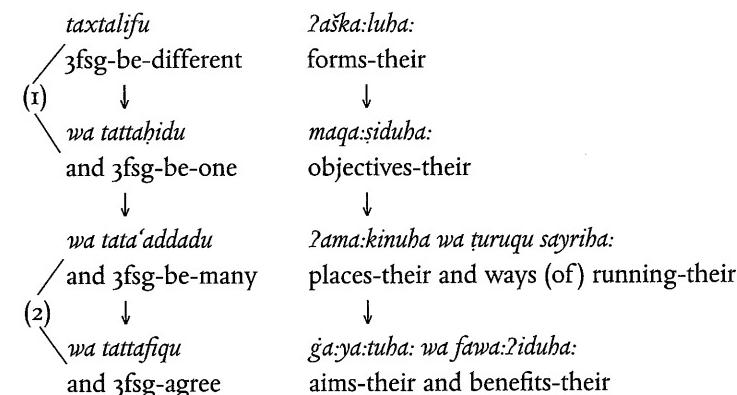
'whose forms differ but whose objectives are one, whose locations and administrative arrangements are many, but whose aims and benefits coincide'.

The rest of the text continues in a similar vein. It has been argued by Barbara Johnstone that this kind of exploitation of the phonic, morphological, and syntactic resources of Arabic is no mere surface ornamentation, but argument by presentation: "an arguer presents truths by making them present in discourse: by repeating them, paraphrasing them, doubling them, calling attention to them. . . . " "Arabic argumentation is structured by the notion that it is the presentation of an idea—the linguistic forms and the very words that are used to describe it—that is persuasive, not the logical structure of proof which Westerners see behind the words."[30] On reflection, however, this claim seems an overstatement. On the one hand, there have been periods in Arabic literary his-

tory when genres of writing involving "argumentation" have come to the fore but in which repetition and paraphrastic parallelism were hardly used at all— one thinks, for example, of early grammatical treatises, of medieval scientific writing, and of the *Muqaddima* of the sociologist and polymath Ibn Khaldūn. On the other hand, this style is not culture specific: the history of rhetoric in many cultures, including English, shows some periods when a similarly repetitious and ornate style was in vogue and others when it was not.[31] And if this claim about Arabic were valid, how would we explain the switch to a more succinct style in contemporary expository writing, unless we perversely assume that modern Arab writers are determined to be less persuasive than their forebears? The function of repetition and parallelism does not, at bottom, seem to have anything to do with "argumentation" as such, and it is worth considering briefly the genesis of this style, because its historical resonances and evolution have relevance to its use today.

The parallelistic, repetitive patternings of Arabic prose, as we have noted, are commonly labeled by westerners (though not Arabs) as "poetic." The first point to make is that the demarcation line between the terms designating "poetry" and "prose" in Arabic does not correspond, either formally or functionally, to the distinction between the equivalent terms in European languages. In the past, *ši'r* 'poetry' referred strictly to verse that , minimally, adhered to one of a limited number of Classical meters and was monorhymed and end stopped. Other popular forms of what would be termed "poetry" in Western cultures, for example, rhymed but nonmetrical verse, were not considered *ši'r* in the narrow sense. In addition to this substandard poetry, there was *saj'*, conventionally translated as 'rhymed prose'. The earliest examples of this are fragmentary, pre-Islamic oracular utterances, and echoes of the same style, nonmetrical but rhyming, and apocalyptic in tone, can be found in the early Meccan *su:ra*s of the Koran. After the coming of Islam, *saj'* continued to be used in oratory and was the normal style in 'sermons' (*xuṭba*, pl. *xuṭab*), the surviving examples of which show a more elaborated species of parallelistic rhyming prose,[32] a prototype of the style that was to be further developed in the ninth and later centuries by literateurs such as Ibn al-Nubātah, who wrote ornate, exhortatory epistles to the Muslim community on a wide variety of religious and religiosocial themes such as the desirability of marriage.

Despite their formal differences, what *ši'r, xuṭba* and *saj'* had in common was rhyme and phrasal balance, which set them apart from ordinary speech and put them into a general category of what might be termed "elevated discourse," which had specific societal (even political) functions. Among these were the affective ones of encomium, satire, exhortation, and persuasion. "Argumentation"

in itself has never constituted a separate genre of speech or writing in Arabic, but to the extent that an argument depends for its effectiveness on the creation or exploitation of an emotional rapport between speaker/writer and audience/reader, it was natural that those parts of texts where a case was being argued should be embellished by the use of an orational style, one of whose chief functions was to persuade and exhort. It comes as no surprise that when, in 1851, the Egyptian translator Rifā'a Rāfi' al-Ṭahṭāwī produced the first Arabic translation of a European novel, Fénelon's didactic tale *Les aventures de Télémaque* (1696), he rendered it using rhyming prose, parallelism, and synonymous pairs of phrases throughout[33]—stylistic features completely absent in the original but apparently considered by the translator to be necessary in Arabic if the translation was to "count" as proper literature at all and if the didactic purpose of the author was to come through clearly to the Arab reader.

Although in the last century or so contact with the West has exposed speakers and writers to other less ornate models, the use of such a linguistically convoluted prose style continues to be favored, even to the present day, in writing (and speaking) intended to persuade. Parallelistic and repetitious discourse— even if using hackneyed and well-worn phrases—is likely to be encountered (and is expected) in formal speech making on such occasions as the opening of universities and assemblies, at passing-out parades and graduation ceremonies, on "national days," and at victory parades. President Nasser of Egypt was particularly fond of this style, albeit in a form sufficiently diluted to allow ordinary Egyptians to understand him.[34] In writing, the style is much used in religious polemic; in reflective, moral-philosophical tracts; and in literary criticism and literary writings in general that start from a nonwestern, nonsecular critical perspective, but not now, as was the case 100 years ago in Muhammad 'Abduh's time, in more neutral bread-and-butter contexts where facts are being presented rather than a case argued for. It is associated with writers who see themselves as continuing the literary traditions of Classical Arabic, and ordinary Arabs recognize it as evidence of the writer's excellent style and command of the language—a fact that is in itself likely to persuade some readers of the validity of the arguments.[35] The use of repetition and parallelism is not exclusive to Muslim writers, despite its religious associations: the Christian Lebanese emigré Mikha'il Nu'ayma was particularly fond of it in his philosophical essays.[36] In modern journalism, political commentators and those with moral axes to grind also occasionally make use of it to lend weight to their arguments.[37] The use of rhyme and parallelism in book titles remains a means by which writers on the most "prosaic" of topics make ritual obeisance to "good style": a recent example is the rhyming title of a propagandistic account of the Gulf War of

1990–1991, *Ḥaqā'iq al-Iḥtilāl wa Watha'iq al-Taḥrīr*[38] 'The Facts of the Occupation and the Documents of the Liberation'.

Notes

1. Fück (tr. Denizeau) 1955 outlines the lexical development of Classical Arabic over the period A.D. 632–1258.
2. Donner 1981, 95.
3. Johnstone 1967, 57–58 for further examples.
4. Somekh 1991, especially 21–46, 93–107, 119–28.
5. See Suleiman 2003, 79 and 162–223 for a detailed discussion of the divergent ways in which nineteenth- and twentieth-century intellectuals in Syria, Egypt, and Lebanon sought to incorporate the Arabic language as a key element in local conceptualizations of Arab nationalism.
6. Hamzaoui 1965, 9.
7. See Chalabi 1984, 287–88 and Ali 1987, 48–55 for a discussion and examples of the different (and sometimes conflicting) coinages and coinage principles adopted by the academies in different countries.
8. In only a few Arab countries, most notably Syria, is all undergraduate teaching and teaching materials, even in the sciences and medicine, in Arabic.
9. For example, the frequent use of the originally purely locative particle *huna:ka* 'there' in an existential sense as an equivalent of English 'there' in 'there is' or French 'y' in *il y a* (e.g., *laysa huna:ka badi:l* 'there is no alternative') does not cause raised eyebrows among the purists. This construction certainly existed in CLA but was nowhere near as common a locution as it has now become in MSA, undoubtedly because of the influence of European languages. Blau (1981b, 6off.) attributes many of the more recent developments in the syntax of written Arabic, which continue this trend, to the influence of what he calls (after Blanc 1957, 400) Standard Average European (SAE) syntax, which enters Arabic via literal translation from (mainly) English and French.
10. Much medieval popular literature, in the oldest recensions we have, is written in a hybrid Arabic of this type (Pinault 1992, 14–15), a product of the need for a consciously elevated and appropriately "literary" style to be reconciled with the conditions of oral performance before a largely nonliterate audience. In ephemeral, nonliterary material, especially that of non-Islamic origin (medieval Judeo and Christian Arabic), the deviations from the norms of written Classical Arabic are even clearer (Blau 1981a, 1966–67 passim).
11. Stetkevych 1970, 32–33.
12. See Ali 1987, 69ff. for numerous examples.
13. Ibrāhīm al-Yāzijī's *Lughat al-Jarā'id* (*The Language of Newspapers*), published in 1901, lists several hundred words and phrases used in the press at that time that the author considered unacceptable and/or incorrect.
14. Somekh 1991, 17–18. A tendency in Egyptian popular journalism to mix written colloquial and MSA in a more deliberate and consistent fashion, apparently for stylistic effect,

has recently been noted by Rosenbaum (2000), a style that he calls *Fuṣḥāmmiyya*. This style is found in narrative sections (i.e., is not merely a means of reporting verbatim colloquial speech) and involves abrupt shifts from one code to the other and back again (see 9.5.3).

15. Two extreme examples of loan translation that have appeared recently in the Jordanian press (which have not been observed elsewhere and caused amusement to a large gathering of educated Arabs from different countries) were provided by Dr. Yousef Badr in a paper on loan translation delivered at the First International Conference on Translation and Technology at Yarmouk University, Jordan, 11–15 April 1992: *baṭṭa 'arja:ʔ*, a literal translation of the English 'lame duck', which appeared in the Jordanian press around the time of the end of Ronald Reagan's presidency, and *jumhu:ri:ya:t mawz*, a literal translation of 'banana republics'.
16. These expressions were all noted as in fairly common use from at least the mid-1980s onward in newspapers such as *al-Ḥayā* (Lebanon), *al-Sharq Al-'Awsaṭ* (Saudi Arabia and London), magazines such as *al-'Arabī* (Kuwait), and even in respectable modern prose literature.
17. Ibrahim 1988, 211.
18. Increasing literacy has also accelerated this kind of change in the dialects: borrowings such as *hafi:z* 'office' and *sbe:tar* 'hospital', which came into the Gulf dialects with the advent of the oil industry, have now all but been replaced in the speech of the literate under-forties by their pan-Arab MSA equivalents *maktab* and *mustašfa:*.
19. On 30 September 1993, the writer saw an advertisement for Yardley's "Lace" (a women's perfume) on Middle East Broadcasting, an Arab satellite television station, in which the following text was spoken and appeared on the screen: *la:šay yuṣbiḥu lmarʔa miθil "Lace"* = 'Nothing becomes a woman like "Lace"'. The verb *yuṣbiḥu*, however, means 'becomes' in the sense of "turns into," not in the sense of "beautifies." This is a not untypical example of the kind of (mis)translation which has become widespread in advertizing.
20. Rosenhouse 1990 presents some interesting diachronic statistics on the incidence of various types of nominal structures in written Arabic, as revealed by an analysis of two sets of prose samples by a dozen writers, half of them writing in the eighth/ninth centuries and the other half in the nineteenth/twentieth centuries. The modern sample is mainly literary but includes some journalistic material. The study shows that (i) noun phrases are more frequent in the modern sample (and finite verbs less), (ii) the noun phrases in the modern material are structurally more complex (greater incidence of annexion for example), and (iii) that the clauses that contain them are longer. These findings apply even to the specifically "literary" modern material.
21. Al-Nafīsī 1978, 26.
22. Ibid., 17.
23. Wehr 1953, 31–34.
24. Somekh 1991, 15.
25. Ibid., 31–33.
26. Al-Ḥuṣarī 1985, 86–90.
27. Somekh 1991 provides a concise, informative overview of the field.
28. Beeston 1983, 180 quoting Eissfeld 1965, 57.

LEXICAL AND STYLISTIC DEVELOPMENTS

29. 'Imāra (ed.) 1972, 7.

30. Johnstone 1991, 117.

31. Beeston 1983, 181–82 notes examples of the use of paraphrastic parallelism in English prose of the Tudor age that have a very similar structure to that supposedly typical of Arabic.

32. See Serjeant 1983, 119 for a famous late-seventh century example of the *xuṭba* genre in which al-Ḥajjāj, the governor of Iraq, excoriates the inhabitants of Kūfa for their dissension, hypocrisy, and low morals.

33. Somekh 1991, 77–78.

34. See, for example, the extract from his speech of 9 March 1958 on "the Arab Revolution," reproduced, with a translation, in Monteil 1960, 314–15. The opening lines give a flavor of how Nasser used repetition and parallelism:

> *haːðihi ʔayyuha: lʔixwa, hiya θθawratu lʕarabiːya lḥaqiːqiːya, ʔaθθawratu lʕarabiːya, ʔallati: tanbudu min dimaːʔikum, wa tanbudu min mašaːʕirikum wa tanbudu min quluːbikum . . .*
> 'This, oh my brothers, is the Arab Revolution, the Arab Revolution which pulsates from your blood and pulsates from your feelings and pulsates from your hearts'

The three conjoined relative clauses make use of the same syntactic structure with the same verb repeated three times, although its prepositional complement is permutated: 'from your blood/feelings/heart', standing for the physical, intellectual, and emotional aspects of the audience's hoped-for commitment to the cause.

35. Beeston 1983, 184–85 gives an excellent example of the functional differentiation of the "western expository" and the "ornate" styles in a single piece of writing. In an article in a journal for educated laymen (*al-ʿArabī*, roughly equivalent in its coverage to the *Readers' Digest*) on the subject of gray hair in poetry, a university lecturer uses a straightforward expository style in describing the physiological causes of gray hair but then switches to "ornate" style in dealing, a couple of paragraphs later, with a more emotionally charged aspect of his subject, the psychological impact of the appearance of gray hair.

36. For example, in the title essay of his 1966 collection called *al-Nūr wa l-Dayjūr* (lit. 'light and darkness', itself a rhyming lexical couplet in which a rare word *dayjuːr* 'darkness' has apparently been chosen over the commoner synonym *ðalaːm* because of its rhyme with *nuːr*), Nuʿayma makes heavy use of parallelism and repetition to make the point that modern societies are much more concerned with preventing their systems of material values being corrupted than to protect their moral, spiritual, and aesthetic ones. We are not living in an age of 'light' (*nuːr*), he claims, but rather in an age of 'darkness' (*dayjuːr*).

37. The veteran political columnist Muṣṭafā Amīn frequently uses this style to make veiled critical comment. See for example Jubouri's analysis (Jubouri 1984, 99–117) of a 1982 text of Amīn's dealing with the virtues and vices of modern political parties.

38. Diyab 1991.

9

Language Level
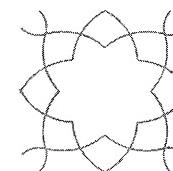
9.1 Introduction

In this concluding chapter, we address the complex question of language level in Arabic: how far is it possible to distinguish discrete levels of Arabic within individual Arab speech communities? If formally discrete levels can be shown to exist, how many are there, and in what circumstances do native speakers and writers make use of them? The answer to this second question should give us some clue as to the symbolic and pragmatic values that different kinds of Arabic have for the native user.

The obvious point of departure for the discussion of language level in Arabic is Charles Ferguson's well-known article on diglossia in which he deals with relationships between form and function in Arabic, Swiss German, Haitian Creole, and modern Greek.[1] Ferguson distinguishes two linguistic varieties within each language that he refers to generically as "H" ("high") and "L" ("low"). These varieties, he claimed, differ from each other formally and functionally in similar ways in all four languages. In the case of Arabic, where H signifies CLA/MSA and L all Arabic dialects as a group, Ferguson distinguishes the two varieties from each other formally in the following general ways:

(a) Phonologically: L is the basic system. The divergent features of H are a parasystem, which are often replaced by their L equivalents in the oral performance of H.

(b) Morphophonologically: the L system is simpler and more regular than the H system (e.g., verb paradigms are more symmetrical, and there is less allomorphy in general).

(c) Syntactically: there is a total loss of case and mood inflections in L, with retention in H; there are fewer categories of person and number in L; concord is stricter[2] in L than in H.

(d) Lexically: there are many paired lexical items, one H, one L, for everyday concepts, in which the range of meaning is roughly the same, but the use of one or other member of the contrasting pair immediately stamps the utterance or written sequence as H or L.

In discussing the stability of H and L in Arabic, Ferguson notes that there is also

> a kind of spoken Arabic much used in semiformal or cross-dialectal situations [which] has a highly classical vocabulary with few or no inflectional endings, with certain features of classical syntax, but with a fundamentally colloquial base in morphology and syntax, and a generous admixture of colloquial vocabulary . . . .[3]

Ferguson gives a sample list of situations, reproduced in table 9.1, in which H or L is typically required.

He adds the comment:

> The importance of using the right variety in the right situation can hardly be overestimated. An outsider who learns to speak fluent accurate L and then uses it in a formal speech is an object of ridicule. A member of the speech community who uses H in a purely conversational situation or in an informal activity like shopping is equally an object of ridicule.[4]

Much of the sociolinguistic work done in Arabic in the forty-odd years since the publication of Ferguson's article has attempted to extend, refine, or refute this outline model of the sociolinguistic structure of Arabic-speaking societies. Some researchers have examined the accuracy of Ferguson's admittedly sketchy characterization of the formal aspects of H and L by analyzing data collected in various (usually speech) contexts in order to discover whether, even in what he claimed were the paradigm cases, the two varieties are in practice kept discrete.[5] Others have turned their attention precisely to the structure of the mixed kinds of Arabic produced in the "semiformal" and "cross-dialectal" situations that were excluded from his model.[6] The enormous economic and demographic

Table 9.1   Diglossia after Fergson: Form and Function

Contexts	H	L
Sermon in church or mosque	x	
Instructions to servants, waiters, workmen, clerks		x
Personal letter	x	
Speech in parliament, political speech	x	
University lecture	x	
Conversation with family, friends, colleagues		x
News broadcast	x	
Radio "soap opera"		x
Newspaper editorial, news story, caption on picture	x	
Caption on political cartoon		x
Poetry	x	
Folk literature		x

changes that have occurred in the Arab world in the last two to three decades have made the kind of social contact that gives rise to hybrid varieties of Arabic much more widespread than was the case at the time Ferguson wrote his article. Although still accepting the notion of discreteness of levels, if not always agreeing on how many levels there are, such researchers have looked for a more delicate descriptive model than Ferguson's simple dichotomy, in the form of continua or scales capable of handling MSA/dialect morphological hybrids and co-occurrent syntactic phenomena.[7]

Turning now to the functional side of Ferguson's model, we note that recent research has cast doubt on the epistemological status of the concepts on which it is based. Quite apart from its accuracy in the paradigm cases—has it ever been true, for example, that only H is used in political speeches and university lectures, and only L in conversation with friends and colleagues?—the model is that of an observing linguist, and the definitions of H and L on which it depends are derived from other linguists' grammatical descriptions, which are always to some degree *idealizations* of (or in the case of H often *prescriptions* for) behavior. It is legitimate to ask how far these a priori assumptions about H and L are shared by linguistically naive native speakers: does everyone agree on which utterances constitute examples of H and which examples of L? The answer to this question seems to be "no" insofar as the definition of H is concerned: recent research suggests that nonlinguist native speakers readily accept as H a wider range of forms than do most linguists, Arab or western.[8] This preliminary observation, if confirmed more widely, will have obvious implications for any work that aims at making sense of stylistic variation in Arabic speech, because

it suggests that native speakers may be less sensitive than linguists assume to some of the formal variation that linguists observe or at least less sensitive to some as yet unspecified aspects of it in some types of situation. What exactly the clinching features are that make a disinterested naive native observer (as opposed to a interested sophisticated linguist) classify a stretch of discourse as H, L, or somewhere between the two is not yet known in any detail. What *is* clear, however, is that the problems of description and explanation are inextricably bound: in any passage of Arabic speech, whether monologue or conversation, one cannot track, still less make sense of, the moment-by-moment, unpredictable changes in language form unless one is also aware of co-occurrent changes in the ideational content of the discourse and the interpersonal relationships of the participants, as perceived by the participants themselves. Changes in the form of what is said are a complex set of signals—the rules of which have yet to be worked out—of these underlying changes.

Before considering in more detail what we mean by a "change in form" (9.4), we will look in some detail at a model that attempts to relate form and function in a more delicate way than Ferguson (9.2) and then at the functional motivations that seem to lie behind changes in form (9.3).

9.2 Language levels in Cairo

How do ordinary Arabs conceptualize large-scale differences in language level? The only terms that they commonly use are *al-fuṣḥa:* 'the pure/ eloquent' (= Ferguson's H), the general term for the learned language of literary and formal writing, and *al-'a:mmi:ya* 'the vernacular' (= Ferguson's L, lit. 'that which pertains to the general (public)'), which is what they learn at their mother's knee and continue to speak for the rest of their lives. Throughout this book we have pointed out the wide differences that do indeed exist between the two at every level of formal linguistic structure. As we have just suggested, however, when it comes to questions of use, the layperson's perception of where the dividing line comes between *fuṣḥa:* and *'a:mmi:ya* does not always correspond with the black-and-white prescriptions of the grammar book. Within an Arabic speech community, there are all sorts of intermediate shades of gray between the plainest dialect and the most formally correct *fuṣḥa:*. The "semiformal" variety mentioned by Ferguson, with its basically L syntax and generous admixture of H vocabulary and phraseology, is only one such. It is also possible to distinguish a kind of substandard H whose syntax, morphology, and lexicon is H but whose phonology is L-influenced and omits most H inflections. Nonetheless, native speakers still think of this as H.[9] Conversely, there is also a species of somewhat

formalized L, a kind of "posh dialect," not quite up to Ferguson's "semiformal" level, in which some of the more saliently L phonological and lexical features are variably replaced by their H equivalents. There are speech contexts where speakers will stick to one of these levels and others where they switch back and forth. Such switches, often signaled paralinguistically by changes in speed of delivery or voice quality, occur at points where there is a change in the topic, the interlocutor, or the purpose of talk—points to which we will return in more detail later. The descriptive difficulty is that the language levels that constitute much of this speech continuum (however one defines "level" and however many levels one proposes) are probabilistic, not absolute: in other words, the levels are constructs, produced by the patterns of simultaneous choices that speakers in a community make, in a consistent and more or less predictable way, on many linguistic variables. But there are few variables where one of the variants that constitute it always occurs in one level and never in the adjacent one: the way most work is that the likelihood of one variant rather than its rival occurring gradually increases in a given range of contexts in more or less the same proportions for all speakers. Levels can thus be identified as discrete only at points where a number of variables have all changed from being typically realized as variant (a) (say, the "formal" variant) to being typically realized as the rival ("informal") variant (b). Not all variables are similarly calibrated to the demands of changing formality/informality of context, however, so there are fuzzy areas between one discrete level and another. Badawi, in his five-level model outlined below, recognizes this and uses the metaphor of the rainbow in which there are clear bright areas of color but other transitional areas where it is impossible to decide where one color ends and another begins. So it is with levels: apart from the changes in paralinguistic performance already alluded to, the formal evidence that a level change has occurred is a large number of simultaneous co-occurrent switches from one variant to another across a number of variables. The analyst's definition of "large number," however, is subjective and switches on some variables may, from the user's point of view, be more salient and significant than switches on others.

Badawi's study of level in Cairene Arabic[10] provides a useful picture both of how variation is structurally organized (which features are more or less used in which levels) and what the social meaning of variation is (which levels are typical of which kinds of speaker and which kinds of speech context) within one major Arab speech community. One limitation is that the data for the study was gathered mainly from the Egyptian media, whose output tends to underrepresent some parts of the continuum—most obviously the "plain dialect" style of speech that hardly occurs at all—and to overrepresent others at the formal

and semiformal end of the scale. Nevertheless, Badawi's study can be accepted as having general validity outside Egypt because, although Arabic dialects differ from one region to another as well as within one region, the status relationships between *fusḥa:* and local dialects are broadly similar everywhere.[11] Badawi distinguishes five language levels:

(1) *fusḥa: al-tura:θ* 'heritage *fusḥa:*'
(2) *fusḥa: al-'aṣr* '*fusḥa:* of the age (we live in)'
(3) *'a:mmi:yat al-muθaqqafi:n* 'dialect of the (well-) educated'
(4) *'a:mmi:yat al-mutanawwiri:n* 'dialect of the literate'[12]
(5) *'a:mmi:yat al-ʔummi:yi:n* 'dialect of the illiterate'

Levels 1 and 2 correspond to Ferguson's H and Levels 4 and 5 to his L, with Level 3 representing a bridge between them and equating to his "semiformal" level. Badawi's terminology points to a fault line in the continuum between Levels 2 and 3: whereas Level 2 is still *fusḥa:*, Level 3 is *'a:mmi:ya.* His explanation is that although Level 2 may show dialectal phonological influences, its morphosyntactic base remains grammar-book *fusḥa:*. Level 3, however, although it may show quite heavy use of *fusḥa:* vocabulary and phraseology and concomitant phonological and morphological influences (see 2.3.1), its syntactic systems—in particular word order, expression of mood and aspect, negation, and concord—remain nonstandard.[13] Levels 1 and 2 are, then, *qualitatively* different from Levels 3–5, which are differentiated *quantitatively* from each other, that is, in the incidence of occurrence of given *'a:mmi:ya* features.

Badawi's terminology also suggests that his model is not one dimensional: whereas the *'a:mmi:ya* levels are defined in terms of the kinds of people who typically use them, the differences between the two *fusḥa:* levels relate to subject matter, "heritage" versus "the contemporary age." Thus the model reflects both style differences demanded by subject matter (i.e., register related) and differences corresponding to categories of social class similar to those employed in western socolinguistic research. In fact, as Badawi makes clear, speakers do not statically occupy one of his levels according to their level of education, at least, not the well-educated ones.[14] Depending on what is being talked about and with whom, well-educated Cairenes may use any of Levels 3, 4, or 5 in spontaneous conversation, although Level 5 only perhaps with illiterate interlocutors with whom they wish to establish a rapport of some sort. The reason for labeling Levels 3–5 in the way Badawi does is that these are the language levels that people with given levels of education (= differing levels of exposure to, and familiarity with, MSA) typically use most of the time. Thus anyone hearing the

Table 9.2   A Hallidayan Model of Language Level in Arabic

Level	Field(s)	Tenor	Mode(s)
1	Islam/Classical Arabic literature/cultural history	ritualized/highly formal	written/prepared spoken monologue/reading aloud
2	Nonreligious serious (e.g., chemistry, law, politics, news bulletin)	formal	written/reading aloud
3	Same as Level 2	semiformal	extempore speech (monologue or conversation)
4	Nonserious (e.g., discussion of television programs, sport, job, fashion)	informal	conversation
5	Nonserious, domestic, uninformed by contact with modern civilization	informal	conversation (esp. with children/ illiterates)

kind of Arabic used in, say, Level 3, would in the absence of any visual or other clues to the contrary conclude that the speaker had a good level of education, and anybody hearing Level 5, that he or she had little or none. Most Cairenes not employed as television newscasters and the like would rarely have cause to regularly attempt to *produce* Level 2 at all in the spoken mode, but they are of course constantly exposed to it through the media.

The multiple factors that operate to produce Badawi's continuum can perhaps be more clearly illustrated (table 9.2.) in terms of the Hallidayan system of field (subject matter), tenor (relationship of participants), and mode (speech versus writing) with some stereotypical examples.

These levels of language are typically represented on television and radio in the following kinds of program:

Level 1:   Recitation of the Koran; dramatic recreation of events in Islamic history
Level 2:   Political speech to the nation, read from a prepared text; news bulletin; voice-over commentary on serious documentary
Level 3:   Studio discussion on any serious topic, for example, literature, the environment; unprepared interview with government minister, scientist, writer
Level 4:   "Vox pop" interviews in the street with ordinary people; "ordinary people" depicted in television/radio plays, serials, soap operas; dis-

cussions, interviews on nonserious topics, especially if involving women (e.g., cooking, fashion); game shows; sports commentary

Level 5: Rarely represented except by speech of stereotypical working-class characters (doormen, porters, messengers, cleaners) in comedies and soap operas.

A linguistic model of the type proposed by Badawi, like the television programs used above to exemplify its levels, is not an exact replica of real life. It describes what usually happens in given sets of circumstances, what native speakers have come to expect will happen. But there are frequently occasions when contextual factors may lead a speaker to switch, sometimes quite consciously, from one level to another. We will now examine some examples of this.

9.3 Level switching

9.3.1 Monologues

Gamal Abdul Nasser (president of Egypt 1954–70) was particularly adept at exploiting the connotative and social meanings of different language levels in order to achieve his rhetorical and, indirectly, his political purposes.[15] More than thirty years since his death, and long after his Arab socialist policies have been abandoned, Nasser is still remembered throughout the Arab world for the brilliance of his rhetoric. He did not achieve this fame by sticking, as Ferguson's model suggests and the Arabs' own mythology about their language might lead one to assume, to the pure CLA of Level 1,[16] nor, as our data-based model derived from "normal" behavior might indicate, to Levels 2 or 3. In his public speeches, broadcast live in Egypt and throughout the Arab world, Nasser regularly used all levels of language from 1 to 5. Although it is now commonplace to hear political leaders such as Muammar al-Gaddafi and, until his fall from power, Saddam Hussein, who superficially espouse similar political principles to Nasser's, speak dialect in public, such behavior went against the grain of the traditions of formal public speaking that lasted until as late as the mid 1950s. It was Nasser who broke this oratorical mold.[17]

The use of parallelism and repetition, as we saw in 8.4, is a traditional and culturally valued stylistic feature of "good" Arabic. Nasser made skillful use of this, albeit in a form simple enough for ordinary (that is, at the time he was in power, mainly illiterate) people to understand him. More important, Nasser constantly and deliberately played on the powerful subconscious connotations that are carried by the different ways of "saying the same thing" for his audience. In many of his speeches, the text is delivered, for the most part, in (in Badawi's

terms) a mixture of a heavily Cairene-accented Level 2 and Level 3. This was the norm for Nasser. But where certain ideational and interpersonal elements were being emphasized, particularly on occasions of acute national crisis, a marked and often sustained switch up or down could occur. During the Suez Crisis of 1956, Nasser gave two speeches at mass rallies a week apart, the first on Friday, 2 November, three days before the Anglo-French invasion force landed, and the second on 9 November, two days after the ceasefire. The first, preinvasion, speech is a national call to arms, expressed in particularly concrete and personal language. It is heavily dialectalized. Nasser describes graphically how "we" will fight "from house to house and village to village," going on to draw a parallel with Egypt's experience in World War I and the 1948 war with Israel in which he served as a young army officer and recounting his personal experiences in a fierce tank battle (*ma: kuttiš ṭabʿan ʔaːʿid fi lxandaʔ, kutt bistimraːr barra* 'I didn't just sit in the trenches, I was out on the battlefield the whole time'). He ends with a fist-shaking personal pledge: "I" will stay with "you," the audience, here in Cairo to fight alongside "you," and nor will "I" allow "my children" to be evacuated from Cairo to avoid the coming bombs. Nasser departs in both form and content from the official text of this speech at this point.[18] The written text simply has the short statement:

wa ʔanaː fiː lqaːhira saʔuqaːtilu maʿakum li ʔaːxiri nuqtati dam
'While I am in Cairo I will fight alongside you to the last drop of blood'

But in the text as performed, Nasser addresses his audience in language that is pure Cairene colloquial in form and delivery, such as might be heard in the alleyways of the poorer districts of the city, and apparently adds the pledge to keep his children in Cairo in the heat of the moment:

awlaːdi mawguːdiːn maʿaːkum fi lqaːhira . . . maː ṭallaʿthumš barra miš ḥaṭallaʿhum barra, wa ana mawguːd maʿaːku hina fi lqaːhira . . . ḥanuqaːtil, zayyima ʔult ilkum imbaːriḥ, li ʔaːxir nuʔṭit dam
'My children are here with you in Cairo . . . I haven't sent them out [of Cairo], nor will I send them out while I'm with you here in Cairo . . . we will fight, as I told you yesterday, to the last drop of blood'

Nasser's phonological, syntactic, and lexical choices make this, in terms of Badawi's model, Level 4/5. Obviously, the same defiant extemporized message *could* have been expressed in the language of Level 2, the "norm" for public speaking, in which case it would have been something like:

Pawla:di: mawgu:du:n ma'akum fi: lqa:hira . . . lam Pursilhum Pila: lxa:rij, wa lan Pursi-
lahum wa Pana: ma'akum huna: fi lqa:hira . . . sanuqa:tilu, kama: qultu lakum Pams, li
Pa:xiri nuqtati dam

The message would have been the same, but not the communicative effect, as Nasser seems to have recognized. In this speech context, the linguistic "correctness" of the Level 2 version sounds pedantic and artificial; it was quite simply an inappropriate vehicle to carry the emotional charge of the message and the personal commitment of the speaker to his audience at a time of acute crisis and physical threat to them.

The second, postceasefire speech, when Egypt was suing for peace through the United Nations, is completely different. The message here is an abstract, moral–philosophical one: that there is a difference between peace, *sala:m*, and surrender, *Pistisla:m* (the words share the same root consonants). A third-person "Egypt" is presented to the audience as a quasimetaphysical essence "defending" its freedom and independence, "calling" for peace, and "knowing" the difference between the two. Its people, whom Nasser is facing as he speaks, are not directly addressed as "you," nor in Nasser's accustomed phrase *Pixwa:ni:* 'my brothers', but as the stylized *Pabna:P misr* 'the sons of Egypt' or the anonymous third person *muwa:tinu:n* 'citizens', which coalesce to form this ethereal essence. Throughout, this appeal to the audience's intellect and sense of national identity is couched in the purest grammar-book *fusha:*: case and mood endings are scrupulously correct, pausal and junctural phenomena are observed, and there are syntactic markers that, in extemporaneously spoken MSA, are strongly associated with the "heritage" or religious *fusha:* of Badawi's Level 1. The delivery is rapid but completely nonconversational, the speech being broken up into roughly chronometrically equal tone groups, each of which builds up to a crescendo on a stressed high final nucleus, which occurs on a long vowel, and then falls, tailing off to a pause before the next tone group. The whole is delivered with a breathless and frenetic rhythm, and it seems at times almost as if Nasser is on the verge of collapse.

Calling a spade a spade (or sometimes a bloody shovel), recounting personal experiences, identifying the audience's hopes and fears—all of these are among the rhetorical tactics that any modern political speaker needs on occasion to resort to if he or she is to communicate effectively. Although it is of course possible to do any of these things in any level of Arabic, all fall naturally within the functional domain of the vernacular, a variety of the language whose use is apparently incompatible with the traditions of public oratory: in fact, Ferguson specifically picks out "political speech" as one of the paradigm cases in which H is supposedly required.[19] However, fervent appeals to national sovereignty

and dignity, and abstract disquisitions on the meaning of peace, which Arab politicians, like others, are expected to make, cannot easily be talked about in colloquial Arabic without recourse to a literary vocabulary that many of the audience would not understand. Nor would it necessarily be felt appropriate, even by an uneducated listener, if they were talked about in this kind of language. The populist public speaker is thus faced by a dilemma: the need to establish and maintain, on occasion, an intimate rapport with his audience requires one kind of Arabic, whereas the "message" sometimes requires quite another. Nasser recognized this sociolinguistic tension in Arabic speech communities and exploited it throughout his career, switching language levels without compunction or regard for the traditions and conventions of public speaking in order to find the most appropriate and effective form for the varying interpersonal role relationships he wished to play out with his audience (pan-Arab leader to all Arabs; Egyptian president to Egyptian citizen; fellow Egyptian to fellow Egyptian; even Egyptian peasant to Egyptian peasant) and the different "packaging" in which he wanted to wrap his message (appeal to abstract ideals, concrete description of problems, testament of personal experience, ridiculing of opponents, etc.).

Sometimes, the same point is made in successive sentences at different language levels or a point is made in Level 2 (the elevated level signaling that it is to be taken as the authoritative "text"), and Level 4 or 5 is used immediately after to parenthetically explain or comment on it before the "text" resumes once more, rather in the manner of a schoolteacher explaining a lesson or an imam a piece of scripture.[20] For example, after the axiomatic assertion that:

bidu:n iddimu:qra:ti:ya la: yumkin li lhurri:ya Pan tatahaqqaq
'Without democracy it is impossible to achieve freedom'

in Level 2, Nasser comments in Level 4, co-opting his audience's opinions:

kullina 'andina tagruba wadha
'We've all got clear experience (of this)'

A similar exegetical ploy, aimed at involving the audience in the development of the argument, is the insertion of rhetorical questions in colloquial Arabic at points where they might not understand the "text," acting out in speech, as it were, the audience's mental reaction. A good example of this kind of switching is in a 1965 speech on the meaning of socialism. Having defined socialism in Level 2 with almost pontifical gravitas as:

ʔattargama . . . ʔaṣṣaḥi:ḥa li kawni θθawra ʿamalan taqaddumi:yan
'The correct . . . interpretation of the revolution in terms of progressive action'

Nasser plays the role of the puzzled listener, replying with a question in staccato Cairene (Level 4/5):

lištiraki:ya yaʿni e:h
'But what's "socialism" mean then, exactly?'

Back comes the reply in Level 2, again delivered in slow, measured tones:

lištira:ki:ya ka kalima maʿna:ha: ʔiqa:mit mugtamaʿ ilkifa:ya wa lʿadl . . . ʔiqa:mit mug-
tamaʿ taka:fuʔ ilfuraṣ . . . ʔiqa:mit mugtamaʿ . . . alʔinta:g wa ʔiqa:mit mugtamaʿ . . .
alxadama:t
'"Socialism," in a word, means the establishment of a society of sufficiency and justice . . . the establishment of a society of equal opportunity . . . the establishment of a society . . . of production and . . . of services'

But Nasser's imaginary listener is still uncomprehending, replying through him in a conversational Level 4/5:

lištiraki:ya yaʿni e:h . . . bnʔu:l ištiraki:ya kilma waḥda bass!
'But what's "socialism" *mean*? . . . We say "socialism," but it's just one (little) word!'

A word or two might perhaps be added here about the use of standard and nonstandard Arabic in another form of solitary performance: song. This is a vast subject, and we include reference to it here merely to illustrate how the connotations of standard and dialectal Arabic, which Nasser exploited so effectively, can also be mobilized in a rather different kind of public performance. In 1972, the Lebanese singer Fayruz issued a long-playing record of songs about Jerusalem and the Palestinians titled *Al-Quds fī l-Bāl* ('Jerusalem on my Mind'). The relationship between the language of these songs and the type of message they convey is instructive. We will look briefly at two. The opening song of this collection, *Zahrat al-Madāʾin* ('Flower of Cities'), which achieved instant fame throughout the Middle East, has become almost a second Palestinian national anthem. The song begins with a verse in which the singer addresses the city directly, as if it were a (female) person:

li ʔajliki ya: madi:natu ṣṣala: ʔuṣalli:
li ʔajliki ya: bahi:yatu lmasa:kin, ya: zahratu lmada:ʔin
ya: qudsu, ya: qudsu, ya: qudsu, ya: madi:natu ṣṣala: ʔuṣalli:

'For your sake, oh city of prayer, I pray
For your sake, oh most splendid of dwelling places, oh flower of cities
Oh Jerusalem, oh Jerusalem, oh Jerusalem, oh city of prayer, I pray.'

Linguistically, this is Level 1 CLA. The song continues in the same high-flown style, with Fayruz addressing the city in the role of a Palestinian exiled from it, by turns conjuring up the colonnades of its mosques and its ancient churches, alluding to religious figures and events associated with Jerusalem ('the child in the grotto' = Jesus, 'the path of he who ascended to the heavens' = Muḥammad), and recounting more recent tragic events ('Justice was defeated before the guns' (sc. of the Israelis in 1967)). The general tone is one of yearning and grief, but the song ends with a call to arms and a defiant prediction that the waters of the River Jordan will wash away all traces of the "barbarians' feet" and put to rights a historic injustice. A full orchestra and choir provide a grandiloquent musical accompaniment.

In stark contrast, the song that immediately follows it, titled "Old Jerusalem," opens in plain Levantine Arabic:

imši:t fi ššawa:riʿ, šawa:riʿ ilʔids ilʿati:ʔa
'I walked through the streets, the streets of old Jerusalem'

The Jerusalem of this song is not the ethereal, Classical Arabic *quds* of "Flower of Cities," the city of millennia of Christian and Muslim history, for whose salvation the singer is "praying" to a Christian or Muslim God, but rather a dialectal *ʔids* of narrow streets and small shops, through which Fayruz has just "walked" in a kind of reverie—a sentimentalized Jerusalem redolent of intimate personal and familial memories, of houses and land passed from one generation to another "under the sun, under the wind." The music is (in a nontechnical sense) in a minor key, the falling cadences underlining the personal reminiscence.

Thus, in songs that deal with what Jerusalem means to Palestinians, just as in speeches that deal with a national political crisis, the choice of medium—standard Arabic on the one hand, dialectal Arabic on the other—is a major element of the message. Either medium can be chosen as a vehicle of expression, but in each case a different set of associations, cultural resonances, and interpersonal relationships between the speaker/singer and the listener is conveyed.

9.3.2 Conversation

The focus of research interest has been both on conversation within a single speech community, where all participants share the same or very similar dialects, and on interdialectal conversation where they do not. We will look at each situa-

tion in turn. Many of the points made about the use of language level in monologue also apply, *mutatis mutandis*, here.

In a relaxed conversation, between, let us say, a group of educated Bahraini acquaintances,[21] the dialectal base will be "Bahraini" in that all the participants, if they have been born and brought up in Bahrain, will share a number of local features or particular combinations of them that are particular to Bahrain. So, for example, the reflex of MSA /q/ will be /g/ in many words, the interdentals /θ/ and /ð/ will be preserved, the second and third person p-stem verbs will have an -/u:n/ ending, a *b*- prefix will be prefixed to p-stem verbs to express future intention, and the word most likely to be used for existential 'there' (in equivalents of 'there is') will be *hast*.[22] This small set of phonological, morphological, and lexical features would already be sufficient to distinguish them as a group from Kuwait and Iraq to the north, where 'there is' is usually /aku/ not /hast/; from Muscat to the southeast where /q/ is preserved, and the p-stem verbs have -/u/ (m.) and -/an/ (f.) rather than a common plural -/u:n/; and, moving further afield, to distinguish them totally from Cairo and Damascus, where all five features would be different.

As this conversation proceeds, what factors will determine the level of language the participants use, and why and at what points will the level rise and fall? These are complex questions. The answers require a careful analysis of the effect of changes in what is being talked about and to what end, what personal relationships exist between the participants, and psychosocial imponderables impossible to recover from simple observation or even from direct questioning of the participants by the observer, such as the degree to which participants, who it will be remembered are educated, continue to identify with the values (and speech patterns) of the sectarian communities in which they were born and brought up.[23] That being said, it is unlikely that any group of Bahrainis talking relaxedly among themselves, or indeed any group from any other single speech community, would deviate markedly from the local linguistic common denominator, that is, the dialectal features that they all share. This means that the phonology, morphology, and sentence syntax would be dialectal virtually whatever they were talking about; choice of vocabulary, however, which depends much more directly on topic, would be more variable.

If our educated Bahrainis are talking about a "serious" topic—Middle Eastern politics, let us say—much of the vocabulary and phraseology they use will be nondialectal, referring to concepts, entities, and processes that belong to the wider world of pan-Arab and international concerns, for which the natural vehicle is MSA. Much of this standard vocabulary, as we saw in chapter 8, comes in prepackaged noun phrases ('peace process', 'shuttle diplomacy', 'new world

order', 'safe havens', 'ethnic cleansing', etc.), a kind of shorthand allowing economical reference to be made to complex processes and ideas and establishing a shared conceptual framework for the conversation. As in English, such phrases are linguistically frozen elements—their pronunciation reflects the norms of MSA rather than those of the dialect of the speaker, and their phrase-internal syntax is invariable. The sentences in which they are accommodated, however, will be dialectal from the point of view of their syntax and morphology. Furthermore, the core vocabulary items that describe nontechnical entities, operations, and relationships—for example, verbs such as 'see', 'come', 'go', 'bring', 'do, make'—will also tend to be dialectal, although in educated speech there is always some degree of "spillover" from MSA linguistic elements into contiguous dialectal words and structures, resulting in hybrid forms of various kinds (see 9.4 below). We are, in other words, talking here about a Bahraini incarnation of Badawi's Level 3, "the colloquial of the educated." But if conversation shifts to a nonserious topic of merely local interest, say last Friday's match between two leading Bahraini soccer teams, the level will shift sharply downward toward plain Bahraini colloquial (Badawi's Level 4/5) with little or no MSA influence on vocabulary or structure. This is partly because the intrinsically nonserious nature of the subject does not justify the effort of speaking in a more "formal" manner, but partly also because, although MSA phrases have been invented for 'corner kick' (*ḍarba rukni:ya*), 'shoot' (*rama:*), 'penalty' (*ḍarbat jaza:ʔ*), etc., and are used in written soccer reports, most soccer-lovers feel more at home with the dialectal forms they learned at an early age when watching or playing the game—respectively *kurnar, šawwaṭ*, and *balanti* in the Bahraini case.

We have so far been describing what occurs in a *relaxed* conversation between educational *equals* on serious and nonserious topics; what happens where the conversational context *formalizes* things, or where there is a *gap* between the educational attainments of the participants? Interesting cases in point are two interviews conducted on Bahrain Radio in the arts program "'*Alā Ṭarīqi l-Fann*" in 1978.[24] In the first, the program presenter interviews Jāsim Khalaf, a well-known Bahraini comic actor, about his stage career, and in the second, the same presenter interviews Mbārak al-Rāʾī, an old man who was famous in his day as a *naḥḥa:m*, a singer of sea songs employed on pearling dhows to encourage the divers. Although the subject matter of both interviews—verbal art in dialect—is quintessentially local in character and all the participants in both interviews are Bahrainis, the fact of being interviewed on radio could be expected to exercise a formalizing effect on the speech of all participants; furthermore, in parts of the interview, the interviewees are asked not merely to reminisce, but to *conceptualize* and *explain* to the listener what makes a good comic actor and a vir-

tuoso singer of sea songs. So, in theory, although all involved speak similar dialects, both the speech context and the treatment of the topic might be expected to push the language level in the direction of MSA (i.e., Badawi's Levels 2/3), compared with what would be normal in a relaxed chat. However, there is a great disparity in the educational level of the interviewees: Khalaf earns his living as a middle-ranking civil servant, whereas al-Rāʿī spent his whole working life at sea and is almost certainly illiterate. A further difference that would be obvious to a Bahraini audience is that Khalaf and al-Rāʿī speak different local dialects: Khalaf speaks the *Baḥārna*, that is, the local Shiʿi, tacitly stigmatized dialect (which he is famous for accentuating on the stage for comic effect), whereas al-Rāʿī, like the interviewer himself, speaks the ʿ*Arab*, that is, local Sunni, nonstigmatized dialect.

The Khalaf interview is particularly interesting because of the contrasting way the interviewer and interviewee handle the problem of language level. Khalaf speaks without any hesitancy in a instantly recognizable *Baḥārna* dialect (equating roughly to Badawi's Level 4), with only sporadic phonological adjustments away from stigmatized features, such as the occasional replacement of /f/ by /θ/ in words such as *fala:fa* 'three' where both MSA and the interviewer's dialect have /θ/. Furthermore, in response to the interviewer's address form *usta:ð* 'teacher', a respectful reference to his mastery of acting, Khalaf uses the markedly familiar and local *yaxu:k* 'brother', an address form reserved in Bahrain society for age peers or those felt by the speaker to be of equal status to himself. The interviewer's performance sounds hesitant and rather uncomfortable throughout, with many false starts and rephrasings as he hovers between using the formal MSA of the questions on his script and a dialectal Arabic equivalent in informality to the interviewee's.

The significance of the hesitation becomes apparent when it is contrasted with the same interviewer's completely fluent, dialectal delivery in the interview with al-Rāʿī, the illiterate retired singer of sea shanties. In this case, the contextual factors—the illiteracy of the interviewee, the shared sectarian (= dialectal) background of the speakers, and the purely local and traditional nature of the interviewee's art—all exerted linguistic pressure in the same direction and were more than sufficient to override the formalizing effect of a radio interview, transforming it into the Bahraini equivalent of a fireside chat between friends. But in the Khalaf interview the pressures are conflicting. Khalaf is clearly a literate and educated man and a serious artist performing in a modern art form, albeit one that uses dialect. Although this perceived status requires the interviewer, in a public context such as a radio interview, to put "serious" questions to him in at least an approximation of *fuṣḥa:*, Khalaf's insistence on treating

the whole business in a relaxed and unstuffy manner—*vide* his unhesitating use of dialect and familiar address forms—faces the interviewer with a dilemma. On the one hand, if he persists with his preprepared *fuṣḥa:* questions in the face of the relaxed dialect of Khalaf, he risks sounding pompous; on the other hand, if he rephrases them into his own (ʿ*Arab*) dialect, this would sound both odd and, in the circumstances, patronizing: odd because of the incompatibility of the dialectal form with the intellectual content of the questions, and patronizing because it might seem as if he were implying by the use of dialect that the interviewee was incapable of understanding the questions in their *fuṣḥa:* form. Hence his rather hesitant performance, in which he vacillates between different language levels, exemplified below in a question he puts concerning the level of public receptivity (for religious reasons, among others) to the introduction of drama in the 1950s. The portions in bold are in the interviewer's dialect and the rest are in a close approximation of *fuṣḥa:*, with saliently nondialectal structures or vocabulary enclosed in square brackets:

> *bi nnisba li [taqabbul iljumhu:r], ah, . . .* **fi ða:k ilwagt** *. . . li ʔayy masraḥi:ya ṭabʿan, yaʿni ka:n ilmustawa, [mustawa taqabbul iljumhu:r] . . . [hal] ka:n, yaʿni, ah . . . bi lmustawa* **lli ḥna** *[alʔa:n]* **nču:fah,** *[mustawa iljumhu:r li taqabbul ʔayy masraḥi:ya tuʿraḍ] li ʔayy masraḥ [ʔin ka:n] fi lbaḥre:n* 'As far as the [receptivity of the public] is concerned, er, . . . **at that time** . . . to any play at all, I mean, [was the level of public receptivity] . . . was it, I mean, er . . . at the same level **as we see** [now], [the public level of receptivity to any play which is put on], of any theater [if it is] in Bahrain?'

In spite of the hesitations and dialectal insertions, the syntactic framework of this question is written MSA (Level 2)—viz. in the repeated construct nominalizations (*mustawa taqabbul iljumhu:r*), the use of the question particle *hal*, and of the internally voweled passive (*tuʿraḍ*). The choice of vocabulary is elevated commensurately with the MSA syntax but it is striking how both dialectal and MSA elements are used as a means of counterbalancing syntactic sequences in the other code: dialectal *fi ða:k ilwagt* 'then' breaks up an MSA complex nominal that sounds rather unwieldy when spoken, whereas the markedly MSA *alʔa:n* 'now' is slotted into the middle of a dialectal sequence *lli iḥna nču:fah* 'which we see' where the Bahraini *alḥi:n* would be expected. Compare this with the opening of the interview—normally its most formal part—with Mbārak al-Rāʿī, reproduced below, in which the interviewer goes straight into dialect, syntactically, morphologically, phonologically and lexically. Moreover, in contrast to the address form *usta:ð ja:sim* 'teacher Jasim' given to Jāsim Khalaf to acknowledge

his expertise as an actor, the interviewer addresses the interviewee by his teknonym *abu sa'ad* 'Father of Sa'ad'. This is also a sign of respect, although quite different in its connotations from the *usta:ð* that the interviewer gives to Jāsim Khalaf. It is typical of the Bahraini society of the preoil age when respect among ordinary people was not dependent on learning or economic status and is highly appropriate to one whose claim to fame is so closely associated with the culture of that society. Teknonyms are still commonly used, but as here, when given by an educated person to an uneducated one not from his family or well known to him, have the flavor of "talking down":[25]

> *fi bida:yatna iḥna ta'awwadna gabil ʔinna dayman ngu:l 'an ḥaya:t ilfanna:n, mita bida, wa immita ila ʔan wuṣal ila a:xir marḥala min ḥaya:tah u šyiku:n fi:ha . . . fa ya: abu sa'ad illi abbi minnak ʔinnak tiḥaddiθni 'an ḥaya:tak* 'At the beginning, what we usually do is talk about the life of the artist, when he started out right up to him reaching the most recent period of his life, and what he's done during it (lit. and what is in it) . . . so, Abu Sa'ad, what I want you to do is tell me about your life'.

Changes in language level are also used to signal changes in the role a speaker is playing or claiming for him/herself in a conversation, just as in the case of the political monologues discussed earlier. This is well illustrated in recordings gathered by the Leeds University team that investigated the structure of educated conversational Arabic on "serious" subjects in the 1970s.[26] In one conversation, a group of two Jordanians (both male diplomats) and a Syrian (female secretary) who all live in Damascus are discussing the shortage of housing in the city. One of the researchers begins by asking the assembled company whether there is, as he has heard, an accommodation crisis. The beginning of the conversation is dominated by the two men, who in turn deliver long but apparently unprepared monologues on the nature of the problem and its causes (rent inflation, lack of prefabricated building materials, influx of refugees from the Lebanese civil war, landlords' fear of sitting tenants, etc.) in a close approximation to *fusḥa:* (i.e., Badawi's Levels 2/3). They treat the issue impersonally as an abstract "social problem" and the tone is one of an academic lecture. After several minutes of this, one of the researchers cuts in to ask the Syrian woman whether she has any comment, to which she replies in a nervous and hesitant *fusḥa:* that she agrees with the previous speaker. But when in reply the researcher jokingly asks her whether she, as a Syrian and a house owner, is causing part of the problem by concealing another empty property, there is general laughter, and the level of language used by the woman in reply immediately shifts to plain and rapidly delivered Damascene colloquial as she gives an account of

her own family's housing problems. The point to note here is that the language level is related to the degree of distance the speakers put between themselves and the topic: the two men, and initially the woman, following their lead, at first treat it in an impersonal way, as perhaps might be expected of diplomats talking on the record. But as soon as this atmosphere of academic detachment is punctured by the researcher's flippant intervention and the topic is given a more *concrete* and *personal* slant, there is an immediate shift toward the dialect. Later on in the conversation, the topic is once more dealt with in a "third person abstract" rather than in a "first person narrative" fashion, and the language level shifts once again back to an approximation of *fusḥa:*.

A related aspect of this form–function relationship is the use of a more dialectal style for what might be termed "organizational discourse." Even in parts of conversations where speakers are, like the two Jordanian men, giving an authoritative statement of "how the world is" in a language level approximating to *fusḥa:*, parenthetical asides and comments based on personal experience and invitations to other participants to state their views on the matter at hand are delivered at a more dialectal level. As we have noted, Nasser used exactly the same ploy in his speeches to mark off what was to be considered "text" from "commentary on text" for his audiences. Some examples of this are the following three extracts from the Leeds University corpus,[27] in which the back-channel "commentary" and "organizational discourse" are in bold.

Extract 1: On the housing crisis in Damascus

> ' . . . ilwa:qi' ʔinnu lʔamr ha:ða ma: ka:n maqṣu:r 'ala: dimašq li waḥdha: bal ka:n mumtadd fi: šatta: ʔanḥa:ʔ ilwaṭan il'arabi . . . *ʔana ba'raf maθalan fi lurdun ʔinnu fi:h ʔazmat sakan rahi:bi jiddan kama:n* . . . nafs il'awa:mil wa ððuru:f ʔallati natakallam 'anha: fi: dimasq hiya mawju:da fi: 'amma:n . . . ' 'The fact is that this [problem] is not limited just to Damascus, but affects various areas of the Arab World . . . **I know, for example, that there's a terrible housing crisis in Jordan too** . . . the same factors and circumstances which we're talking about in Damascus are also found in Amman . . . '

Extract 2: On the existence of a "shared" pan-Arab dialect

> . . . bi nnisba li llahja lmuštaraka, ya'ni laha: wuju:d, bi dali:l ʔannana nastaṭi:' ʔan nafham 'ala: lixwa:n ilmiṣri:yi:n wa 'ala: lixwa:n illi:bi:yi:n wa l'ira:qi:yi:n wa lurduni:yi:n aw issa'u:di:yi:n, ya'ni huna:k juðu:r muštaraka bi nnisba li lmufrada:t ilmusta'mala fi llahja:t il'a:mmi:ya, fi: 'adad kabi:r minnah, juðu:ruh fusḥa:, li ða:lika ya'ni fi:h ʔimka:ni:yat ilfaham, wa xa:ṣṣatan 'ala: mustawa il "standard", ya'ni 'ala: mustawa — (interruption) **ma:**

he:k, ya'ni niḥna biniqdar nifham 'ala: ka:ffat il?iða:'a:t aw ka:ffat iṣṣuḥuf, maẓbu:ṭ willa la?'

'As far as a "shared dialect" is concerned, it *does* exist, which is proved by the fact that we can understand Egyptians, Libyans, Iraqis, Jordanians or Saudis . . . there are shared (linguistic) roots as far as the vocabulary items used in the dialects is concerned, many of them have Classical roots, and for that reason there is the possibility of understanding, especially at the level of the "standard", I mean at the level—(interruption) **Isn't that so? We can understand all the radio stations and newspapers, right or not?'**

Extract 3: on Arabic language education at school

fa ?iða kunna naḥnu ḥari:ṣi:n 'a lluḡa fi'lan, 'a lluḡa l'arabi:ya, alluḡa l?umm, ya'ni illi hiya min al?usus al?u:la: li lwaḥda l'arabi:ya, fa yajibu ttarki:z bi lmarḥala libtida:?i:ya 'ala: ?aṣwa:t illuḡa lfuṣḥa, ya'ni likay la: taðu:b bi ?aṣwa:t illuḡa l?a:mmi:ya . . . ma: ba'raf ya'ni šu: ra?y ilixwa:n? 'So if we really want [to teach] the language, the Arabic language, the mother tongue, which is one of the basic foundations of Arab unity, we must concentrate at the primary stage (of education) on the sounds of the Classical language, so that they do not get mixed up with the sounds of the dialect . . . **I don't know, what do other people think?**

In all of these extracts, the text that precedes (in the case of (1) both precedes and follows) the section in bold is delivered at a relatively elevated language level: one notes for example in (1) the *ma: . . . bal* 'not only . . . but also' syntactic structure, the literary phrase *fi: šatta: ?anḥa:? ilwaṭan il'arabi*, 'in various areas of the Arab world', and, phonologically, the standard pronunciation of /q/ and the interdental fricatives, where Damascene, the speaker's dialect, has /?/ and dental stops. Extracts 2 and 3, spoken by another Syrian, are if anything even fuller of classicisms, possibly because the topic of the conversation—the structure and teaching of the Arabic language itself—exercises its own especially formalizing influence. But in all cases there is a sudden downward switch in the section in bold. In (1), the switch occurs in an aside made on the basis of the speaker's personal experience, which he makes to support a generalization; in (2), it is a rephrasing of a general point in a more personalized way ('we can understand . . .') and a soliciting of another speaker's agreement, triggered by an apparently challenging interruption; in (3), the switch to dialect occurs when other speakers' views are sought on a point that has just been made. In every case the switch to dialect happens at a point where speech function changes: where personal opinion or knowledge is being expressed or sought.

Spoken communication between Arabs who do not share the same dialectal background, as in the artificially engineered case just described, is of course an inevitable byproduct of the greater degree of geographical mobility that the economic and political developments in the Arab world have brought about during the last fifty years, and especially in the last thirty. But it is now no longer just the case of the well-educated Palestinian or Egyptian university lecturer who resides for years in one of the developing Gulf states, but of illiterate Sudanese, Yemeni, or Egyptian migrant workers laboring for years on the building sites or farms of Saudi Arabia, or of the Tunisian scholarship student studying in Kuwait. How do these various groups cope linguistically? Surprisingly, few studies have been done. The main data-based ones are Blanc's study of a single extended conversation between four expatriate Arabs studying in the United States,[28] Talmoudi's replication of Blanc's study but with North African Arabs from different areas,[29] Ezzat's monograph on interdialectal intelligibility,[30] and several studies done by the Leeds-based research group of the 1970s[31] from whose data the example discussed immediately above was derived. None of these studies, however, deal with illiterate or poorly educated speakers, and the Leeds study was specifically aimed at describing the emergence of a supranational Educated Spoken Arabic, roughly equivalent in its circumstances of use, in an interdialectal context, to Badawi's Cairene Level 3 *'a:mmi:yat al-muθaqqafi:n.* From a general sociolinguistic perspective, only the research by the Leeds group was set up in such a way, was sufficiently wide ranging in geographical coverage, and generated sufficient data for any conclusions to be generally valid; unfortunately, however, the results of this project are still largely unpublished.

The major finding of Blanc's study, a landmark in the study of variation in Arabic, was that interdialectal conversation is subject to two tendencies: "classicizing" and "leveling." "Leveling" refers to the tendency of speakers to eschew dialectal elements that are highly localized in favor of alternatives that have greater areal currency. This can affect all linguistic levels from phonology to lexicon, but it is particularly marked in the latter, the level of language structure at which speakers are themselves most conscious of dialect differences. A simple example will illustrate the principle. In a conversation between educated Arabs from the Gulf, Baghdad, Cairo, and Jerusalem, there are available to the speakers at least three dialectal ways of expressing "existential" 'there', as in 'there are people who think . . . ', that are geographically distributed as follows:

	GULF	BAGHDAD	CAIRO	JERUSALEM
'there is/are'	*hast/aku/* *fi:*	*aku*	*fi:*	*fi:*

In this case, *fi:* is the variant likely to be used by all of the speakers because it has no association with any particular area and represents the nearest thing to a dialectal "common denominator" for this particular group. Speakers in a heterogeneous group tend to "level" their speech in the direction of what they recognize as a pan-Arab dialectal mean even if this sometimes involves, as it does here for the Iraqi, using a dialectal form that is not Iraqi at all. The preparedness of speakers to shift to dialectal forms that are not their own does vary, however. On the one hand, a Bahraini or Qatari would be most unlikely to use *hast* in a cross-dialectal situation because he or she might not even be understood by speakers from outside the Gulf, so localized is this word. On the other hand, Egyptians in particular seem much less inclined to shift away from Egyptianisms not found in other dialects, perhaps because of the dominant position that their dialect has established for itself over many decades in the educational systems and media of most Arab countries.[32] We are here talking about "leveling" in a situation in which the dialects of the collocutors are markedly different from each other, but the same principle holds even where they are from less widely dispersed geographical areas. Speakers from less well-known or prestigious communities in a given country will tend to accommodate their speech to the dialect of a more prestigious area or group when the occasion demands, for example, peasants from the Nile Delta or Upper Egypt will make an effort to approximate to Cairene usage when they visit the capital; schoolchildren who use a "Bedouin" dialect with family will switch to an "urban" one at school.[33]

At the same time as the tendency to "level," one observes in interdialectal conversation between educated speakers a tendency to "classicize." Whereas "leveling" seems to be a reaction to the background dialectal differences between speakers, whose aim is to emphasize shared elements and zero out localisms, the "classicizing" tendency reflects a deliberate stylistic shift—which may be sustained for a short phrase, several sentences, or much longer—beyond this. Hence it is found in monodialectal as well as cross-dialectal contexts. Reverting to the case of possible Arabic equivalents of "existential 'there'," the choices *huna:k* or *yu:jad(u)* would be examples of classicization in any extempore speech context, because no dialect, only MSA, uses these words as its normal way of expressing 'there is'. As we saw in the context of Nasser's speeches and in the behavior of the two Jordanian men in the example discussed earlier, "classicizing" is more often than not a reflection of the treatment of topic per se. Where speakers are treating a topic in a "formal," "serious," or "academic" way (for whatever reason), and *a fortiori* when the topic or situation is intrinsically serious or formal (e.g., a discussion on religion; a job interview), they "classicize" their dialect

to signal this, using nondialectal forms that may range from a mildly formalizing switch from dialectal to standard phonology (e.g., switching from Cairene /ʔ/ and Baghdadi /g/ to MSA /q/) to, in the most extreme case, the total replacement of dialectal forms by MSA syntax and lexicon. However, even in what appears to be the "same" situation and on the "same" topic, individual speakers' (as well as individual national groups') propensities to classicize differ, and the differences seem to depend as much on attitudinal factors as on their objective knowledge of MSA.[34]

9.4 Co-occurrence phenomena and level

So far we have been talking about "levels" as if they were discrete linguistic objects consistently and measurably different from each other. However, we have yet to demonstrate this. Although no one would dispute that the pure *fuṣḥa:* of an Azhari cleric and the colloquial of an illiterate laborer are distinguishable from each other in virtually every particular, there is much less agreement among linguists about the defining characteristics, or even the number, of levels between these two extremes.[35] A recent study has shown that, at least in one limited area at the top end of the cline (close-to-*fuṣḥa:*), ordinary Egyptians show a significant level of agreement in their ranking of different versions of the same speech extracts as more or less "classical,"[36] but more wide-ranging research on what makes conversational texts seem more "dialectal" or more "classical" to native speakers (and what significance, if any, they attach to different kinds and sizes of difference) has yet to be done. In this section we will therefore attempt, on the basis of actual examples, to give a linguistic characterization of selected aspects of these "middle" stylistic levels and try to articulate the rules that limit the combinatorial possibilities.

Standard and dialectal elements interact linguistically with each other at several levels. Let us begin at the minimal level—the word—and consider the phenomenon of hybridization, the cover term we will use for the process by which words are formed that are neither textbook *fuṣḥa:* nor plain dialect. Hybrids occur in all conversation that moves even marginally outside the concerns of the everyday. The susceptibility of a given dialectal or MSA word to hybridization by acquiring elements "borrowed" from its analog depends in the first place on the degree of distance between the competing forms, which of course differs from dialect to dialect (see below). Consider the following pairs, of which ((1)–(3)) illustrate gradually increasing degrees of MSA–dialect morphophonological difference and ((4)) an MSA–dialect lexical difference, for one dialect:

	MSA	CAIRENE	
(1)	*rama:*	*rama*	'he threw'
(2)	*qa:la*	*ʔa:l*	'he said'
(3)	*'arafa*	*'irif*	'he knew'
(4)	*raʔa:*	*ša:f*	'he saw'

In (1), the only difference between the citation forms is the length of the final vowel. But in spoken MSA this long vowel is neutralized in prepause position, making it identical with the dialect form, and in both the MSA and the Cairene it is pronounced long if the verb has an object suffix, that is, *rama:ni:* 'he threw me' for both. This is a case where the competing forms are so close to each other as to be for all intents and purposes identical. In (2) there are two differences, one in the initial consonant, the other in the final vowel of the MSA form, which inflects the form for 3msg. The hybrid form here is *qa:l*, which is, as it were, stylistically bivalent: if it occurs in prejunctural contexts where the final -/*a*/ of MSA *qa:la* would normally be sounded, *qa:l* can be regarded as "dialectalized MSA" (loss of person marker); however, it could equally well be a token of "standardized dialect" (substitution of /*q*/ for /*ʔ*/). Only the surrounding context (i.e., whether the neighboring linguistic features and vocabulary suggest "toned-down" MSA or "elevated" dialect) would make it clear which was the most likely interpretation of the speaker's intention in using such a form. In (3), there are three differences: the vowels in the first and second syllables—which could be considered a single discontinuous vowel morpheme (or "vowel prosody")—and the presence/absence of the final inflection, as in (2). The possible hybrids are *'araf* (loss of inflection, as in (2)) and *'arif*, which combines two of the dialectal features (loss of inflection and change in the quality of the second vowel) with one of the MSA ones (quality of the first vowel). *'araf* is nearer to "textbook" MSA, *'arif* nearer the ordinary dialect. As in (2) it is impossible to decide out of context whether such forms are the product of dialect style raising or a "toning down" of MSA.

It is also worth bearing in mind that in cases like (1)–(3) the extent of the differences between MSA and different dialects is not always the same: compare the Cairene forms with, for example, Gulf *'araf* and urban Omani *qa:l*, both "plain dialect" forms to their users and both nearer to MSA than their respective Cairene equivalents.

Pair (4) illustrates the case where hybridization ceases to be a possibility: either the MSA or the dialectal lexeme has to be selected, with no possibility of compromise. This last case applies to virtually all Arabic speakers, because *ša:f* is the near-universal dialect norm. Consequently, a lexical shift of the type

ša:f → *raʔa:* (or vice versa) is seen by native speakers as a stylistically more significant move, in whichever direction it occurs, than hybridization of the type illustrated in (1)–(3).[37]

Hybridization thus involves the marrying of features drawn from competing linguistic variants to form a new variant. This is not a random, but a rule-governed, process. Just as dialect base forms vary from one area of the Arab world to another (reflected in different preferred syllable structures, variation in the stability of unstressed short vowels, etc.), so too do the local rules for hybridization. In the case of Pair (3), for example, dialectal *'irif* can be hybridized in Cairo by changing the first vowel but not, apparently, the second without the first also (viz. **'iraf* does not occur). This is probably for two reasons: first, in Cairene, the *CiCaC* concatenation where all *C*s are root consonants is a *nominal*, never a verbal template; second, in all *CvCvC(v)* concatenations the prominent syllable is the first one in both Cairene and MSA, and a change in the quality of the vowel in this syllable is a more perceptually obvious "move" from Cairene toward MSA than would be a change in the vowel of the unstressed syllable. Of course, this is not to say that *'iraf* could not occur in some other location as a hybrid or *a fortiori* a "base" dialect form (it does so in Baghdad, for example) where the constraints are different. However, it seems to be the case that variants that are distinguished by internal short vowel differences only, that is, of the *'araf/'irif/'arif/'iraf* type, do not carry all that much stylistic significance for speakers, who may vary between two or even more of them in an unpredictable way.

The rule-governed nature of hybridization can be illustrated by presenting speakers with some types of hybrids that could theoretically occur if hybridization were random, but which it turns out they will not accept as possible forms. Nonoccurring examples rejected as impossible derived from (2) and (4) above would be **ʔa:la* for 'he said' or **ša:fa:* 'they two saw'. These are unacceptable because they violate not a phonological constraint, but a principle of stylistic compatibility: the final -/*a*/ inflection in **ʔa:la* and the dual inflection -/*a:*/ in **ša:fa:* are classicisms too elevated in the stylistic "pecking order" to be combinable with forms as markedly dialectal as /*ʔ*/ in the first case and the lexeme *ša:f* in the second. Out of context, a form such as *ʔa:la* would be understood by a native-speaker only as a noun from a different root meaning "tool" or "machine"; **ša:fa:* would be interpreted as a joke form, perhaps in the mouth of an illiterate person in a comedy trying unsuccessfully to "talk posh." This example illustrates a general point touched on earlier in 2.3.2, viz. that there are selectional restrictions between linguistic elements and that these restrictions are asymmetrical: a lexical item such as *ša:f* that is marked as dialectal rules out non-

dialectal choices such as MSA inflections and morphophonological realization, but the converse does not apply so strictly—a lexical item such as *ðahaba* 'to go', which is marked as MSA, and *a fortiori* lexical items that are shared by both MSA and the dialect, admit of dialectal morphophonological realization(s) and can be combined with dialectal grammatical morphemes (e.g., those indicating mood and aspect). Hybridization is not, in other words, an unrestrictedly bidirectional process.

Not only does the compatibility principle operate across morpheme boundaries within phonological words, but also across word boundaries within phrases and sentences. In general, however, the strength of the co-occurrence restrictions decreases as the closeness of the syntactic relationship between the constituents decreases. Where major syntactic boundaries occur—at clause and sentence boundaries, and even between matrix sentences and embedded elements—the co-occurrence constraints are weaker, and it is here that switches—from elevated MSA to plain dialect and vice versa—can occur. In order to illustrate the relative strengths of different kinds of boundary and to keep the discussion within manageable limits, we will examine (a) selectional constraints *within* some types of verb phrase, and (b) selectional constraints *between* the verb phrase and the rest of the sentence in which it occurs. The data here are drawn mainly from my own analyses of excerpts from the Leeds University corpus and from the published papers of members of the Leeds team.

9.4.1 Constraints within p-stem verb phrases

Recall that, expressed in its most general form—that is, covering MSA and the dialects—the p-stem finite main verb in Arabic has the following morphological elements (optional elements in parentheses):

(NEG) + (mood/aspect) + prefix + stem + suffix + (O) + (NEG)

Thus the Cairene *ma: byiḥibbuha:š* 'they don't like her' fills all these slots:

ma:	+ *b*	+ *yi*	+ *ḥibb*	+ *u*	+ *ha:*	+ *š*
NEG	aspect	prefix (3m)	p-stem 'like'	pl	O (3fs)	NEG

The equivalent MSA form is *la: yuḥibbu:naha::*

la:	+ *yu*	+ *ḥibb*	+ *u:na*	+ *ha:*
NEG	prefix (3m)	p-stem 'like'	mpl	O (3fs)

Aside from differences in the morphophonology, such as the different shape of the verb prefix, it can be seen from this particular pair of examples that MSA

lacks an overt aspect marker where the verbal aspect is habitual or timeless and that the MSA negative particle is a single preposed element, as opposed to a "fore-and-aft" discontinous one, like French '*ne . . . pas*'. MSA does of course have a tense marker that could fill the second position in this VP complex, but only for future tense, when *sa-* or *sawfa* is prefixed, for example *sayuḥibbu:naha:* 'they will like her'. The question is, where and how within this complex does a "dialectal" or "MSA" selection restrict selections in the rest of it? In answering this question, we should make it clear that selection does not proceed in any particular (left–right, right–left) direction. Selections made *within* VP complexes produced in normal conversation are envisaged as *co-occurrent*, not consecutive. However, there nonetheless seems to be a hierarchy of features such that selections at certain points *imply* co-occurrent selection of others. A simple illustration of this is provided by the following four possibilities in Cairene:

STYLE	FORM	
most formal	*saʔaqu:lu*	
↑	*saʔaqu:l*	
↕		'I will say'
↓	*ḥaʔaqu:l*	
least formal	*ḥaʔu:l*	
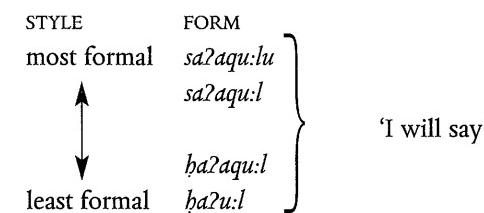
Starting from the "most formal" (MSA) end of the spectrum, the selection of final -*u* as a mood marker (indicative) implies the co-occurrent selection of *sa-* as a tense marker and /*q*/ in the phonological realization of the stem. The next most restrictive element is *sa-*, which, while not requiring final -*u*, still requires /*q*/ in the stem. Ascending from the bottom now, the omission of the stem prefix *ʔa-* is the most restrictive dialectal feature, requiring the simultaneous selection of dialectal *ḥa-* as the mood/aspect marker, and the realization of the initial consonant of the stem as /*ʔ*/, not /*q*/; but if this *ʔa-* is not deleted, either *ḥa-* or *sa-* can co-occur, with *ḥaʔaqu:l* representing a less formal choice stylistically than *saʔaqu:l*. Several possibilities are ruled out by the principle of stylistic compatability, for example, *saʔu:l*, *ḥaʔaqu:lu*.

In fact, there seem to be strong restrictions on the type of elements co-occurring in the VP when MSA particles are selected: MSA elements are favored or even obligatory elsewhere in the VP; but there are only weak restrictions where the dialectal particles are selected: dialectal or MSA elements, with a few exceptions, are possible in all slots. A two-minute contribution by an educated Egyptian to the conversation on housing problems,[38] from which we have taken earlier examples, illustrates this well. There are some sixteen examples of p-stem

VPs functioning as main verbs, marked in five cases according to the MSA tense/aspectual system (which in cases (iv) and (v) below means being zero marked) and in eleven cases according to the dialectal system. The tokens can be divided into three groups:

Group 1: MSA marker + MSA verb

(i)	*sayamurr*	'he will pass'
(ii)	*sayuwa:gih*	'he will face'
(iii)	*qad yastamirr*	'it may continue'
(iv)	*yas?al*	'he asks'
(v)	*yagid*	'he finds'

All of these examples contain sets of co-occurrent features that are MSA. In all cases, the vowel of the *y*- prefix follows MSA, not Cairene, norms, and in four out of five cases (the exception is (iv), which is lexically "shared" by the MSA and dialectal lexica), the verb is an MSA, not a dialectal, choice. For example, (i) would be *ḥayfu:t* in plain Cairene, and (v) would be *byil?a*. Apart from the lack of the final vowel indicating mood, (i)–(v) are, lexically and morphophonologically speaking, MSA. Compare these now with (vi)–(xvi), all of which have the dialectal aspectual/modal marker *b(i)-*. These eleven examples can be divided into two distinct groups.

Group 2 contains examples that have a dialectal aspect marker but in which the verb prefix + stem component is, like (i)–(v), basically MSA in character. Examples (vi)–(viii) are internally voweled passives, whereas (ix), (x), and (xi) contain stems that from the lexical point of view are again MSA rather than dialectal, with *y*- or *t*- verb prefixes to match. So (vi)–(xi), from the point of view of the morphophonology and lexical status of their verb prefix + stem, are also MSA, and in this respect indistinguishable from examples (i)–(v):

Group 2: dialect marker + MSA verb

(vi)	*byutrak*	'he is left'
(vii)	*byuqbal*	'he is accepted'
(viii)	*byu:jad*	'there is' (lit. 'it is found')
(ix)	*bitattasim*	'it is characterized'
(x)	*byastaṭi:'u*	'they are able'
(xi)	*btu?ajjir*	'it will lease'

Group 3, containing examples (xii)–(xvi), combines the same dialectal marker *b(i)-* with stems and other morphophonological elements that are dialectal:

Group 3: dialect marker + dialect verb

(xii)	*ma: bitḥibbiš*	'it doesn't like'
(xiii)	*biyfaḍḍalu*	'they prefer'
(xiv)	*bitku:n*	'it is' (auxiliary verb)
(xv)	*bitibni*	'it builds'
(xvi)	*byimla*	'it fills'

It would be rash to draw firm conclusions from such a small amount of data as this, and it must also be borne in mind that the speaker is an Egyptian whose speech may in some ways make him untypical of educated Arabs as a whole (see n. 32). But what is clear is that whereas the selection of MSA aspectual/tense particles requires the coselection of MSA lexical items and morphophonology in the same VP, the selection of dialectal particles is not necessarily associated with other dialectal features: *b(i)-* is "multilevel" (as is the future/intentive particle *ḥa-*) in ways that *sa-* and *qad* are not. In fact, the examination of larger samples of conversational Arabic than we have space to describe here suggests that this is one of the main strategies speakers follow in educated conversational Arabic: they marry together *clitic elements* and other functionals drawn from the dialect, covering basic syntactic functions such as negation, aspect, and tense in the VP, and possession and deixis in the NP, with *vocabulary* from MSA. So a hybrid combination such as *biyadhab* (dialectal *bi-* + MSA stem) 'he goes' is possible, but *sayiru:ḥ (MSA *sa-* + dialectal stem), with the same meaning, is not. The morphophonology of the hybrids is to some extent a matter of (Labovian inherent) variability: if we take Cairene *b(i)-* and MSA (minus its final inflection) *yankasir* 'it breaks', there are several possible variants: *biyankasir* (with "MSA" voweling), *byinkasir* (hybrid voweling), and *byinkisir* (dialect voweling).

9.4.2 Constraints within the sentence

The fact that the main verb occurs early in the sentence in spoken Arabic and that it often has no overt nominal or pronominal subject preceding it gives it an important role in setting the stylistic key of what follows. If, for example, in talking about a schoolteacher's responsibilities a Syrian wishes to say 'he can prepare what he needs in the way of teaching aids',[39] the main verb corresponding to 'he can', which begins the sentence, can be realized in a number of ways (cf. the examples adduced in section (a) immediately above)[40] that restrict the choices made in the rest of the sentence:

	MAIN VERB	COMPOSITION OF VERB	STYLE
(1)	*yastaṭi:'u*	(MSA lexis, inflected for mood)	high-flown
(2)	*yastaṭi:'*	(MSA lexis, no inflection)	formal

(3) *byastaṭiːʿ*   (MSA lexis, dialectal marker)   careful
(4) *byiʔdar*   (dialectal lexis, dialectal marker)   informal

The choice made here would constrain verb complementation in a fairly predictable way:

Choice (1) would require both the MSA complementizer *ʔan* corresponding to English 'that' and a correspondingly high-flown morphophonological realization (*yu-* prefix, with optional a-set inflection) of the verb corresponding to 'prepare'. That is:

| *yastaṭiːʿu* | *ʔan yuẓahhiza* |
| | *ʔan yuẓahhiz* |

Choice (2) would also require the MSA complementizer, but the omission of the mood inflection in the main verb makes it unlikely that the governed verb would be inflected:

yastaṭiːʿ ʔan yuẓahhiz

Choice (3) might be followed by a complementizer, which could be high-flown *ʔan* or more likely dialectal *ʔinnu*; alternatively (and more dialectally) the complementizer could be omitted. Whichever complementizer choice is made, the following verb is likely to harmonize: *ʔan* with *yu-* verb prefix, *ʔinnu* or zero with *yi-*:

	ʔan	*yuẓahhiz*
byastaṭiːʿ	*ʔinnu*	*yiẓahhiz*
	ϕ	*yiẓahhiz*

Choice (4) might or might not have a complementizer, but if it did it would be *ʔinnu*; the realization of the following verb would be correspondingly dialectal:

| *byiʔdar* | *ʔinnu* | *yiẓahhiz* |
| | *ϕ* | *yiẓahhiz* |

Two general points are suggested by this example. First, the principle of compatibility whose operation we illustrated within the internal structure of the main verb is operating here too, although across word boundaries: there are no violent switches from one extreme to the other, for example, *byiʔdar ʔan yuẓahhiza* would sound distinctly odd, as would (although perhaps less so,)[41] *yastaṭiːʿu ʔinnu yiẓahhiz*. Second, note the "multivalency" of forms such as *yuẓahhiz/yiẓahhiz*. Such forms, which are formally similar to each other and which between

them cover the whole of the stylistic range, provide crossover points that allow speakers to change direction relative to the direction they were moving in before. For example, after *yuẓahhiz* in our example sentence, the noun phrase 'what he needs in the way of teaching aids' could be *ma: yalzamu min wasaːʔil taʿliːmiːya* (continuing high) or *illi yilzamu min wasaːʔil taʿliːmiːya* (crossing over and going lower), and the same choices are available after *yiẓahhiz*, in this case signaling, respectively, crossing over and going higher, or continuing low.[42] The existence of such "multivalent" forms that are morphophonologically neutral as between higher and lower levels is obviously vital to the flexibility and indeed the basic functioning of the system.

The above example, which illustrates the syntagmatic constraints within the constituents of major sentence parts and the indeterminacy of the system that crossover points necessarily provide in order to allow level switches, may seem somewhat theoretical. A very striking example of the mechanism in operation in actual conversation is provided by Meiseles, cited by Mitchell.[43] After a lengthy contribution to a conversation by a senior army officer, a collocutor is asked by one of the Leeds University researchers if he has any comment and replies:

——dialect——		crossover		——(high-flown) MSA ——
ma: fiːš ʿandi ḥaːga	*ʔuḍifha*	*ʔila ma: qaːlahu lginiraːl*
'I have nothing	to add	to what the general said'

In the first part of this sentence the selections at all linguistic levels are in plain Cairene, the speaker's dialect, but in the third part they are not merely MSA, but "high-flown" MSA. The crossover from dialect to high-flown MSA occurs via the "multivalent" hybrid *ʔuḍifha* (lit. 'I add it') at the beginning of the relative clause that modifies *ḥaːga* 'anything'. Lexically, the verb used for 'to add' is perhaps more associated with MSA and written Arabic than with dialect. From the point of view of its morphophonology, the form is mixed: the *ʔu-* first person prefix is MSA, but the rest of the form, notably the shortening of the second (stressed) vowel from MSA /iː/ to /i/, reflects Cairene syllable structure preferences and morphology. It thus provides a potential "bridge" to a higher level, or the speaker could continue low. In the event, the speaker crosses the "bridge," building on the *ʔu-* of the prefix and ascending further the stylistic cline, in order, no doubt, to reflect "the respect felt by the speaker to be due to the general" when referring to him directly.[44]

The points at which sudden level shifts are most likely to occur, as in the above example, are major grammatical boundaries within the sentence or at the

boundaries of sentences themselves. Their function, as we saw in 9.3, is to signal changes in the formality of the situation or in the rhetorical function of what is being said, and they may be accompanied by pauses and paralinguistic phenomena such as changes in voice pitch and quality, changes in the speed of delivery, clearings of the throat, laughter, etc. Within sentences, these stylistic shifts typically occur via the crossover mechanism just exemplified. But there is another type of shift, whereby isolated highflown elements are slotted into a syntactic framework that is at a (sometimes considerably) lower level, in an apparently unpredictable and unmotivated way. Consider the following examples:[45]

(a)————————informal——————|–high-flown–|——informal|———
baʔaːli   ḥawaːli sabiʕ siniːn   lam aqraʔ   našrit ʔaxbaːr
remains-to-me about seven years I didn't read a news bulletin
'It's about seven years since I last read a news bulletin' (Lebanese actor interviewed on Jordan Radio)
(b)——formal——|———high-flown(+++)—|
taḥmil turaːθ   ʔarbaʕata ʕašara qarnan
carries culture   fourteen   centuries
' . . . transmits the culture of fourteen centuries' (Arabic specialist explaining the aims of teaching Arabic at school)

In (a) a negative verb phrase, which is saliently MSA in its morphology, has been inserted into a syntactic environment that is, in terms of Badawi's model, two stylistic levels below it: a Level 1/2 structure slotted into a Level 4 framework. In (b) a noun phrase involving number—a phrase category that even in the formal scripted speech of television news bulletins is usually realized in its dialectal form, such is the complexity of the MSA rules involved and the fear readers have of getting them wrong[46]—is given its full, correct "high-flown" treatment (hence the '+++' marking). In spoken Arabic, such faithful observance of case endings in this type of phrase is typical only of highly formal, often religious discourse, and it contrasts in the example with the inflectionless "toned-down" MSA that immediately precedes and follows it.

We do not seem here to be dealing with the "crossover" phenomenon: the shift to a higher level does not occur via an MSA–dialect hybrid, but is sudden and short-lived, limited to a single phrase, after which the style level falls back to what it was before. Nor, *a fortiori* does the suddenness of the shift signal a complete change in rhetorical function like those we illustrated from monologues and general conversation in 9.3. However, it does seem to have a more localized rhetorical function. By using isolated MSA forms, the speaker selec-

tively emphasizes elements of his message more than he would if he "performed" these meaning elements in the dialect. Thus in (a) the use of MSA negative *lam ʔaqraʔ* 'I didn't read' draws attention to itself because of the dialectal environment in which it is located. By selecting this saliently MSA negative, and thereby implicitly appealing to the connotations of "authoritativeness" and "formality" that MSA carries for any educated Arab listener, the speaker states more categorically than he would if he used the equivalent dialectal expression that he has *definitely not* read the news for seven years. The emphatic tone of *lam ʔaqraʔ* is thus *contextually* derived: the form is not inherently emphatic but becomes so here because of the sharp contrast with the informal dialectal framework in which it is used. The use of an MSA element in example (a) is thus less a stylistic shift in the sense in which we have been using the term than a means by which the speaker can overtly commit himself to the truth value of what he is saying.

The use of the high-flown number expression in (b) can be similarly explained. Here the speaker's choosing to inflect fully a phrase type that, even in many kinds of scripted MSA speech (let alone, as here, extempore conversation) is realized dialectally, draws the listener's attention to the meaning that lies behind this form. Because the speaker's general aim in the conversation is to point up the educational value, richness, and cultural continuity of Arabic literature (and no doubt incidentally to display, as an Arabic specialist, his effortless control of a tricky area of Classical Arabic morphosyntax!) it is an appropriate tactic for him to single out the number phrase "fourteen centuries" for special treatment. As in (a), the strategy is one of inserting markedly MSA syntactic elements to create local contextual contrasts and thereby emphasize the meaning of the inserted element. In (b), however, it is not so much a question of emphasising the speaker's commitment to the truth of what he is saying but rather his opinion of its importance.

9.5 Dialogue and dialect in literature and journalism

We conclude this chapter with some observations on the use of dialectal Arabic in "serious" (that is, nonfolkloric) literature and journalism—types of writing that are normally thought of as the province of MSA par excellence. In particular, we shall look at how the problem of writing dialogue in plays and narratives has been handled in a language like Arabic in which the gap between the norms of speech and writing is wide, and in which only the latter has traditionally been considered a medium for literature that is worthy of the name.

9.5.1 Literature

The idea that vernacular Arabic might be a suitable medium for serious literature arose in the 1950s following the widespread move to political independence in the Arab world in the aftermath of World War II. Political independence was accompanied by the beginnings of distinct national Arabic literatures in which the plots, settings, and subject matter reflected contemporary local preoccupations. After the coming to power in 1954 of Gamal Abdul Nasser and the Arab Socialist Union with its avowedly populist agenda, the debate in Egypt concerning the appropriateness of using dialect in serious literature became particularly bitter. The dispute was between, on the one side, the old literary establishment, many of whom doubled as members of the Arabic Language Academy, the bulwark of linguistic conservatism and the defender of linguistic purity, and on the other a loose assemblage of mainly young, mainly leftist writers.

The problem of what kind of Arabic could and should be used in literature loomed particularly large, for obvious reasons, in the theater. The theater as we know it in the West was introduced into the Arab world only in the middle of the nineteenth century but did not take root as an accepted form of indigenous art until well into this century. As a genre of "serious" literature, drama, like poetry and prose, had to be scripted in *fuṣḥa:* in which, of course, Arabs do not normally converse. As long as the setting was not meant to be realistic (e.g., the theater of the absurd, allegory), or especially if it involved the recreation of events from Islamic history, this problem could be ignored: in the first case, palpably unreal worlds do not need to pay attention to the detail of linguistic reality (in fact, the unreality is increased by having the characters speak *fuṣḥa:*), and however eighth-century Arabs actually spoke to each other, no playwright, nor any audience, was going to quarrel with the idea that it must have been in something like pure Classical Arabic (the linguistic strategy employed by the Egyptian Mahmud Taymour in his historical dramas *Ibn Jalā* and *Ṣaqr Quraysh*, both written in the 1950s). But plays set in the modern world posed a problem if the language that the actors were required to speak was to bear any relationship to observed reality outside the theater. Using *fuṣḥa:* would give the proceedings an air of stilted unreality, whereas dialectal Arabic was still felt by many to be an unsuitable vehicle for any form of writing that had something serious to say.

Another Egyptian writer of the same generation as Taymour, Tawfiq al-Hakim (although, unlike him, more given to experimentation) attempted to come to grips with these linguistic difficulties in two of his plays, *al-Ṣafqa* (*The Deal*, 1956) and *al-Warṭa* (*The Dilemma*, 1966). *The Deal* is written in what al-Hakim termed "the third language": a species of Arabic that, as far as possible, conforms to the syntactic rules of written MSA on the printed page, avoiding lexical and syntactic choices that are markedly dialectal on the one hand and "high-flown" on the other. However, because MSA orthography underspecifies the morphophonological realization of words, the text can be performed in something akin to the dialect of any group of actors via the omission of most grammatical inflections, the pronunciation of short vowels and consonants according to local dialectal rather than MSA norms, etc. In *The Dilemma*, al-Hakim's solution to the problem of writing natural-sounding dialogue is somewhat different. This time, the written text does incorporate (Cairene) dialectal features, although these are limited mainly to high-frequency clitics and functionals and their associated nonstandard word orders, for example, interrogative pronouns, demonstrative expressions, and relative clauses. As we noted in 9.4, these are among the dialectal elements most resistant to replacement by their MSA equivalents in natural conversation, and hence it is easier for the actors to speak the dialogue in this play with something approaching the rhythms of natural speech—provided, that is, that they are Egyptians. The drawback of this approach compared to the earlier experiment with "the third language" is that the use of dialect in the script inevitably locates the dialogue in a specific place even when the action and context of the play, as in *The Dilemma*, do not require it. To perform this play, Syrian or Iraqi actors would either have to fake Egyptian accents or recast the dialogue into their own dialect.[47] Significantly, these are the only plays of al-Hakim's vast output in which he attempts to write dialogue even in approximate conformity with the linguistic realities of everyday life. Both his solutions to the problem are the uneasy compromises of a theatrical innovator but a linguistic conservative, for whom Classical Arabic remained the vehicle of choice for "Literature with a capital L," however unnatural the results of this axiom turned out when applied to the live theater.

Occupying the opposite end of the literary–linguistic spectrum, we find the quintessentially Egyptian short-story writer and playwright Yusuf Idris (1927–1991). Idris was born in a Nile village and trained (and briefly practiced) as a doctor in the Qasr al-'Aini hospital in Cairo before giving up medicine to begin his career as a writer. Perhaps more than any other Arab literary figure, Idris's direct personal knowledge and sympathetic understanding of the hard life of ordinary citizens—in his case, Egyptian peasants and the lower echelons of Cairene society—dominated both the content of his writing and the linguistic form it took. From the early 1950s until the mid-1960s (roughly until the Six Day War of 1967), Idris published collections of short stories and some

plays in which the invariably Egyptian characters speak dialogue that is not in some carefully calculated compromise between MSA and dialect, but in a broad and colorful Egyptian. Even the narrative sections of some of the stories in his first collection 'Arkhaṣ Layālī (*The Cheapest Nights*, 1954) contain plentiful examples of dialectal lexis and even syntax.

Idris's championing of the vernacular had first and foremost an artistic motive. His stories evoke a starkly and palpably Egyptian world of sights, sounds, smells, social types, and human foibles, and the choice of a linguistic medium appropriate to this milieu was central if they were to convince. Idris himself insisted that his aim at this early point in his career was to break away from derivative western models of the short story and create a genre of literary story telling that was distinctively Egyptian from the point of view of plot, setting, and language.[48] The use of dialect was to him a natural consequence of this, but it frequently brought him into conflict with the literary establishment. Throughout his career, in many newspaper articles, Idris fulminated against what he saw as the linguistic pedantry and narrowness of view of the Arabic Language Academy and the Higher Council for Arts and Literature, which he felt had blighted the careers of many writers.[49] Thus, although the continued use of dialect for the dialogue in his early work was necessary primarily to establish a sense of social reality, it can also be seen as a deliberate slap in the face for the literary establishment and an indirect expression of sympathy for the socialist populism of Nasser, who, as we have seen, did not hesitate to use Cairene dialect in speeches to the people.[50]

Idris occasionally used MSA in his early stories and plays as a device to illustrate, often comically, the gap in education and world view of different social classes, or to satirize pedants and authority figures generally. In *Gumhūrīyat Farahāt* (*Farahat's Republic*, 1957) a policeman in a working-class quarter in Cairo is taking a statement from an illiterate young woman. At the end of the interview, the policeman, Farahat, puts his final question in the prescribed officialese of the form he is filling in, which the young woman completely fails to understand. A farcical exchange ensues (bold sections in MSA):

FARAHAT: *ya: bitt . . .* **hal ladayki ʔaqwa:lun ʔuxra:**
              'Now girlie . . . **have you any further statements to make?**'
WOMAN: *ʔuxra e:h ya: si:di*
              'What's a "**further**", then?'
FARAHAT: *'ayza tʔu:li ha:ga tanya ya'ni*
              'Got anything else to say!?'[51]

In the story *al-Nās* (*The People*, 1957) language differences are also used to underline educational and social divisions, here to mock the misplaced missionary zeal of some university students, who attempt to convince a group of illiterate villagers that the tamarisk tree they think has magical therapeutic properties is in fact harmful. The students are depicted as launching a *giha:d* ('holy war') against this superstition, which they preach through pedantic-sounding exhortations in Classical Arabic in the village and its mosque, a form of Arabic that they deem suitable for the loftiness of their purpose. The peasants listen politely and compliment them on the fine things they say, but completely fail to understand what they are talking about. The language division serves as a metaphor for the complete disparity in the belief systems of the two groups: faith in "progress" and modern science on the one side, superstitious belief on the other. Idris here seems to be asking the elite to climb down from their ivory tower and speak to ordinary people in a language they can understand.[52] The use of language to maintain authority is also satirized in the short story *al-Tamrīn al-'Awwal* (*The First Exercise*, 1956) in which Idris at one point provides an amusing inventory of the insults that the more pedantic and old-fashioned members of the Egyptian teaching profession are in the habit of heaping on the heads of their pupils, starting with the most Classical of Classical expressions, but descending to the insults of the street and alleyway when they are put under stress. This descent has the effect of losing them the spurious air of superiority that they can maintain only as long as they speak in *fuṣha:*.

Idris experimented in other ways with the dramatic exploitation of language difference. In *al-Laḥẓa al-Harija* (*The Critical Moment*, 1956), a play about the Suez invasion, he faced the problem of writing dialogue in Arabic for British as well as Egyptian characters. His solution is to make the British characters talk in *fuṣha:*, whereas the Egyptians speak Egyptian. It cannot be said that this is an altogether convincing ploy dramatically, but it illustrates once again that for Idris *fuṣha:* is the "marked" variety of Arabic that ordinary people—as opposed to the assortment of pedants and outsiders in whose mouths he puts it—do not normally speak on the stage.

From the end of the 1960s, however, a change is noticeable in Idris's language. The other-worldliness, symbolism, and introspection of some of the stories in collections such as *Bayt min Laḥm* (*House of Flesh*, 1971) is matched by a corresponding move in language away from the color and directness of the Egyptian dialect and toward the greater formality and nonlocal values evoked by *fuṣha:*, in dialogue as in narration. Characters become less flesh-and-blood people conversing in naturalistic dialogue than emblematic figures whose ac-

tions are laden with symbolism. However, in one story in this collection, *Snūbizm* (*Snobbery*), Idris once again pokes fun at the donnish airs and linguistic pedantry of the teaching profession. When a professor of anthropology attempts to conduct, in purest *fuṣḥa:*, an on-the-spot investigation into the reasons why a busful of passengers turns against a middle-aged woman who had accused a man on the bus of making lewd advances to her, the passengers, enraged at his long-winded pretentiousness and in a hurry to get on with their business, unceremoniously throw him off the bus.

Whereas the espousal of the dialect as a literary resource by Yusuf Idris in Egypt and others such as Fu'ad al-Takarli in Iraq coincided with the rise of grassroots nationalism in the 1950s and, in Egypt at least, with the adoption of the dialect as a public political idiom, the political fragmentation of the Arab world in the last three decades has been accompanied, seemingly paradoxically, by a move in the opposite direction toward public linguistic homogeneity and standardization. Partly, perhaps, this is a reassertion of cultural unity at a time of political adversity—the downside of "public" dialect use, whether in "serious" literature or elsewhere, is that it looks divisive to some and hence supportive of imperialist designs. But mainly it reflects the much higher literacy levels than existed thirty or forty years ago. Although the regional dialects still thrive, no one would now question the overwhelmingly dominant position of MSA as the language of public life, the media, and serious literature. This is the result of the spread of education, greater urbanization, and the ubiquitousness of the electronic media and the simple homogenized *fuṣḥa:* that they purvey. In consequence, *fuṣḥa:* is today more of a working idiom for a much greater proportion of the population in all countries than it was in the 1950s. The result is that the "language issue" is now a less emotive one in the literary sphere. Writers have felt freer to develop their own patterns of standard/dialect usage without regard for the pressures and pronouncements of councils and committees.

A case in point is provided by the divergent approaches of two leading novelists of the 1980s and 1990s, the Saudi Abdurrahman Munif[53] and the Egyptian leftist Yusuf al-Qa'id. Both have recently written multivolume works set in defined locations at specified points in recent history, yet their approach to the problem of a language for dialogue could not be more different. Munif's five-volume *Mudun al-Milḥ* (*Cities of Salt*, 1983–88) tells the story of a nameless but archetypal Arabian state evolving painfully from illiterate seminomadism to a settled technocracy; al-Qa'id's trilogy *Shakāwā al-Miṣrī al-Faṣīḥ* (*The Complaints of the Eloquent Egyptian*, 1989) is set in Sadat's Cairo at the time of the mid-1970s food riots and recounts the circumstances that force the wretched unnamed "family," the novel's focus, to leave the graveyard it has been living in for want

of an alternative and put itself up for auction in a public square. Both novels chart in detail the social and physical conditions against which the characters act out the story, although the point of view is different. Munif's stance is that of the omniscient third-person narrator, whereas al-Qa'id's is partly third-person, partly first, because one strand in the story is an autobiographical account of the author's own struggle to get his antiestablishment work published in the teeth of opposition from government censors and a politically timid literary establishment.

It might be surmised, on literary as well as political grounds, that Munif would be more likely to use *fuṣḥa:* and al-Qa'id to use dialect. In fact, the reverse is the case: Munif's Saudis of the 1930s from the lowest social rank up to the emir use dialect, whereas al-Qa'id puts a generally undiluted MSA into the mouths of even the most wretched of his cast of low-life characters, as he does in reporting his own direct speech. The different ways the dialogue is handled seem to be connected with differences in the authors' conception of how their works are to be understood and the literary traditions to which they belong. Munif's story is a discursive social and historical document on a Tolstoyan epic scale, with minute observation and attention to characterization, and a mission to describe the trauma of the unfolding events from the point of view of the inhabitants of his imaginary but all-too-recognizable Arabian state. Much of the social and political conflict that occurs is told by means of dialogue, and what the characters say and how they say it is one of the main means Munif uses to differentiate character. This is more easily done by reconstructing the natural rhythms and idioms of actual speech, hence the naturalistic use of dialect, with *fuṣḥa:* being reserved for the narrative framework.[54] Al-Qa'id's trilogy, however, although apparently grounded in recent historical and (apparently) autobiographical fact, is presented via an unnamed and emblematic family that, by selling itself to the highest bidder, symbolizes the bankrupt moral and material state to which ordinary Egyptians have been reduced in the new age of economic laissez-faire. The interweaving of the autobiographical strand of the novel with the story (such as it is) is apparently an experiment in narrative technique. The tone is one of understated moral outrage mixed with biting satire. There is no real plot to speak of: the "story" is a modern-day fable in which dialogue is used sparingly. When it is, it is often less as a means of characterization (there are no flesh-and-blood, only symbolic, characters with symbolic names) than as a way of relieving the experimental narrative by, for example, illustrating the casual corruptness of officialdom and its automatic support for the interests of the wealthy. Dialogue is thus a less important structural element in this work than in Munif's, and it does not have the same function of characterization. The

enigmatic title—*faṣiːḥ* means 'eloquent' but also 'pure' in a linguistic sense, and is from the same root as *fuṣḥaː*—seems to be a reference to the author as a generic, "essential" Egyptian standing out from the crowd and speaking up on behalf of all "actual" Egyptians, which ultimately, perhaps, is the point of the novel.[55] On this abstract, symbolic plane also, the choice of *fuṣḥaː* as a means of rendering speech seems an appropriate linguistic abstraction from reality. The choice of *fuṣḥaː* for dialogue here is in marked contrast to al-Qaʿid's more naturalistic novels, set mainly in the Egyptian countryside, in which he uses a broad rural dialect to render speech. More recently, he has taken the bold step of writing a complete novel in Cairene dialect (*Laban al-ʿAṣfūr*)[56] in the form of a first-person narrative. The story is a modern fable, recounting in a poor woman's own vividly idiomatic words the fate her family suffers after her son finds a bag in the street containing a million pounds. In order to explain the author's unorthodox choice of linguistic code, the publisher's apologetic blurb on the back cover describes the family as "human marginalia" that has at its disposal only "the minimal linguistic lexicon." But here, the author could hardly have used any medium other than the colloquial, if he was not to sacrifice all semblance of verisimilitude.

The oldest generation of novelists, dominated by the Nobel Prize–winning elder statesman of Egyptian letters, Naguib Mahfouz, have been consistent in their linguistic conservatism through to the present day. Like al-Hakim in the theater, Mahfouz, most of whose huge output is as uncompromisingly Egyptian in setting as Idris's, has nonetheless clung to at least the outer morphosyntactic trappings of *fuṣḥaː* in his dialogue. On one occasion he described the use of dialect in literature as a "disease,"[57] even though, on examination, the dialogue in his own later work betrays increasingly an underlying dialectal structure and rhythm—a species of dialect dressed in *fuṣḥaː* clothes. The practical advantage of this approach, as we noted in reference to al-Hakim's theatrical experiments, is that it makes the work more immediately accessible to a wider nonlocal audience and, to the extent that the characters are middle-class and educated, as they are in much of Mahfouz's work, does not require such a suspension of disbelief as when illiterates are required to speak *fuṣḥaː*.

9.5.2 Journalism

Let us now turn briefly to the use of dialect in journalism. As we noted in chapter 8, written dialect in newpapers and magazines is limited to nonserious topics such as sport and fashion, and even here it is only in Egypt that this is at all common. With very few exceptions, politicians who are interviewed at length and required to answer off the cuff do so in the kind of educated dialect we

have been describing in this chapter, but what they say is almost always "translated" into *fuṣḥaː* in a written magazine or newspaper report, even where the original questions and answers are supposedly being reported verbatim. The kind of Arabic that emerges from this process conforms, in the way that we have just described that Mahfouz's characters' dialogue does, to the surface structure of *fuṣḥaː*. Very occasionally, however, dialect *is* deliberately quoted in feature articles to serve a specific purpose. A good example is in an article published in the daily newspaper *al-Sharq al-'Awsaṭ* on 27 August 1992 by Basim al-Jisir, a Lebanese journalist, discussing the upcoming Lebanese general elections. In the article, al-Jisir claims that political life has become highly ideological and divorced from the everyday concerns of ordinary voters. This point is graphically underscored when he momentarily switches from the *fuṣḥaː* of his argument to quote the words of an unnamed (and possibly invented) Lebanese parliamentary deputy from the pre–civil-war period, who is supposed to have said to him:

> *allah yirḥam abaːk, šu biddak min kull halfalsafa, wa leːš ʿam tṣaʿibna bi haššakal? yaʿni biddha kull hannazariːyaːt wa lliff wa ddawaraːn ḥatta yiʿruf ilinsaːn maṣlaḥtu wa tiʿruf ilḥukuːma waːjbaːtha?! xalliːna ʿalbasiːṭ! . . .*
>
> 'For goodness sake, what's the point of all this philosophizing, and why d'you complicate everything? Does it need all this theorizing and tying yourself in a knot for anyone to know what's in his own interest and for the government to know what its job is?! Let's keep things simple! . . . '

The point here is that the form of the words is very much a part of the message—that plain-speaking simple folk can give a lesson in practical politics to airy-fairy theoreticians anytime. The exasperated tone is exactly that of the only kind of newpaper text in which the use of dialect is invariable throughout the Arab world: the handwritten words that emerge from the mouths of cartoon characters commenting satirically on current political and social events.

9.5.3 "Mixed" written style

Rosenbaum (2000) describes a kind of "mixed" written Arabic that is increasingly used in a range of Egyptian writing of the less formal type—"opinion" magazine articles, personal reflections, short stories, and other kinds of first-person narrative in which an Egyptian social setting or political milieu is in the foreground—for example, the first-hand account of a young Egyptian working with a group of others in a London hotel or a discussion in a popular magazine of the Egyptian population explosion. Typically in this style, and unlike the way colloquial Arabic is reserved by certain novelists for speech, writers use

a mixture of MSA and Egyptian Colloquial Arabic (ECA) in the *narrative* sections of the their texts, and there is much seemingly random switching back and forth between the two, with frequent hendiadys (e.g., *za"alni:* (ECA) *wa ʔaḥzanani:* (MSA) 'it upset me and saddened me') and phrasal parallelism, even in the expression of affect, where one might have expected ECA to be the natural code choice (e.g., *wa ya: ʔasfa:h wa ya: xaybata:h* (MSA) 'how sad and disappointing!' followed immediately by *ḥaswa fi 'e:n il-ḥasu:d* (ECA) 'may they be protected from the evil eye!' in a discussion about the population explosion in Egypt[58]). It does not seem to be the case that MSA and the ECA are allocated to specific and separate types of rhetorical or discourse function, as we noted often to be the case in the speeches of Gamal Abdul Nasser. Code switching in this style can occur from word to word, within one sentence, or between sentences and goes in either direction, apparently unpredictably. The mutability of codes seems to be complete: when Rosenbaum rewrote some of his sample texts so that the portions written in MSA and ECA were the exact opposite of the original, the rewritten version was just as acceptable to the Egyptians he asked (provided, of course, that they were not so prescriptivist in outlook that they did not reject *any* "mixed" text).[59] "Mixed" written style of this kind is evidently an intentional choice and can be seen as an attempt by the writer to bring himself/herself closer to the reader by tapping into the "folksy" resonances of everyday speech. It appears to be confined to Egypt and points up once more the different attitude that Egyptians have to their native speech compared with that of other Arab nations.

Notes

1. Ferguson 1971 (repr. from 1959). For an overview and critique of how this notion, as applied to Arabic, has been developed and conceptualized in subsequent western linguistic research, see Eisele 2002.
2. In Ferguson's terms, "strict" concord means that some category that is present in the head or subject noun is copied in the verb or noun modifier, while "deflected" agreement indicates a surface-level mismatch of features, such as the requirement in MSA that non-human plural nouns take feminine singular agreement in adjectives, pronouns, etc. (see Ferguson 1989, 9).
3. Ferguson 1971, 5.
4. Both these claims seem exaggerated. In September 1986, at the opening of the new Sultan Qaboos University in Oman, I witnessed an American university teacher give a fluent, idiomatic fifteen-minute presentation in Damascene Arabic on the organization of English courses in the University Language Center to a group of about 600 Omani freshmen who would be taking these courses. He was obliged to do this because the standard of their English was too low for them to understand a presentation in English, and

he did not feel confident enough of his spoken command of H. The audience reaction was a mixture of amusement and delight: amusement at the incongruity of an American giving a speech in a Syrian dialect to a group of Omanis, and delight at hearing a foreigner able to perform with this level of competence in *any* kind of Arabic, L or H. No doubt if he had delivered his speech in faultless H they would have felt that they had been paid an even bigger cultural compliment, but there was certainly no one in the audience who regarded the speaker as an object of ridicule, as the tremendous applause at the end of his performance testified. The use of H in non-H contexts by native speakers is also attested. The following is a personal communication from Professor Mona Baker, a linguist, translator, and native of Cairo, commenting on a draft version of Holes 1993, a paper that is concerned with explaining level switching in political discourse: "My ex–father-in-law, who is very religious, uses H in everyday contexts—not all the time, but very often. I've noticed that a number of 'religious' men (don't know of any women) do the same in Egypt. It sounds funny to ordinary people like myself but I do recognize it as an expression of 'Muslim' identity even though the context in which it is used often has nothing to do with religion." In some cases, the deliberate use of H in non-H contexts is a conscious effort to "set an example," based on the belief that it would be better for all Arabs if they abandoned dialect and spoke only pure H. A case in point is the Egyptian poet Farouq Shusha, who is in the habit of using H even in relaxed conversation.

5. See Harrell 1960 for variation in Egyptian radio news broadcasts; Diem 1974 (esp. 67–87) on variation in broadcast monologues and interviews of various types from various radio stations; El-Hassan 1977 on variation in mosque sermons; Meiseles 1979 on the substandard variable MSA of personal letters; Holes 1993 and Mazraani 1997 on level switching in political speeches; and Rosenbaum 2000 (see 9.5.3) on the "mixed" written style of Egyptian magazine articles. In a 2002 article, Al-Batal shows how LBCI (Lebanese Broadcasting Corporation International), a privately owned television company that developed out of the Christian-supporting LBC founded during the Lebanese civil war, now routinely uses a distinctly and distinctive "Lebanized" MSA to report on local Lebanese affairs in its news bulletins, whereas it uses normative MSA to report non-local items of news.
6. Cross-dialectal variation and leveling is analyzed in Blanc 1960 for speakers from the eastern Arab world on the basis of one long text and similarly by Talmoudi 1984 for a group of North African speakers. Particular aspects of mixed, "semiformal" Arabic have been extensively investigated by (for example) Mitchell 1978, 1986; El-Hassan 1977, 1978; Sallam 1979, 1980; Meiseles 1980; and Mitchell and El-Hassan 1994.
7. Badawi 1973 (Cairo); El-Hassan 1978 (Cairo and the Levant); Holes 1986a and 1987 (Bahrain); and Owens and Bani-Yasin 1987 (Jordan) have experimented with various scaling devices and graphs to display ranges of individual grammars or parts of grammars that coexist within a community.
8. See Parkinson 1991, 1993.
9. Parkinson (1991, 61) states at the conclusion of his report on his experimental investigations into Egyptians' attitudes to *fuṣḥa:*, dialectal, and "mixed" texts: "there is a wide range of acceptable oral performance styles for a text whose written form would be ac-

cepted as *fuṣḥa:*, and that while increasing 'authenticity' or 'classicisation' of phonological variables increases acceptance of text, completely full vowelling and proper observation of pause form, and the like, can actually decrease it, making a 'middle' solution, with partial vowelling but correct (Egyptian) phonology, a clear preference over other varieties."

10. Badawi 1973.

11. However, "semiformal" Arabic (in Badawi's model, Level 3) seems to be a more stable and recognized variety in Cairo and the cities of the Levant than in areas of the Arab world that have only recently started to undergo the social changes that go with the general diffusion of education and literacy through society. In the latter areas (e.g., the Gulf) a "semiformal" code is still in the process of formation, with the result that in many of the contexts where a group of educated Egyptians would tend to use a semiformal variety (a television discussion on a serious subject such as health care), Gulf Arabs would tend to attempt more consciously to speak an approximation of MSA (in Badawi's terms, Level 2).

12. *mutanawwir* literally means 'enlightened', that is, having received a modicum of primary education, as opposed to *muθaqqaf*, which indicates an education to at least high school graduation standard.

13. Badawi 1973, 166–74.

14. Ibid., 207–13.

15. For a detailed analysis of six extracts from Nasser's speeches covering the period 1956 to 1965, see Holes 1993, from which the extracts in this section have been taken.

16. The commonly held Arab belief in the inherent beauty and superiority of Classical Arabic is aptly summed up by Ferguson (1968, 376):

> First of all, Arabs feel that their language is beautiful, making its most direct appeal in the recitation of classical poetry and in formal or semi-formal oratory. There is no mistaking the emotional involvement of both the reciter or the speaker and the audience on these occasions. It has been pointed out that Arabs often respond more deeply to the rhythms and phonetic symbolism of the classical language under these conditions than to the semantic content of the poem or speech . . . For many purposes even the illiterate peasant will prefer a classical-sounding highly literary Arabic which he only half understands to a pure conversational Arabic which he understands perfectly.

17. It is no coincidence that, at the same time as Nasser was "democratizing" Arabic political rhetoric by making greater use of the idiom of everyday life, left-wing writers and supporters of the Egyptian revolution such as Yusuf Idris were doing something similar in the short story and drama.

18. *Collected Speeches* 1:547.

19. Ferguson 1971, 5.

20. El-Hassan 1977, 114 provides an excellent illustration of functionally motivated level switching in a mosque sermon recorded in Upper Egypt. The "text" of the sermon is delivered in pure *fuṣḥa:* (Badawi's Level 1), but the imam constantly interrupts himself

to provide explanations, comments, asides, and "translations" in the local colloquial (Level 5) in order to ensure that all of the audience understand the "text." For example:

"text" (Level 1)
samma:hu nnabi:yu ṣalla lla:hu 'alayhi wa sallam 'abdulla:h, kama: ka:na yutlaqu 'alayhi 'ati:q . . .
'The prophet, peace be upon him, called him 'Abdullah. He was also called 'Atiq . . . '
↓
"comment" (Level 5)
le:h
'Why?'
↓
"text" (Level 1)
fa qi:la li ḥusni wajhih
'It was said that it was because he was good-looking'
↓
"translation" (Level 5)
wiššu ka:n ḥilu kida
'He had a handsome face, kind of '

A similar type of switching is described for mosque sermons in Mauritania in Taine-Cheikh 2002.

21. The linguistic behavior of the speakers in this imaginary conversation is based on an analysis of tape recordings of large numbers of actual conversations made in Bahrain over an eight-month period of field work in 1977–78.

22. Note that this statement applies to *educated* speakers talking to other educated speakers. There is a range of dialects in Bahrain as in any other place—in Bahrain's case correlating closely with historical origin/confessional allegiance—and there are dialectal features that educated speakers would be likely to use with family or kin but would be likely to zero out in the circumstances described here. For example, many educated Sunni speakers would use an affricate /j/ instead of /g/ in some words when speaking "off the record" with friends and family, and educated Shi'i speakers, particularly if from village backgrounds, would likewise use /k/ instead of /g/ when, say, talking to an uneducated mother or relative. Such variants are recognized by Bahrainis as markers of particular social identities within Bahraini society but are normally suppressed by educated speakers in contexts where such social identities are not a relevant factor (but see n. 23 below).

23. The importance of this latter point was graphically brought home to me when gathering conversational data in Bahrain in the 1970s. Two young men of the same age and from the same sect, social background, and neighborhood, and both in the same year at the then Bahrain Men's Teacher Training College, were separately interviewed in the same physical context and asked the same kinds of questions on their likes and dislikes, hobbies, interests, and ambitions. I was initially puzzled by the fact that the first of the two interviewees consistently used dialectal forms that educated speakers would normally avoid in the formal atmosphere engendered by tape recording (and with a foreigner to boot). In fact, his speech sounded virtually indistinguishable from that typical of the gen-

eration of his illiterate grandfather. The other speaker spoke quite differently, not merely avoiding completely the stigmatized forms his classmate had used so unselfconsciously, but using a mixture of nonlocal pan-Arab forms and dialectal features associated with the locally prestigious social group to which he did not belong. It became clear in the course of the conversation that the first speaker spent all his time when not at his studies feeding, cleaning, riding, and generally looking after horses. Moreover, he associated exclusively with the ordinary (i.e., illiterate) older men who earned their livelihoods in farming and animal husbandry. He was extremely proud of the prowess he had developed in horsemanship and the knowledge of its lore he had gained from the stablemen. The other speaker, it turned out, had quite opposite and wholly nonlocal interests—cars, electronic consumer goods, and western pop music—and he spoke disparagingly and at length about the "old-fashioned" ideas of his parents' generation.

There seems in this case to be the probability of a connection between language choice and self-image. That is, it is not just the different speech networks that each speaker's interests caused him to participate in regularly that were affecting his speech, but also the diametrically opposed attitudes of each toward the traditional values of the social group from which they both came—supportive in the first case and dismissive in the second. It is possible that the projection of attitude is also the explanation of the unpredictable use of a highly classicized Arabic by some devout Muslims in nonreligious contexts reported in note 4 above: by using such language they are averring or emphasizing the Muslim aspect of their identity.

24. I am grateful to the late Jasim Sharidah of Bahrain Radio for the tape recordings on which these observations are based.

25. Parkinson 1985, 58–59 makes a similar point about the use of teknonyms in modern Egypt.

26. I am grateful to David Barber for allowing me access to some portions of this material.

27. Extract (1) is from tape 11, side 2, lines 12–14 of the accompanying transcription; (2) is from tape 23, side 1, lines 6–11; and (3) is also from tape 23, side 1, lines 83–86.

28. Blanc 1960.

29. Talmoudi 1984.

30. Ezzat 1974.

31. El-Hassan 1977, 1978; Meiseles 1980; Mitchell 1978, 1986; Sallam 1979, 1980; and Mitchell and El-Hassan 1994. A recent comparative study of the syntax of four widely dispersed geographical dialects—Moroccan, Egyptian (= Cairene), Syrian (= Damascene), and Kuwaiti—confirms the observation of Mitchell and El-Hassan (1994, 2) that regional dialect differences are lexical and phonological before they are grammatical (Brustad 2000, 4). The syntactic and morphosyntactic substructures are largely shared, and it is the relatively more superficial elements—the lexical material that fills the syntagms and the phonological rules for performing them—that differ. This would explain why, all things being equal, shifts by speakers during the course of unscripted speech to conspicuously MSA, nondialectal syntactic structures and morphosyntactic forms tend to be seen by others as more significant stylistically than lexical or phonological switches.

32. El-Hassan notes in his detailed cross-dialectal analysis of variation in the use of demonstratives: "This (= statistics of the use of Egyptian versus non-Egyptian forms) suggests

that educated Egyptians, whether talking to other educated Egyptians or to non-Egyptians tend to use far more stigmatised than prestigious demonstrative tokens. Moreover, in situations where the other collocutors are non-Egyptian, the tendency is for educated Egyptians to use stigmatised forms more frequently than in situations where all the speakers are Egyptian" (1978, 42). El-Hassan explains these findings as evidence of "the desire to sound like an Egyptian" in cross-dialectal conversations (i.e., the assertion of national loyalty) against the background of a situation where "Egyptian Arabic has spread far and wide through the mass media." The asymmetrical, even unilateral nature of cross-dialectal accommodation and convergence is also brought out clearly in a study of workplace-based conversational transcripts between Tunisians on the one hand and a mixture of Lebanese, Syrians, Palestinians, and Egyptians on the other (S'hiri 2002), whom the Tunisians generically refer to as "Sharqis" ("Easterners"). It is not merely for reasons of comprehensibility and efficient communication in the work place that Tunisians accommodate to the Easterners, and not vice versa. By doing so they show the friendliness, savoir-faire, and aptitude for language learning that Tunisians think of as national characteristics and seek to gain social approval by publicly identifying with the speech patterns of Easterners, the culturally and politically dominant group in the Arabic speaking world (S'hiri 2002, 169–73).

33. See Cadora 1970 on this kind of switching as it was occurring in the 1960s among the younger and more educated generations of extended families who had recently moved to the (then) predominantly Christian town of Ramallah, Palestine.

34. A case in point is the speech habits of some political figures, which are cultivated as part of their public image. The late King Hussein of Jordan, for example, rarely used anything but pure *fuṣḥa:* on television and radio, whether he was speaking from a script or was being interviewed. He was widely admired by Arabs from many countries, not just Jordanians, for this practice, which is often cited as evidence of his learning, dignity, and purity as an Arab: hoardings with his picture on in 1980s' Amman played up this image and bore the legend *Ṣaqr Quraysh* ('The Falcon of Quraysh')—a reference to his claimed descent from the Prophet's tribe. The record shows that Nasser was also perfectly capable of speaking correct *fuṣḥa:* when the occasion demanded, but, as we have shown, he frequently made a point of using dialect to cultivate a "man of the people" image. In each case language choice seems to have been governed at least to some extent by public relations and political ends.

35. El-Hassan 1977 *passim* for a critique of the five-level models of Blanc and Badawi, especially 125ff.

36. Parkinson 1991.

37. An interesting and amusing example of this lexical "pecking order" was provided by King Fahad of Saudi Arabia, interviewed on Bahrain television in December 1986 at the opening of the elevated causeway linking Bahrain to the eastern province of his country. Having cut the ribbon and driven across the causeway that his government had paid for, the king emerged from his limousine to be asked for an off-the-cuff reaction to the realization of a plan mooted for more than twenty years and now finally achieved. Visibly enthusiastic about what he had seen, he responded:

šufna (pause) *wa raʔayna* (pause) *wa* (pause) *laːḥazna* . . .
    'We have seen . . . and we have seen . . . and . . . we have observed'

All of these verbs are being used to mean 'we have seen'. The first is plain dialect, the second is MSA, and the last is lexically "high-flown" MSA, at least when used, as here, to describe the mere physical act of 'seeing' a bridge (rather than, for example, 'making an observation/remark', which is its normal MSA use). Having initially reacted to the interviewer's question in his unguarded natural speech, the king clearly felt that the solemnity of the occasion demanded a more formal linguistic treatment, which he achieved by the use of these stylistically ascending synonyms. Having found the appropriate level with *laːḥazna*, the king continued speaking in high-level MSA.

38. Leeds University Corpus of ESA tape 11, side 2, lines 58–73 of the transcription.
39. This example is based on Mitchell 1986, 23–24.
40. N.B. that the final *-u* of the Egyptian speaker's *byastaṭiːʿu* 'they are able' in Example (x) is a *dialectal* marker of plurality, whereas the final *-u* of *yastaṭiːʿu* 'he is able' is an MSA marker of indicative mood, the form itself being singular.
41. Less so because speakers who feel forced by context to attempt to speak in a very high-flown MSA sometimes begin successfully but then, as the sentence develops, begin to "miss the target" because of insufficient knowledge of the rules of MSA and/or dialect interference.
42. In either case, *wasaːʔil taʕliːmiːya* 'educational aids' is a frozen MSA phrase like those we illustrated in chapter 8 that is not susceptible to variation and thus contributes little to the stylistic tone.
43. Mitchell 1986, 24–25.
44. Ibid., 25.
45. El-Hassan 1977, 122, 124.
46. Badawi 1973, 140–41.
47. It seems odd that al-Hakim chose to experiment with his nondialectal "third language" in *The Deal*, a credible and exciting story set in the Egyptian countryside, yet reserved his toned-down version of the Egyptian dialect for *The Dilemma*, a play that does not depend on any particular social or geographical context and is largely a vehicle for al-Hakim's view that scientific research cannot be pursued in isolation from ethical considerations (Badawi 1987, 81). One would have thought that the linguistic experiment would have been more natural if conducted the other way around.
48. Idris stated this in an Egyptian television commentary on his life made in 1986.
49. When in 1960 the Higher Council for Arts and Literature decided to exclude from a competition plays written in the vernacular, forcing the well-known dramatist Nuʿman ʿAshur to withdraw, Idris wrote in the Egyptian daily newspaper *al-Gumhūrīya:*

> Let the Council with its venerable committees, meet and disperse and report that the literary movement has stagnated and died down. Let them report that the present generation is corrupt and that the only remaining good writers are those who are past seventy. Let them linger upon the colloquial and the classical, the rules of metre and rhyme. Let them alone be the men of letters, and let them speak in literature's name. Let them put on their sombre clothes, let them assume those serious dignified airs,

> their celebrations and prizes. Even let them install themselves as the guardians of language, and may they continue to act as such. But may you [Nuʿman ʿAshur] keep on producing and writing and may your work last! (quoted in Kurpershoek 1981, 116).

50. In an article in *al-Gumhūrīya* (18 June 1966) titled *Al-balāgha al-siyāsīya taṭawwarat* ('Political rhetoric has moved on'), Idris argued that the revolution of 1952, which advanced the lot of the peasants and lower classes, required a new, more demotic idiom between leader and led, in contrast to the flowery eloquence of the leaders of the last Egyptian revolution of 1919, who had been highly educated members of a middle-class elite.
51. Idris 1981, 101.
52. Kurpershoek 1981, 122–23.
53. Munif was stripped of his Saudi citizenship because of his literary activities and moved to Jordan. He died in Damascus in 2004.
54. In a more recent (1999) and similarly epic-scale work on the political history of early nineteenth-century Iraq, *'Ard al-Sawād,* Munif again makes heavy use of a dialect, in this case Baghdadi, to give vividness and color to the dialogue.
55. The title is a double literary allusion. On the one hand, it echoes the title of a short story by al-Qaʿid's contemporary Gamal al-Ghitani *Shakāwā al-Jundī al-Faṣīḥ* (*Complaints of the Eloquent Private Soldier*), published in the literary journal *al-Hilāl* in August 1971. This piece tells the story, largely in the first person, of a young conscript posted to a remote corner of Egypt who finds that he has been dismissed from his civilian job. The general social and political tenor is similar to al-Qaʿid's full-length work. (I am indebted to Paul Starkey for this observation.) Ultimately, however, both al-Qaʿid's and al-Ghitani's titles hark back to an ancient Egyptian Middle Kingdom fable "The Eloquent Peasant," which is "a serious disquisition on the need for justice, and a parable on the utility of fine speech" in which an Egyptian peasant presents complaints to a magistrate, but in fine language. The magistrate is so pleased with the peasant's eloquence that he tells the king, who forces the magistrate to goad the peasant to produce ever-finer language. The peasant ultimately receives a just reward. (Lichtheim 1975, 169).
56. Literally 'Sparrow's Milk', an Egyptian colloquial proverbial expression for something impossible or unattainable.
57. Somekh 1991, 27.
58. Rosenbaum 2000, 73.
59. Ibid., 78.

Appendix: The Arabic Script

Arabic is written and read from right to left. The script is cursive, that is, the letters are joined to each other, even in print. There are twenty-eight separate letters (see the chart below), although not that many separate shapes, because some letters are distinguished from each other only by the presence/absence, position or number of superscript or subscript dots. The shape of a letter changes slightly depending on whether it is in initial, medial, or final position in the word. Six letters of the twenty-eight do not join onto a following letter.

CHART OF ARABIC LETTERS (*ḥuru:fu lhija:ʔ*)

name of letter	unconnected	with preceding letter	with following letter	with both	symbol in transcription
ʔalif	ا	ا	...	...	: (vowel length)
ba:ʔ	ب	ب	ـ	ـ	b
ta:ʔ	ت	ت	ـ	ـ	t
θa:ʔ	ث	ث	ـ	ـ	θ
ji:m	ج	ج	ـ	ـ	j
ḥa:ʔ	ح	ح	ـ	ـ	ḥ
xa:ʔ	خ	خ	ـ	ـ	x
da:l	د	ـ	...	...	d
ða:l	ذ	ـ	...	...	ð
ra:ʔ	ر	ـ	...	...	r
za:ʔ	ز	ـ	...	...	z
si:n	س	س	ـ	ـ	s

| name of letter | SHAPE | | | | symbol in transcription |
	unconnected	with preceding letter	with following letter	with both	
ši:n	ش	ش	شـ	ﺸ	š
ṣa:d	ص	ص	صـ	ﺼ	ṣ
ḍa:d	ض	ض	ضـ	ﻀ	ḍ
ṭa:ʔ	ط	ط	ط	ﻄ	ṭ
ða:ʔ	ظ	ظ	ظ	ﻈ	ḏ
ʻayn	ع	ع	عـ	ﻌ	ʻ
ġayn	غ	غ	غـ	ﻐ	ġ
fa:ʔ	ف	ف	ف	ﻔ	f
qa:f	ق	ق	ق	ﻘ	q
ka:f	ك	ك	ﻛ	ﻜ	k
la:m	ل	ل	ل	ﻠ	l
mi:m	م	م	مـ	ﻤ	m
nu:n	ن	ن	نـ	ﻨ	n
ha:ʔ	ه	ﺔ	هـ	ﻬ	h
wa:w	و	و	...	...	w or : (vowel length)
ya:ʔ	ي	ي	يـ	ﻴ	y or : (vowel length)

Apart from the standard letters there are a number of other letters that are sometimes used to represent sounds that Modern Standard Arabic does not have and that are used in some parts of the Arab world to represent dialectal Arabic sounds or borrowings from other languages. In the Gulf, for example, one sometimes sees:

پ = /p/   چ = /č/   گ = /g/

Orthographic conventions

The written Arabic word consists of consonants and long vowels only. Long /a:/ is marked by an *ʔalif,* /u:/ by a *wa:w,* and /i:/ by a *ya:ʔ.* These signs of length are written after the short vowel that they lengthen. The short vowels themselves, which are of course phonemic, are written above (/a/ or /u/) or below (/i/) the consonant they follow. However, it is not normal practice to mark short vowels at all, except in the text of the Koran, where they are obligatory; other than this, they appear only in pedagogical material, such as grammars of MSA and children's reading primers. Here are some examples of voweled words:

كَتَبَ   عَلِمَ   كَبُرَ

kataba   *ʻalima*   *kabura*

'he wrote'   'he knew'   'he grew up'

كِتَاب   عُلُوم   كَبِير

kita:b   *ʻulu:m*   *kabi:r*

'book'   'sciences'   'big'

As well as the short vowels, there are two other orthographic marks written above the word that serve a phonological function: *suku:n* 'pause', which marks the closure of a syllable, and *šadda* 'strengthening', which marks consonantal length. Like the short vowels, neither is usually marked in unvoweled writing. For example:

suku:n   *šadda*

صَبْر   عَلَّم

ṣabr   *ʻallama*

'patience'   'he taught'

There are one or two very common words in which the *ʔalif* that marks long /a:/ is not usually written in the body of the word but (if it is written at all) above the line, as if it were a short vowel, for example, in *ʔalla:h* 'God' and the first /a:/ in *ha:ða:* 'this'.

As well as its use as a sign of vowel length, *ʔalif* has several other orthographic uses. Along with *wa:w* and *ya:ʔ* (which, unlike *ʔalif,* may have an inde-

pendent consonantal value (/w/ and /y/ respectively)), it is used as a carrier (Ar. *kursi:y* 'chair') of the glottal stop /ʔ/, *hamza*, which, although a fully functioning consonant, is not written in the body of the word as other consonants, but above or below it. There are historical reasons for this anomaly (see 2.4.1.1). When *ya:ʔ* acts as a *kursi:y*, it has no dots. For example:

carrier *ʔalif* with *hamza* written above:	carrier *ʔalif* with *hamza* written below:
رَأْس	إِرْسَال
raʔs	*ʔirsa:l*
'head'	'sending'
carrier *wa:w*:	carrier *ya:ʔ*:
سُؤَال	رَئِيس
suʔa:l	*raʔi:s*
'question'	'president'

A prosthetic *hamza* that sits on an *ʔalif* and is used in order to avoid an unacceptable postpausal CC syllabic onset when /l/, the definite article, occurs utterance-initially (أَلْ = *ʔal*). This *hamza*, which has a purely phonological function, is not normally marked in unvoweled script (see 2.4.1.1). The definite article itself is sometimes pronounced as /l/, and sometimes via assimilation to a following coronal consonant (= dentals, alveolars, and alveopalatals except /j/) as consonantal length. In either case, however, it is written as *la:m*:

following consonant = noncoronal	= coronal
أَلْبَيْت	أَلشَّمس
ʔal-bayt	*ʔaš-šams*
'the house'	'the sun'

Other than when it occurs in utterance-initial position, the definite article has no glottal onset (called in Arabic *hamzatu lqatʕ* 'the cutting *hamza*'), although the *ʔalif* on which it sits is nonetheless written in all but a few fixed syntactic contexts (such as after the preposition *li*). In fully voweled texts, a special or-

thographic sign (*hamzatu lwaṣl* 'the joining *hamza*'), which indicates its elision, is written above the *ʔalif*. For example:

فِي ٱلْبَيْتِ

fi: l-bayti

'in the house'

There is one other orthographic anomaly that, like the writing of *hamza*, has its roots in ancient dialectal differences. In certain types of form, a final *ya:ʔ* after /a/ was pronounced as a glide in the Hejaz but was treated as if it were a vowel-lengthening *ʔalif* in eastern Arabia. Modern spelling retains the final spelling *ya:ʔ* (= /y/) thus reflecting the ancient Hejazi pronunciation, but modern pronunciation everywhere has "eastern" /a:/. This final vowel-lengthening *ya:* is written without its dots and is called (along with all other forms of final /a:/) *ʔalif maqṣu:ra* 'shortenable *ʔalif*', so called because its vowel is pronounced short when it occurs in a closed syllable as a result of juncture. For example:

رَمَى

rama:

'he threw'

Special conventions exist for marking certain morphosyntactic phenomena in the script:

(a) Case inflection

The so-called *tanwi:n* ('adding an "n"'), which indicates indefiniteness in certain categories of nouns (2.4.1.3), is marked by a set of special superscript signs, basically a doubling of the corresponding short vowel signs. The sign for the accusative *-an* is written on top of an *ʔalif* (although not if the noun ends in *ta:ʔ marbu:ṭa* 'tied "t"' (see (c) below). With the exception of this *ʔalif*, which cannot be omitted from the spelling of any word in which it occurs, the other signs, like the short vowels, are not normally written. For example:

بَيْتٌ	بَيْتاً	بَيْتٍ
bayt-un	*bayt-an*	*bayt-in*
'a house (NOM)'	'a house (ACC)'	'a house (GEN)'

(b) Redundant *ʔalif*

In certain verb forms that would otherwise end in *wa:w*, a redundant *ʔalif* that serves simply to mark the end of the word is written (see 2.4.1.7). For example:

كَتَبُوا   لَمْ يَكْتُبُوا   أُكْتُبُوا

katabu: *lam yaktubu:*   *ʔuktubu:*

'they(m) wrote'   'they(m) didn't write'   'write (pl)!'

(c) Feminine noun suffix *-a*

A special letter, *ha:ʔ* with two dots above it, marks the feminine noun suffix. Historically, such feminine nouns were sometimes pronounced with final -/*ah*/, sometimes -/*at*/, hence the spelling that combines features of both *ha:ʔ* and *ta:ʔ*. In today's MSA, this sign, the so-called *ta:ʔ marbu:ta* ('tied "t"') is pronounced -/*a*/ in pause, /*t*/ in juncture. For example:

مَدْرَسَةٌ   مَدْرَسَةُ ٱلْبَنَاتِ

madrasa   *madrasatu l-bana:ti*

'school'   'the girls' school'

Script styles

The commonest handwriting style is known as *ruqʿa*. In newspaper headlines, on advertising hoardings, and in "urban calligraphy" in general one encounters a wide variety of other styles. Some of the most ornate examples are seen in mosques and religious buildings. The phrase *bismi lla:hi rrahma:ni rrahi:m* 'In the name of God the Beneficent, the Merciful' is shown here in five of the commonest styles:

ruqʿa   بسم الله الرّحمن الرّحيم

θuluθ   بسم الله الرّحمن الرّحيم

nasx   بسم الله الرّحمن الرحيم

di:wa:ni: بسم الله الرّحمن الرّحيم

ku:fi: بسم الله الرحمن الرحيم
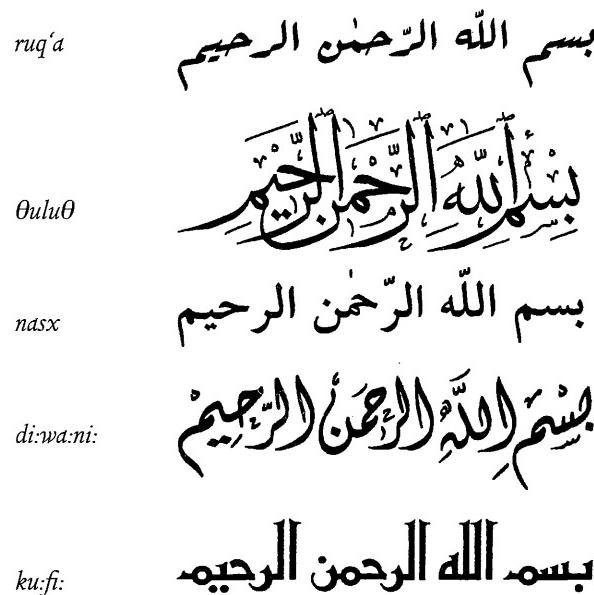
Bibliography

Journal abbreviations

AB	*al-Abhāth*
AL	*Anthropological Linguistics*
ALA	*al-ʿArabiyya*
AO	*Acta Orientalia*
AR	*Arabica*
ARLI	*Archivum Linguisticum*
BO	*Bibliotheca Orientalis*
BSOAS	*Bulletin of the School of Oriental and African Studies*
EI	*Encyclopaedia of Islam*
IJSL	*International Journal of the Sociology of Language*
IOS	*Israel Oriental Studies*
IS	*Islam*
JL	*Journal of Linguistics*
JNES	*Journal of Near Eastern Studies*
JPCL	*Journal of Pidgin and Creole Languages*
JSS	*Journal of Semitic Studies*
LAN	*Language*
LING	*Linguistics*
LNG	*Lingua*
LS	*Language in Society*
MAS-GELLAS	*Matériaux Arabes et Sudarabiques de la groupe d'études de linguistiques et littératures Arabes et Sudarabiques*
MEJ	*Middle East Journal*
MSOS	*Mitteilungen des Seminars für Orientalische Sprachen zu Berlin*
MUSJ	*Mélanges de l'Université Saint Joseph (Beirut)*
NAS	*New Arabian Studies*
SI	*Studia Islamica*
SLS	*Studies in the Linguistic Sciences*
TPS	*Transactions of the Philological Society*
WO	*Word*
WZKM	*Wiener Zeitschrift für die Kunde des Morgenlandes*
ZAL	*Zeitschrift für arabische Linguistik*
ZDMG	*Zeitschrift der deutschen morgenländischer Gesellschaft*

Abboud, P. F. 1964. *The syntax of Najdi Arabic*, Ann Arbor, University Microfilms International.

Abdel-Jawwad, H. 1986. The emergence of an urban dialect in the Jordanian urban centers. *IJSL* 61:53–63.

Abdel-Jawwad, H., and M. Awwad. 1987. Reflexes of Classical Arabic interdentals: A study in historical sociolinguistics. *AB* 35:75–102.

Abu Haidar, F. 1988. Male/female linguistic variation in a Baghdad community. In *A miscellany of Middle Eastern articles: In memoriam Thomas Muir Johnstone (1924–83)*, 151–62. Harlow, UK: Longman.

———. 2002. Arabic and English in conflict: Iraqis in the UK. In *Language contact and language conflict in Arabic*, ed. A. Rouchdy, 286–96. London: RoutledgeCurzon Press.

Abun-Nasr, J. M. 1975. *A history of the Maghreb*. Cambridge: Cambridge University Press.

Agius, D. A. 1991. Precedence of VOS over VSO in Modern Standard Arabic. In *Semitic studies in honor of Wolf Leslau,* vol. I, ed. A. Kaye, 39–55. Wiesbaden, Harrassowitz.

Agius, D. A., and A. Harrak. 1987. Auxiliary particles preceding the imperfective aspect in Arabic dialects. *AR* 34:164–80.

Al-Batal, M. 2002. Identity and language tension in Lebanon: The Arabic of local news at LBCI. In *Language contact and language conflict in Arabic*, ed. A. Rouchdy, 91–115. London: RoutledgeCurzon.

Al-Ḥuṣarī, S. 1985. *Fī l-Lugha wa l-'Adab wa 'Alāqātuhumā bi l-Qawmīya (On language and literature and their relationship to nationalism)*. Beirut: Markaz Dirāsāt al-Waḥda al-'Arabīya.

Al-Khoury, C. 1956. *Qiṣṣat Finiānos (The story of Finianos)*. Beirut: Imprimerie Catholique.

Al-Nafīsī, A. 1978. *Al-Kuwayt: al-ra'y al-'ākhar (Kuwait: The alternative view)*. London: Tā-Hā Publishing.

———. 1982. *Majlis al-Ta'āwun al-Khalījī (The Gulf Co-operation Council)*. London: Tā-Hā Publishing.

Al-Nassir, A. A. 1993. *Sībawayh the phonologist*. London: KPI.

Al-Shaykh Durra, I. I. 1988. *Iqtiṣādīyāt al-'Iskān (Housing economics)*. Kuwait: 'Ālam al-Ma'rifa.

Al-Toma, S. J. 1974. Language education in Arab countries and the role of the academies. In *Advances in Language Planning*, ed. J. Fishman, 279–313. The Hague: Mouton.

Al-Wer, E. 2002. Education as a speaker variable. In *Language contact and language conflict in Arabic*, ed. A. Rouchdy, 41–53. London: RoutledgeCurzon Press.

Al-Yāzijī, Sheikh I. 1901. *Lughat al-Jarā'id (The language of newspapers)*. Cairo: al-Maktaba al-Maḥmūdīya al-Tijārīya.

Ali, A-S. M. 1987. *A linguistic study of the development of scientific vocabulary in standard Arabic*. London: KPI.

Angoujard, J.-P. 1990. *Metrical structure of Arabic*. Dordrecht: Foris.

Arberry, A. J. 1957. *The seven odes*. London: George Allen & Unwin.

Ayalon, A. 1987. *Language and change in the Arab Middle East: The evolution of modern political discourse*. Oxford: Oxford University Press.

Badawi, El-S. 1973. *Mustawayāt al-'Arabīya al-Mu'āṣira fī Miṣr (Levels of contemporary Arabic in Egypt)*. Cairo: Dār al-Ma'ārif.

Badawi, El-S., M. G. Carter, and A. Gully. 2004. *Modern Arabic: A comprehensive grammar*. London and New York: Routledge.

Badawi, M. M. 1987. *Modern Arabic drama in Egypt*. Cambridge: Cambridge University Press.

Barlow, M. 1988. A situated theory of agreement. Ph.D. diss. Stanford: Stanford University.

Bauer, L. 1926. *Das palästinische Arabisch (Palestinian Arabic)*. Leipzig: Hinrichs'sche Buchhandlung.

Beeston, A.F.L. 1968. *Written Arabic: An approach to the basic structures*. Cambridge: Cambridge University Press.

———. 1970. *The Arabic language today*. London: Hutchinson.

———. 1973. Co-ordination in literary Arabic. *MUSJ* 48:167–74.

———. 1983. The role of parallelism in Arabic prose. In *Arabic literature to the end of the Umayyad period*, Beeston et al., 180–85. Cambridge: Cambridge University Press.

Belnap, R. K. 1993. The meaning of deflected/strict agreement variation in Cairene Arabic. In *Perspectives on Arabic linguistics V*, ed. M. Eid and C. D. Holes, 97–117. Amsterdam: Benjamins.

Belnap, R. K., and O. Shabaneh. 1992. Variable agreement and non-human plurals in Classical and Modern Standard Arabic. In *Perspectives on Arabic linguistics IV*, ed. E. Broselow, M. Eid, J. McCarthy, 245–62. Amsterdam: Benjamins.

Bensallama, Al-B. 1971. *Al-Lugha l-'Arabīya wa Mashākil al-Kitāba (The Arabic language and problems of writing)*. Tunis: Dār al-Tūnisīya li l-Nashr.

Blanc, H. 1957. Hebrew in Israel: Trends and problems. *MEJ* 11:397–409.

———. 1960. Style variations in Arabic: A sample of interdialectal conversation. In *Contributions to Arabic linguistics*, ed. C. Ferguson, 81–156. Cambridge, Mass.: Harvard University Press.

———. 1964. *Communal dialects in Baghdad*. Cambridge, Mass.: Harvard University Press.

———. 1970 Dual and pseudo-dual in the Arabic dialects. *LAN* 46:1, 42–57.

Blau, J. 1960. *Syntax des Palästinischen Bauerndialekts von Bir-Zeit (Syntax of the Palestinian rural dialect of Bir-Zeit)*. Walldorf-Hessen: VFO.

———. 1966–67. *A grammar of Christian Arabic*. Louvain.

———. 1973. Remarks on some syntactic trends in Modern Standard Arabic. *IOS* III:172–231.

———. 1976. Some additional observations on syntactic trends in Modern Standard Arabic. *IOS* VI:158–90.

———. 1981a. *The emergence and linguistic background of Judaeo-Arabic*. Jerusalem: Ben-Zvi Institute.

———. 1981b. *The renaissance of Modern Hebrew and Modern Standard Arabic*. Los Angeles: University of California Press.

———. 1988. *Studies in Middle Arabic and its Judaeo-Arabic variety*. Jerusalem: Magnes Press.

Bloch, A. 1986. *Studies in Arabic syntax and semantics*. Wiesbaden: Harrassowitz.

Blohm, P. 1990. Gibt es im Arabischen Aktionsarten? (Is there grammatical aspect in Arabic?) *ZAL* 22:7–22.

Bobzin, H. 1980. Zur Häufigkeit von Verben in Neuhocharabischen (On the frequency of verbs in Modern Standard Arabic). *ZAL* 5:35–69.

Boucherit, A. 2002. Algérie: de l'arabe à l'arabisation (Algeria: From Arabic to arabization). In *Language Contact and Language Conflict in Arabic*, ed. A. Rouchdy, 54–69. Richmond, UK: RoutledgeCurzon.

Boumans, L., and J. J. de Ruiter. 2002. Moroccan Arabic in the European diaspora. In *Language contact and language conflict in Arabic*, ed. A. Rouchdy, 259–85. London: RoutledgeCurzon Press.

Bravmann, M. M. 1977. *Studies in Semitic philology*. Leiden: Brill.

Britti, A. 1980. A history of right dislocation in certain Levantine Arabic dialects. *SLS* 10, no. 2:121–39.

Brockelmann, C. 1908–13. *Grundriss der vergleichenden Grammatik der semitischen Sprachen (Sketch of the comparative grammar of the Semitic languages)*. Berlin: Reuter & Reichard.

Broselow, E., M. Eid, J. McCarthy, eds. 1992. *Perspectives on Arabic linguistics IV*. Amsterdam: Benjamins.

Brustad, K. E. 2000. *The syntax of spoken Arabic*. Washington, D.C.: Georgetown University Press.

Cadora, F. J. 1970. Some linguistic concomitants of contactual factors of urbanisation. *AL* 12, no. 1:10–19.

———. 1979. *Interdialectal lexical compatability in Arabic*. Leiden: Brill.

Cantarino, V. 1974–75. *The syntax of modern Arabic prose*. Bloomington: Indiana University Press.

Cantineau, J. 1960. *Etudes de linguistique arabe (Studies in Arabic linguistics)*. Paris: Klincksieck.

Carter, M. 1983. Language control as people control in mediaeval Islam: The aims of the grammarians and their cultural context. *AB* XXXI:65–84.

Caubet, D. 1986. Systèmes aspecto-temporels en arabe maghrébin (Aspect-tense systems in Maghrebi Arabic). *MAS-GELLAS* 3/4:97–132.

Chalabi, S. 1984. Modern Arabic terminology and bilingual lexicography: activities and problems. In *English for specific purposes in the Arab world*, ed. J. Swales and H. Mustapha, 280–91. Aston, UK: University of Aston Press.

Chejne, A. G. 1965. Arabic: Its significance and place in Arab Muslim Society. *MEJ* 19:447–70.

———. 1969 *The Arabic language: Its role in history*. Minneapolis: University of Minnesota Press.

Cohen, D. 1975. *Le parler arabe des Juifs de Tunis: Vol II Étude Linguistique (The Arab dialect of the Jews of Tunis)*. The Hague: Mouton.

Collected Speeches of Gamal Abdul-Nasir. n.d. (Ḥadīth al-Baṭal al-Zaʿīm Gamāl ʿAbdul-Nāṣir ʾilā l-ʾUmma), 4 vols. Cairo: Dār al-Taḥrīr.

Corriente, F. 1976 From Old Arabic to Classical Arabic through the Pre-Islamic koinè: Some notes on the native grammarians' sources, attitudes and goals. *JSS* 21:62–98.

Cowell, M. 1964. *A reference grammar of Syrian Arabic*. Washington, D.C.: Georgetown University Press.

Cuvalay, M. 1991. The expression of durativity in Arabic. In *Proceedings of the colloquium on Arabic grammar*, ed. K. Devenyi and T. Ivanyi (= Budapest Studies in Arabic 3/4). Budapest.

Diem, W. 1974. *Hochsprache und Dialekt im Arabischen (Standard language and dialect in Arabic)*. Wiesbaden: Franz Steiner Verlag.

Diyab, W. 1991. *Ḥaqāʾiq al-Iḥtilāl wa Wathāʾiq al-Taḥrīr (The facts of the occupation and the documents of the liberation)*. Beirut: al-Muʾassasa al-ʿArabīya al-Urubbīya li l-Ṣaḥāfa wa l-Nashr.

Donner, F. 1981. *The early Islamic conquests*. Princeton: Princeton University Press.

Doss, M. 1987. Further remarks on the use of the *b*-imperfect in Spoken Literary Arabic. *ZAL* 17:93–95.

Eid, M., and C. D. Holes, eds. 1993. *Perspectives on Arabic linguistics V*. Benjamins: Amsterdam.

Eisele, J. C. 1999. *Arabic verbs in time, tense and aspect in Cairene Arabic*. Harrassowitz: Wiesbaden.

———. 2002. Approaching diglossia: authorities, values and representations. In *Language contact and language conflict in Arabic*, ed. A. Rouchdy, 3–23. London: RoutledgeCurzon Press.

Eissfeld, O. 1965. *The Old Testament: An Introduction*. Tr. P. R. Ackroyd. Oxford: Blackwell.

El-Hassan, S. 1977. Educated spoken Arabic in Egypt and the Levant: A critical review of diglossia and related concepts. *ARLI* 8, no. 2:112–32.

———. 1978. Variation in the demonstrative system in educated spoken Arabic. *ARLI* 9, no. 1:32–57.

El-Yasin, M. K. 1985. Basic word order in Classical Arabic and Jordanian Arabic. *LNG* 65:107–22.

Emeneau, M. B. 1964. Oral poets of South India—The Todas. In *Language in culture in society*, ed. D. Hymes, 330–43. New York: Harper & Row.

Erwin, W. 1963. *A short reference grammar of Iraqi Arabic*. Washington, D.C.: Georgetown University Press.

Ezzat, A. 1974. *Intelligibility among Arabic dialects*. Beirut: Beirut Arab University.

Feghali, M. 1928. *Syntaxe des Parlers Arabes Actuels du Liban [Syntax of the modern Arabic dialects of Lebanon]*. Paris: Imprimerie Nationale.

Ferguson, C. A. 1959. The Arabic koinè. *LAN* 35:616–30.

———. 1968. Myths about Arabic. In *Readings in the sociology of language*, ed. J. A. Fishman, 375–81. The Hague: Mouton.

———. 1971. Diglossia. In *Language structure and language use: Essays by Charles Ferguson*. Stanford: Stanford University Press (reprint of *WO* 15:325–40).

———. 1978. The emphatic /l/ in Arabic. In *Contributions to Arabic linguistics*, ed. S. al-Ani. Bloomington: Indiana University Press.

———. 1989. Grammatical agreement in Classical Arabic and the modern dialects: A response to Versteegh's pidginisation hypothesis. *ALA* 22:5–17.

Fischer, W. 1959. *Die demonstrativen Bildungen der neuarabischen Dialekete (The demonstrative forms in modern Arabic dialects)*. The Hague: Mouton.

———. 1961. Die Sprache der arabischen Sprachinsel in Uzbekistan (The language of the Arabic speech island in Uzbekistan). *IS* 36:232–63.

Fischer, W., and O. Jastrow. 1980. *Handbuch der Arabischen Dialekte (Handbook of Arabic dialects)*. Wiesbaden: Harrassowitz.

Fleisch, H. 1968. *L'Arabe Classique: Esquisse d'une Structure Linguistique (Classical Arabic: Outline of the linguistic structure)*. Beirut: Dar al-Machreq.

———. 1979. *Traité de Philologie Arabe, Vol. II: Pronoms, Morphologie Verbale, Particules (Treatise on Arabic philology, Vol. II: Pronouns, verbal morphology, particles)*. Beirut: Dar al-Machreq.

Fück, J. 1955. Trans. Denizeau. *ʿArabīya*. Paris: Didier.

Furayha, A. 1955. *Naḥwa ʿArabīya Muyassara (Towards a simplified Arabic)*. Beirut: Dār al-Thaqāfa.

Garbell, I. 1958. Remarks on the historical phonology of an Eastern Mediterranean Arabic dialect. *WO* 14:303–37.

Gibson, M. 2002. Dialect leveling in Tunisian Arabic: Towards a new spoken standard. In *Language contact and language conflict in Arabic*, ed. A. Rouchdy, 24–40. London: RoutledgeCurzon Press.

Goodman, M. 1986. Review of Versteegh (1984). *JPCL* 1, no. 1:165–70.

Greenberg, J. 1978. The patterning of root morphemes in Semitic. In *Readings in Arabic Linguistics*, S. Al-Ani. Bloomington: Indiana University (repr. of *WO* 6:162–81).

Grotzfeld, H. 1983. Language hierarchy and speaking Arabic: Language constancy, variation and tolerance in an Arabic dialect area. *AB* XXXI:85–97.

Gully, A. 1993. The changing face of Modern Written Arabic: An update. *ALA* 27:19–59.

Haarmann, U. 1974. Religiöses Recht und Grammatik im klassischen Islam (Religious law and grammar in classical Islam). XVIII Deutscher Orientalistentag, ed. Voigt W. *ZDMG* Supp. II:149–69.

Haddad, G., and M. Kenstowicz. 1980. A note on the parallels between the definite article and the relative clause marker in Arabic. *SLS* 10, no. 2:141–47.

Haeri, N. 1996. *The sociolinguistic market of Cairo: Gender, class and education.* London: KPI.

Hamzaoui, R. 1965. *L'Académie Arabe de Damas et le Problème de la Modernisation de la Langue Arabe (The Arabic language academy of Damascus and the problem of modernizing the Arabic language).* Leiden: Brill.

Harrell. R. S. 1960. A linguistic analysis of Egyptian Radio Arabic. In *Contributions to Arabic linguistics,* ed. C. A. Ferguson, C.A., 1–77. Cambridge, Mass.: Harvard University Press.

———. 1962. *A short reference grammar of Moroccan Arabic.* Washington, D.C.: Georgetown University Press.

Heath, J. 1989. *From code-switching to borrowing: A case study of Moroccan Arabic.* London: KPI.

Henkin, R. 1992. The three faces of the Arabic participle in Negev Bedouin dialects: Continuous, resultative and evidential. *BSOAS* 55:433–44.

———. 1996. Negev Bedouin vs. sedentary Palestinian narrative styles. *IOS* 16:169–91.

Hinds, M., and El-S. Badawi. 1986. *A dictionary of Egyptian Arabic.* Beirut: Librairie du Liban.

Holes, C. D. 1983a. Bahraini dialects: Sectarian differences and the sedentary/nomadic split. *ZAL* 10:7–38.

———. 1983b. Patterns of communal language variation in Bahrain. *LS* 12, no. 4:433–57.

———. 1984. Bahraini dialects: Sectarian differences exemplified through texts. *ZAL* 13:27–67.

———. 1986a. The social motivation for convergence in three Arabic dialects. *IJSL* 61:33–51.

———. 1986b. Variation in the morphophonology of Arabic dialects. *TPS* 84:167–90.

———. 1986c. Review of Versteegh (1984). *BO* 43, nos. 1/2:218–22.

———. 1987. *Language variation and change in a modernising Arab state.* London, KPI.

———. 1990. *Gulf Arabic.* London: Routledge.

———. 1991. *Kashkasha* and the fronting and affrication of the velar stops revisited: A contribution to the historical phonology of the peninsular Arabic dialects. In *Semitic studies in honor of Wolf Leslau, vol. I,* ed. A. Kaye, 652–78. Wiesbaden: Harrassowitz.

———. 1993. The uses of variation: a study of the political speeches of Gamal Abdul-Nasir. In *Perspectives on Arabic linguistics V,* ed. M. Eid and C. D. Holes, 13–45. Amsterdam: Benjamins.

———. 1995a. The structure and function of repetition and parallelism in spoken Arabic: A sociolinguistic study. *JSS* 40:57–81.

———. 1995b. Community, dialect and urbanization in the Arabic-speaking Middle East. *BSOAS* 58:270–87.

———. 2001. *Dialect, culture and society in eastern Arabia. Vol I: Glossary.* Leiden: Brill.

Holm, J. 1988. *Pidgins and Creoles.* Cambridge: Cambridge University Press.

Holt, M. 1994. Algeria: Language, nation and state. In *Arabic sociolinguistics: Issues and perspectives,* ed. Y. Suleiman, 25–41. Richmond, UK: Curzon Press.

Hopkins, S. 1984 *Studies in the grammar of early Arabic.* Oxford: Oxford University Press.

———. 1988. Review of Versteegh (1984). *ZAL* 18:98–99.

Hourani, A. 1967. *Arabic thought in the liberal age, 1789–1939.* Oxford: Oxford University Press.

Husayn, T. 1966. *Al-'Ayyām (The days).* Cairo: Dār al-Ma'ārif.

Ibn Khaldūn. n.d. *Al-Muqaddima (Prolegomena).* Cairo: Maṭba'at Muḥammad 'Āṭif.

———. 1986. *The Muqaddima.* Trans. F. Rosenthal. London: RKP (repr. of 1958).

Ibrahim, M. H. 1986. Standard and prestige language: A problem in Arabic sociolinguistics. *AL* 28, no. 1:115–26.

Ibrahim, S. 1988. *Bayrūt, Bayrūt (Beirut, Beirut).* 2d ed. Beirut: Dār al-Mustaqbal al-'Arabī.

———. 1992. *Dhāt (Self).* Beirut: Dār al-Mustaqbal al-'Arabī.

Idris, Y. n.d. *'Arkhaṣ Layālī (The cheapest nights).* Beirut: Dār al-'Awda.

———. 1981. *Malik al-Quṭn (The cotton king) and Gumhūrīyat Farahāt (Farahat's republic).* Cairo: Dār Miṣr li l-Ṭibā'a.

'Imāra, M. ed.. 1972. *Al-'A'māl al-Kāmila li l-'Imām Muḥammad 'Abduh 2: al-Kitābāt al-Ijtimā'īya (The complete works of Imam Muḥammad 'Abduh, vol. 2: Social writings).* Beirut: al-Mu'assasa al-'Arabīya li l-Dirāsāt wa l-Nashr.

Ingham, B. 1982. *Northeast Arabian dialects.* London: KPI.

———. 1994. *Najdi Arabic: Central Arabian.* Amsterdam: Benjamins.

Janssens, G. 1972. *Stress in Arabic and word structure in the modern Arabic dialects.* Louvain: Peeters.

Johnstone, B. 1991. *Repetition in Arabic discourse.* Amsterdam: Benjamins.

Johnstone. T. M. 1961. Some characteristics of the Dōsiri dialect of Arabic as spoken in Kuwait. *BSOAS* 24:249–97.

———. 1963. The affrication of *kāf* and *gāf* in the Arabic dialects of the Arabian peninsula. *JSS* 8, no. 2:209–26.

———. 1967. *Eastern Arabian dialect studies.* London: Oxford University Press.

Jubouri, A. 1984. The role of repetition in Arabic argumentative discourse. In *English for specific purposes in the Arab world,* ed. J. Swales and H. Mustapha, 99–117. Aston, UK: University of Aston Press.

Kaye, A. 1994. Formal and informal Arabic: Diglossia, triglossia, tetraglossia etc., polyglossia—multiglossia viewed as a continuum. *ZAL* 27:47–66.

Khan, G. 1988. *Studies in Semitic Syntax.* Oxford: Oxford University Press.

Kofler, H. 1940–42. Reste altarabischer Dialekte (Remnants of the ancient Arabic dialects). *WZKM* 47:60–130, 233–62; 48:52–88, 247–74; 49:15–30, 234–56.

Kurpershoek, P. M. 1981. *The short stories of Yusuf Idris.* Leiden: Brill.

Labov, W. 1972. *Sociolinguistic patterns.* Philadelphia: University of Pennsylvania Press.

Levin, A. 1987. The particle *la* as an object marker in some Arabic dialects of the Galilee. *ZAL* 17:31–40.

Lichtheim, M. 1975. *Ancient Egyptian literature.* Los Angeles: University of California Press.

Lyons, M. C. 1962. *A classical Arabic reader.* Cambridge: Cambridge University Press.

Mahfouz, N. 1966. *Al-Qāhira al-Gadīda (Heliopolis),* 6th ed. Cairo: Dār Miṣr li l-Ṭibā'a.

Marçais, P. 1956. *Le Parler Arabe de Djidjelli (The Arabic dialect of Djidjelli).* Paris: Maisonneuve.

Masliya, S. 1997. The diminutive in spoken Iraqi Arabic. *ZAL* 33:68–88.

Matar, A. A. 1981. *Lahjat al-Badū fī al-Sāḥil al-Shamālī li Jumhūrīyat Miṣr al-'Arabīya (The Bedouin dialect of the north coast of the Arab Republic of Egypt).* Cairo: Dār al-Ma'ārif.

Mazraani, N. 1997. *Aspects of language variation in Arabic political speech-making.* Richmond, UK: Curzon Press.

Meiseles, G. 1979. Informal written Arabic. *IOS* 9:272–314.

———. 1980. Educated spoken Arabic and the Arabic language continuum. *ARLI* 11, no. 2:118–48.

Meissner, B. 1901. Neuarabische Sprichwörter und Räthsel aus dem Iraq (Modern Arabic proverbs and riddles from Iraq). *MSOS* 4:137–74.

Miller, A. M. 1986. The origin of the modern Arabic sedentary dialects: An evaluation of several theories. *ALA* 19:47–74.

Mitchell, T. F. 1952. The active participle in the Arabic dialect of Cyrenaica. *BSOAS* 14:11–33.
———. 1978. Educated spoken Arabic in Egypt and the Levant, with special reference to participle and tense. *JL* 14. no. 2:227–58.
———. 1986. What is educated spoken Arabic? *IJSL* 61:7–32.
———. 1990. *Pronouncing Arabic I*. Oxford: Oxford University Press.
Mitchell, T. F., and S. El-Hassan. 1994. *Mood, Modality and Aspect in Spoken Arabic*. London: KPI.
Monteil, V. 1960. *L'Arabe Moderne*. Paris: Klincksieck.
Moscati, S., A. Spitaler, E. Ullendroff, and W. von Soden. 1964. *An Introduction to the Comparative Grammar of the Semitic Languages*. Wiesbaden: Harrassowitz.
Munif, A-R. 1988. *Mudun al-Milḥ* (*Cities of Salt*). Beirut: al-Mu'assasa al-'Arabīya li l-Dirāsāt wa l-Nashr.
———. 1999. *'Arḍ al-Sawād* (*Black Earth*), Beirut, al-Mu'assasa al-'Arabīya li l-Dirāsāt wa l-Nashr.
Nicholson, R. A. 1969. *A literary history of the Arabs*. Cambridge: Cambridge University Press.
Noeldeke, T. 1909–19. *Die Geschichte des Qorâns* (*The history of the Koran*). Rev. F. Schwally. Leipzig: Dietrich'sche Verlagsbuchhandlung.
Nu'ayma, M. 1966. *Al-Nūr wa l-Dayyūr* (*Light and dark*). Beirut: Dār Ṣādir.
Nuṣṣ, A. A. 1930. *Mudhakkirāt Nashshāl* (*Diary of a pickpocket*). Cairo: al-Maktaba al-Yūsufīya.
Oestrup, J. 1897. *Contes Arabes de Damas* (*Arabic tales from Damascus*). Leiden: Brill.
Olmsted, G. J., and S. Gamal-Eldin. 1982. *Cairene Egyptian colloquial Arabic*. Lingua Descriptive Studies. Amsterdam: North-Holland.
Owens, J. 1988. *The foundations of grammar: An introduction to mediaeval Arabic grammatical theory*. Amsterdam: Benjamins.
Owens, J., and R. Bani-Yasin. 1987. The lexical basis of variation in Jordanian Arabic. *LING* 25:705–38.
Parkinson, D. 1985. *Constructing the social context of communication: Terms of address in Egyptian Arabic*. Berlin: Mouton de gruyter.
———. 1991. Searching for modern *fuṣḥa*: real-life formal Arabic. *ALA* 24:31–64.
———. 1993. Knowing standard Arabic: Language ability in its social context. In *Perspectives on Arabic linguistics V*, ed. M. Eid and C. D. Holes C.D., 47–73. Amsterdam: Benjamins.
Pasch, H., and R. Thelwall. 1986. *The development of Nubi*. Unpublished mimeo. Coleraine: Universities of Cologne and Ulster.
Persson, M. 2002. *Sentential object complements in Modern Standard Arabic*. Stockholm: Almqvist & Wiksell International.
Pinault, D. 1992. *Story-telling Techniques in the Arabian Nights*. Leiden: Brill.
Prochazka, T. 1988. *Saudi Arabian dialects*. London: KPI.
Prokosch, E. 1983. *Osmanisches Wortgut im Aegyptischen-Arabischen* (*Ottoman vocabulary in Egyptian Arabic*). Berlin: Klaus Schwarz Verlag.
Rabin, C. 1951. *Ancient West Arabian*. London: Taylor's Foreign Press.
———. 1955. The beginnings of Classical Arabic. *SI* 4:19–37.
Rafī' 1971. *Qaṣā'id Sha'bīya* (*Popular odes*). Kuwait: al-Maṭba 'a al-'Aṣrīya.
Rammuny, R. 1978. Functional and semantic developments in negation as used in Modern Literary Arabic prose after World War II. *JNES* 37, no. 3:245–64.
Reckendorf, H. 1895. *Die Syntaktischen Verhältnisse des Arabischen* (*Syntactic Relations in Arabic*). Leiden: Brill (repr. 1967).

Reinhardt, C. 1894. *Ein arabischer Dialekt gesprochen in Oman und Zanzibar* (*An Arabic dialect spoken in Oman and Zanzibar*). Amsterdam: Philo Press (repr. 1972).
Rosenbaum, G. M. 2000. '*Fuṣḥāmmiyya*: Alternating style in Egyptian prose. *ZAL* 38:68–87.
Rosenhouse, J. 1978. Circumstantial clauses in some Arabic dialects. *ZDMG* 128:226–37.
———. 1984. *The Bedouin Arab dialects*. Wiesbaden: Harrassowitz.
———. 1990. Tendencies to nominalisation in Modern Literary Arabic as compared with Classical Arabic. *ZAL* 22:23–43.
Rouchdy, A. ed. 1992. *The Arabic language in America*. Detroit: Wayne State University Press.
———. 2002. *Language contact and language conflict in Arabic*. London: RoutledgeCurzon Press.
Sa'id, N. Z. 1964. *Tārīkh al-Da'wa 'ilā l-'Āmmīya wa ' Āthāruhā fī Miṣr* (*The history of the call for the use of colloquial and its effects in Egypt*). Cairo: Maṭba'at Dār Nashr al-Thaqāfa.
Sallam, A. M. 1979. Concordial relations within the noun phrase in ESA. *ARLI* 10, no. 1:20–56.
———. 1980. Phonological variation in ESA: A study of the uvular and related plosive types. *BSOAS* 43:77–100.
Serjeant, R. B. 1983. Early Arabic prose. In *Arabic literature to the end of the Umayyad period*, ed. A. F. L. Beeston, T. M. Johnston, R. B. Serjeant, and G. R. Smith, 114–53. Cambridge: Cambridge University Press.
S'hiri, S. 2002. Speak Arabic please!: Tunisian Arabic speakers' linguistic accommodation to Middle Easterners. In *Language Contact and Language Conflict in Arabic*, ed. A. Rouchdy, 149–74. London: RoutledgeCurzon Press.
Singer, H-R. 1958. Neuarabische Fragewörter (Modern Arabian Interrogatives). Inaugural diss. Erlangen: Erlangen University.
———. 1984. *Grammatik des Arabischen von Tunis* (*Grammar of the Arabic of Tunis*). Berlin: de gruyter.
Sirat, A-S. 1973. Notes on the Arabic dialect spoken in the Balkh region of Afghanistan. *AO* 35:89–101.
Somekh, S. 1991. *Genre and language in modern Arabic literature*. Wiesbaden: Harrassowitz.
Sowayan, S. A. 1985. *Nabati Poetry*. Los Angeles: University of California Press.
Stekevych, J. 1970. *The modern Arabic literary language*. Chicago: Chicago University Press.
Suleiman, Y. 1993. The linguistic situation in Jordan and code-switching. *NAS* 1:1–20.
———. ed. 1994. *Arabic sociolinguistics: Issues and perspectives*. Richmond, UK: Curzon. Press.
———. 2003. *The Arabic language and national identity*. Edinburgh: Edinburgh University Press.
Tahar, A. 1975. *Al-Shi'r al-Malḥūn al-Jazā'irī* (*Algerian popular poetry*). Algiers: Société Nationale d'Edition et de Diffusion.
Taine-Cheikh, C. 2002. De la variation linguistique dans la prêche populaire mauritanien (Concerning linguistic variation in Mauritanian popular preaching). In *Language contact and language conflict in Arabic*, ed. A. Rouchdy, 177–202. London: RoutledgeCurzon Press.
Talmoudi, F. 1984. *The diglossic situation in North Africa*. Gothenburg: Acta Universitatis Gothoburgensis.
Thomason, S. G., and A. Elgibali. 1986. Before the lingua franca: pidginized Arabic in the eleventh century A.D. *LNG* 68:407–39.
Tomiche, N. 1964. *Le Parler Arabe du Caire* (*The Arabic dialect of Cairo*). The Hague: Mouton.
'Umar, A. M. 1970. *Tārīkh al-Lugha al-'Arabīya fī Miṣr* (*The history of the Arabic language in Egypt*). Cairo: Al-Hay'a al-Miṣrīya al-'Āmma li l-Ta'līf wa l-Nashr.

Uth, T. 1987. Der II und V Stamm im arabischen Verbalsystem und ihre Beschreibung mit se-
mantischen Valenzmodellen (Themes II and V in the Arabic verb system with a descrip-
tion of their semantic valency). Leipzig: Ph.D. diss.

Versteegh, K. 1983. Arabic grammar and the corruption of speech. *AB* 31:139–60.

———. 1984. *Pidginization and creolization: The case of Arabic.* Amsterdam: Benjamins.

———. 1997. *The Arabic language.* Edinburgh: Edinburgh University Press.

Vial, C. 1983. *L'Egyptien tel qu'on l'écrit (Egyptian as a written language).* Cairo/Paris: Institut
Français d'Archéologie Orientale du Caire.

Violet, B. 1901–2. Ein zweisprächiges Psalmfragment aus Damaskus (A bilingual psalm frag-
ment from Damascus). Berlin: Berichtigter Sonderabzug aus der *Orientalischen Litter-
aturzeitung.*

Voigt, R. 1988. *Die infirmen Verbaltypen des Arabischen und das Biradikalismus-problem (Arabic weak
verbs and the problem of biradicalism).* Stuttgart: Franz Steiner Verlag.

Vollers, K. 1906. *Volkssprache und Schriftsprache im alten Arabien (Popular language and written lan-
guage in ancient Arabia).* Strassburg, Amsterdam: APA-Oriental Press (repr. 1981).

Wansbrough, J. 1977. *Quranic Studies.* Oxford: Oxford University Press.

Wehr, H. 1953. Zur Funktion arabischer Negationen (On the function of Arabic negatives).
ZDMG 103:27–39.

Wise, H. 1975. *A transformational grammar of spoken Egyptian Arabic.* Oxford: Blackwell.

Woidich, M. 1975. Zur Funktion des activen Partizips im Kairenisch-Arabischen (On the func-
tion of the active participle in Cairene Arabic). *ZDMG* 125:273–93.

———. 1991. Die Formtypen des Zustandssatzes im Kairenischen (Circumstantial clauses in
Cairene Arabic). *ZAL* 23:66–98.

Wright, W. 1896–98. *A grammar of the Arabic language.* 2 vols. Cambridge: Cambridge Uni-
versity Press (repr. 1964).

Younes, M. 1993. Emphasis spread in two Arabic dialects. In *Perspectives on Arabic linguistics
V*, ed. M. Eid and C. D. Holes, 119–45. Amsterdam: Benjamins.

Zaydān, J. n.d. *Bunāt al-Nahda (Builders of the Arab renaissance).* Cairo: Dār al-Hilāl.

Ziadeh, F. J. 1964. *A reader in modern written Arabic.* Princeton: Princeton University Press.

Zwettler, M. 1978. *The oral tradition of classical Arabic poetry.* Columbus: Ohio State Univer-
sity Press.

Index