The Oxford Handbook *of*
ARABIC
LINGUISTICS

THE OXFORD HANDBOOK OF

ARABIC
LINGUISTICS

THE OXFORD HANDBOOKS IN LINGUISTICS

THE OXFORD HANDBOOK OF

ARABIC LINGUISTICS

JONATHAN OWENS

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide.

Oxford   New York

Auckland   Cape Town   Dar es Salaam   Hong Kong   Karachi
Kuala Lumpur   Madrid   Melbourne   Mexico City   Nairobi
New Delhi   Shanghai   Taipei   Toronto

With offices in

Argentina   Austria   Brazil   Chile   Czech Republic   France   Greece
Guatemala   Hungary   Italy   Japan   Poland   Portugal   Singapore
South Korea   Switzerland   Thailand   Turkey   Ukraine   Vietnam

Oxford is a registered trademark of Oxford University Press
in the UK and certain other countries.

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

To the once and future study of the Arabic language

This page intentionally left blank

Contents

Contributors

Enam Al-Wer Essex University

Elabbas Benmamoun University of Illinois

Ramzi Baalbaki American University of Beirut

Peter Behnstedt PD Hamburg University

Abdelali Bentahila Abdelmalek Essaadi University, Tangier

Sami Boudelaa United Arab Emirates University

Tim Buckwalter University of Maryland

Lina Choueiri American University of Beirut

Peter T. Daniels Independent Scholar

Eirlys Davies Abdelmalek Essaadi University, Tangier

Everhard Ditters Nijmegen (Radboud) University

Lutz Edzard Oslo University

Mohamed Embarki University of Besançon

Sam Hellmuth University of York

Clive Holes Oxford University

Maarten Kossmann Leiden University

Pierre Larcher Université de Provence

Stefano Manfredi University of Naples "L'Orientale" and University of Turin

Daniel L. Newman University of Durham

Jonathan Owens Bayreuth University

Dilworth B. Parkinson Brigham Young University

Robert Ratcliffe Tokyo University of Foreign Studies

Jan Retsö University of Göteburg

Karin Christina Ryding Georgetown University

Solomon I. Sara Georgetown University

Yasir Suleiman Cambridge University

Mauro Tosco University of Turin

Manfred Woidich University of Amsterdam

Abbreviations[1]

AH/CE, e.g., Ibn Sīda (458/1066) = Islamic calendar = AH 458/Gregorian calendar = CE 1066,

(*X) token not correct with material in brackets

*(X) token not correct without material in brackets

< > encloses root consonants or specified phonological element

⟨ ⟩ orthographic unit

e.g. [Larcher, "ALT II"] Refer to the article "Arabic Linguistic Tradition II" written by Pierre Larcher in this volume.

1	first person
2	second person
3	third person
4M	4M model of codeswitching
ACC	accusative
ADV	adverb
AGT	Arabic grammatical tradition
ALT	Arabic linguistic tradition
ANT	anterior
AP	active participle
AP	adjective phrase
ATR	advanced tongue root
AUX	auxiliary
C	consonant
CA	Classical Arabic
CL	computer linguistics
CLLD	clitic-left dislocation
CMC	computer-mediated communication
CMPL	completive
CollA	colloquial Arabic
COMP	complementizer

CONJ	conjunction
COP	copula
CP	complementizer phrase
CS	codeswitching
CS	construct state
CV	consonant vowel
D	determiner
DEF	definite
DP	determiner phrase
DG	dependency grammar
e	in syntax, indicates structurally relevant gap
EEG	electroencephalography
EgA	Egyptian Arabic
EL	elative
EL	embedded language
EMA	electromagnetic midsagittal articulometry
ENS	educated native speaker
EXS	existential
F1 (etc.)	first formant
F	feminine
FP	functional projection
FT	future
fMRI	functional magnetic resonance imaging
Fr.	French
GEN	genitive
GPA	Gulf Pidgin Arabic
HAB	habitual
Hz.	Herz
IA	Iraqi Arabic
IC	immediate constituent
IDF	indefinite
IMPF	imperfect
INDC	indicative
IP	inflectional phrase
IPA	International Phonetic Alphabet
IRR	irrealis
JA	Jordanian Arabic
JSV	jussive
L1	native language
L2	second language
LA	Lebanese Arabic
LAD	language acquisition device
M	masculine

MA	Moroccan Arabic
MEG	magnetoencephalography
ML	matrix language
MLD	Moroccan-flavored Dutch
MLF	Matrix Language Frame model
ms.	milleseconds
N	noun
NA	Nigerian Arabic
NEG	negative
NegP	negation phrase
NLP	natural language processing
NOM	nominative
NP	noun phrase
Num	numeral
NumP	number phrase
OA	Old Arabic
OBJ	object
OT	optimality theory
P	phrase
PA	Palestinian Arabic
P/C	Pidgin/Creole
PET	positron emission tomography
PL	plural
P	preposition
PP	prepositional phrase
PRE	preformative vowel of imperfect
PRED	predicate
PRES	present
PROG	progressive
Pron	pronoun
PRT	participle
PSSD	possessed
PSSR	possessor
PST	past
PSV	passive
PT	processability theory
Q	quantifier
QP	quantifier phrase
R&P	root and pattern (theory)
REL	relative
RP	resumptive pronoun
SA	Standard Arabic (also known as Modern Standard Arabic)
SBJ	subjunctive

SG	singular
SILL	strategy inventory for language learning
SLA	second-language acquisition
Spec	specifier
SQUID	superconducting quantum inference device
SUBJ	subject
SVO	subject–verb–object
T	tense
TAD	traditional Arabic dialectology
TAFL	teaching of Arabic as a foreign language
TJA	Tripolitanian Jewish Arabic (Libya)
TL	target language
TMA	tense, mode aspect marker
TMS	transcranial magnetic stimulation
TP	tense phrase
UG	universal grammar
UPSID	University of California–Los Angeles Phonological Segment Inventory Database
V	verb
V	vowel
VOS	verb–object–subject
VOT	voice onset time
VS	verb–subject
WAD	*Wortatlas der arabischen Dialekte*
WFR	word formation rules
WSA	Western Sudanic Arabic

Journals, Book Series, and Organization Abbreviations

ACTFL	American Council for the Teaching of Foreign Languages
AIDA	Association Internationale de la dialectologie Arabe
AL	*Anthropological Linguistics*
BSOAS	*Bulletin of the School of Oriental and African Studies*
CERES	Centre d'Études et de Recherches Économiques et Sociales
EALL	*Encyclopedia of Arabic Language and Linguistics*
EI2	*Encyclopedia of Islam/Encyclopédie de l'Islam, nouvelle édition*
GLECS	Groupe Linguistique d'Études Chamito-Sémitique
GURT	Georgetown University Round Table on Languages and Linguistics
HSK	Handbücher zur Sprach- und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science
IFAO	Institut Français d'Archéologie orientale.
ILR	Interagency Language Roundtable

INALCO	Institut National des Langues et Civilisations Orientales
IULC	Indiana University Linguistics Club
JAOS	*Journal of the American Oriental Society*
MAS Gellas	Matériaux Arabes et Sudarabiques
PAL	Perspectives on Arabic Linguistics (book series, Benjamins)
ZAL	*Zeitschrift für arabische Linguistik*
ZDMG	*Zeitschrift der deutschen morgenländischen Gesellschaft*

This page intentionally left blank

TRANSCRIPTION AND TRANSLITERATION EQUIVALENCES

aa = ā
ii = ī = iy
uu = ū = uw
x = ḫ = kh
ħ = ḥ
ʕ = ʻ
ʔ = ʼ
θ = ṯ = th
đ = ḏ = dh
ʃ = š = sh
ʒ = ž
dʒ = dž = ǧ = dj
ɣ = ġ = gh = gh

"Emphatics"

ṣ = ş = sˤ = ص
ṭ = ŧ = tˤ = ط
ḍ = đ = dˤ = ض
ḍ = đ = dˤ = ظ
ẓ = ƶ = ظ (depending on reflex value)
Other emphatic consonants indicated by dot under consonant, for example, ṃ, ṛ, ḷ

This page intentionally left blank

Maps

This page intentionally left blank

THE OXFORD HANDBOOK OF

ARABIC

LINGUISTICS

This page intentionally left blank

CHAPTER 1

A HOUSE OF SOUND STRUCTURE, OF MARVELOUS FORM AND PROPORTION

an introduction

JONATHAN OWENS*

1.1 THE INTEREST OF ARABIC: PROPOSITION 1

ARGUABLY, for the linguist, Arabic is the most interesting language in the world. I will term this "Proposition 1." This claim will certainly strike most as either arrogant or woefully wrong-headed, otiose, and lacking any measurable basis of substantiation. It furthermore runs afoul of deeply embedded beliefs in linguistics itself.

In particular is the assumption that all languages are, for purposes of linguistic analysis and insight they give into the universal properties of language, equal. Indeed, on this basis one can agree only that there is no a priori reason to think that the structure of Arabic will tell us more about language than will, say, the structure of Dɣweɗe, a Central Chadic language spoken by perhaps 40,000 speakers. In terms of its grammatical properties alone, Arabic has no more claim to the attention of linguists than does any other language.

To hypothetically formulate a second objection, it might be argued in some circles that Arabic should have a special linguistic place due to being the language of Quranic

revelation. While this position may have its partisans among some, it in fact has no inherent connection to its status within linguistics, as indeed was recognized by many of the Classical Arabic grammarians themselves (e.g., Ibn al-Nadim, cited in [Owens, "History"]).

A third objection is simply that there is no basis for defining what "interesting" means. This brings us to the defense of the proposition.

1.1.1 The Geographical, Demographic, Chronological, Cultural Gestalt

First, and most basically, once one factors away the grammatical, semantic, pragmatic, and formal aspects of languages, it is clear that not all languages are equal.

This can be measured first of all with simple quantitative standards. There are large languages and small languages. Arabic is one of the world's largest, spoken natively by about 300 million speakers, and as a second language (L2) by perhaps another 60 million. It is by a large margin the largest language in Africa (nearly 200 million speakers) and one of the biggest in Asia (120 million). It has been estimated to be the fifth largest language in the world in terms of native speakers. Strength of numbers alone guarantees it communicative centrality in the world language system (de Swaan 1998, 2001).

Arabic is equally spoken over one of the largest land areas of any native language. It is spoken continuously in the east from Iraq and Khuzistan in southwest Iran, all the way to Morocco and to northeastern Nigeria in the west, an area covering nearly a seventh of the latitudinal distance of the globe. In addition, a number of Arabic-speaking Sprachinseln can be found outside of this area (see Map 1.4 in the Appendix at the end of this chapter).

Arabic is furthermore the language of the Quran, Islam being the only one of the large religions whose holy book is revealed in a specific language. Hence, it is learned to one degree or another for religious, ritual, cultural, and legal purposes by nearly all Muslims,[1] and equally important, is therefore revered as the purveyor of God's word. It is the language of the great texts of Arabic–Islamic culture. "Arabic" thus binds the

---

[1] For native speakers, Procházka's (2006) estimate of 280 million strikes us as reasonable, if perhaps slightly low. In addition, Arabic is spoken fluently as a second-language lingua franca in particular in Algeria, Morocco, Mauretania, Libya, Yemen, Chad, Tunisia, and the Sudan.

An estimate of 452 million "total" speakers, such as found at http://en.wikipedia.org/wiki/List_of_ languages_by_number_of_native_speakers#30_to_50_million_native_speakers, should be treated with great caution. Estimating total number of speakers in a language like Arabic begs the question of what a language speaker is. In a survey carried out among Kanuri, one individual reported to me that she uses Arabic "often" (Owens 1995). When I thereupon addressed her in Arabic, she could not understand a word. She explained that she began many acts with *bi sm illaahi* ("in the name of God"). Defining "total" (of what?) is no less a slippery task than defining "often."

communicative, intellectual, and emotional in one linguistic gestalt, in a way perhaps no other language in the world today does.

The history, both written and orally reconstructible, of the Arabic-speaking peoples is, compared with most languages, well documented, even if from the specialist's perspective gaps in the history are perhaps more prominent than what is available. The first reference to Arabs, which may be inferred to be a reference to Arabic speakers, dates from 853 BCE, North-Arabic clan names are mentioned even earlier (Lipiński 2000: 101, 457), and Arabic begins spreading with great rapidity out of its core Middle East location in the Arabian peninsula, Iraq, and Syrian and Jordanian desert at the beginning of the Islamic era (nominally, 622 CE). By 92/711, relatively large and self-contained groups of Arabic speakers stretch from Uzbekistan in the east to Spain (Andalusia) in the west. A further significant expansion out of Upper Egypt into the Lake Chad area around 800/1400 extends this region. With the exception of Spain, and allowing for modern, "global" diasporas, this essentially defines the limits of the Arabic-speaking world until today (see Owens 2009, chapter 1, for broader summary).

The linguistic consequences and challenges of this geo-history are self-evident. While Arabic has even in pre-Islamic times always been dialectally diverse (Rabin 1951), this diversity has probably increased in the wake of the great Arab–Islamic expansion. If till today simple models for classifying Arabic dialects elude us [Behnstedt and Woidich, "Dialectology"], it is no doubt in large part because an originally diverse proto-situation has continued to diversify across the vast geographical region where Arabic is spoken. Hand in hand with cataloguing the dialectal diversity goes the challenge of developing an historical linguistic model that accounts for the present-day situation. If, as argued in this volume [Owens, "History"], traditional accounts of Arabic language history have generally failed to provide linguistically adequate models of historical development, work on a comprehensive account is largely in its incipient stages.

Not surprisingly, in its expansion across a seventh of the earth's latitudinal distance, speakers of Arabic have come into contact with a large number of languages. The degree to which spoken Arabic itself has been globally affected by this contact is a matter of ongoing debate, with some scholars, such as Versteegh (particularly 1984), arguing that the effects have been profound, whereas others, including Kossmann ["Borrowing"], see Arabic often as the dominant, hence imposing, language in contact situations. Certainly the latter perspective receives support from those well- or fairly well-documented extreme situations where unquestionably, or arguably, new varieties arise from the contact. One of these concerns the emergence of Pidgin and Creole varieties in the Sudanic region and East Africa, varieties that emerged from a common ancestor in the 19th century, today variously known as Turku, Juba Arabic, and Nubi or Kinubi. Since Versteegh's (1984) argument that the structure of Arabic dialects is to be accounted for by having passed through a stage of Pidginization, a counterconsensus ([Tosco and Manfredi, "Creoles"]) has developed that these Pidgin/Creole varieties are indeed entirely new languages, following the classical model of creolization, with little implication for understanding mainstream historical developments of contemporary Arabic dialects. Relatively underdebated are Uzbekistan and Afghanistan Arabic, spoken by

very small populations. Whereas these varieties have classic features of Arabic verbal morphological structure, in other areas of grammar they display marked deviations from any other variety of Arabic, for instance, in having a fixed subject–object–verb (SOV) word order. All deviations are readily explicable as influence from the Dari, Tajik, and Uzbek adstrates, and therefore the question can be raised as to whether these varieties are typologically mixed languages ([Tosco and Manfredi, "Creoles"]).

Before adducing more evidence in favor of Proposition 1, it is relevant here to take stock of the argument to this point. Beginning from older, classical perspectives on language, issues in Arabic dialectology and language history are multifarious, the challenge of building a comprehensive descriptive database remains high, and questions of language contact all along the vast geographical expanse of Arabic are open. Each of these domains represents a significant linguistic challenge, certainly descriptively but also methodologically and theoretically: what is the role of contemporary dialects in reconstructing language history; what determines direction of influence (van Coetsam 2000); what domains of language are more liable to contact influence; why do ostensibly similar global social conditions among communities of Arabic speakers lead to radically different linguistic outcomes (Owens 2000: 23); indeed, does a definable construct "Arabic" exist [Retsö, "Arabic?"]. But matters become even more interesting linguistically when the two peripheral varieties, Juba Arabic/Nubi and Uzbekistan Arabic, are added. Arabic is the only language in the world from which have emerged both Creole varieties and, arguably, mixed-language varieties. Arabic thus provides a living model for linguistics as a whole to address classic questions of historical and contact linguistics: what happens structurally to a language in the case of normal transmission (in general, the end product of the contemporary dialects) versus, by way of comparison, extreme situations of sociopolitical upheaval or cases of intense contact in a minority situation (Thomason and Kaufman 1988). Interim positions along the continuum formed by these poles can be integrated into linguistic typologies (e.g., Maltese, Kormakiti Arabic in Cyprus, Anatolian Arabic). Certainly, in the domains of phonology and morphology and also to some degree syntax, rigorous measures of core (necessary, not sufficient) Arabic could be constructed. Lurking in the background is the question of how inferences can be drawn from today's situations to interpret issues of Arabic historical linguistics and how, proceeding from contemporary sociolinguistics methodologies, determining factors in such developments can be extrapolated.

1.1.2 The Classical Language, the Linguistic Tradition

The factors summarized in the previous section alone are of enticing interest to linguistics, without mention even having been made of what is unquestionably the most central icon of Arabic: the classical language. It is remarkable that what today is for some *the* form of Arabic—the ʕArabiyya, or the Fuṣħaa, popularly known as Standard or Modern Standard Arabic—is by and large identical to the form of Arabic broadly described by the late 2nd-/8th-century grammarian Sibawaih.

The functions of the ʕArabiyya are legion. Most centrally, it is, roughly, the variety of Quranic revelation. It is the variety that came to symbolize the remarkable intellectual and cultural flowering in the Islamic era and the variety around which the Arabic script developed [Daniels, "Writing"]. It is the variety that became a central cultural and political pillar of the Arabic nahḍa "renaissance" movement of the 19th century ([Newman, "Nahḍa"]) and enjoys the status of official language in 23 nation states today (see Map 1.3 in the chapter Appendix) with its concomitant importance in modern educational systems, it is the variety typically taught in non-Arab universities [Ryding, "Acquisition"], and it continues to be an essential element in any debate on Arab identity [Suleiman, "Folk Linguistics"].

Each and every one of these associations implies linguistic issues of different types: descriptive, historical, political, second-language acquisition. What is most remarkable, however, is the Arabic linguistic tradition itself, which was built on the basis of one of the true classics of linguistics, the *Kitaab* of Sibawaih (Baalbaki 2008; ["ALT I"]). The very first book on Arabic grammar (so far as our documented record of transmission goes) is a comprehensive (nearly 1,000 pages) descriptive work built on a highly elaborated grammatical theory. While opinions differ as to the origin of the post-Sibawaih Arabic linguistic tradition, it is clear that a highly sophisticated and differentiated theoretical grammatical and pragmatic discourse continued to develop for at least the next 500 years [Larcher "ALT II"]. No less interesting and significant is the voluminous lexicographical tradition that developed in tandem with the grammatical [Sara, "Classical Lexicography"].

Students of Arabic therefore deal not only with the varieties of Arabic themselves but also with a metadiscourse, as it were, which was established within Arabic–Islamic culture. Arabic texts were passed down to us, along with a theoretical framework for analyzing them, constitutive of the Arabic–Islamic tradition, which continues to be of central importance in the contemporary teaching of Arabic and which challenges the interpretive acumen of linguists studying this tradition.

Thus, with respect to Proposition 1, it is not only that Arabic is one of the few languages of the world within which developed a linguistic tradition; also, it is a tradition that continues to exercise its influence on today's Arabs and Arabic society and beyond to Islamic society.

1.1.3 Arabic and Arab Identities

The two previous points set the stage for the inherent language tension that exists in contemporary Arabic societies. Arabic, the mother tongue of its approximately 300 million speakers, is not the same Arabic as the Arabic that is codified and has official political status and cultural centrality through its association with the Quran and with pan-Arab identity.

On one hand, these two broadly defined varieties can be represented as mutually opposed: official versus unofficial, written versus spoken, formal versus

informal, pan- versus local, learned formally versus acquired as a first language (L1). The functional contrasts were made famous by Ferguson (1959). Equally, one can emphasize the complementarity of the codes. The native colloquial is the language not only of home and friends but also of all that is informal, unofficial, spontaneous, and intimate. The growing entertainment industry in its diverse media manifestations is thus wholly dominated by the colloquials, as is the informal world of texting and twittering [Holes, "Orality"]. Blogging, a domain awaiting comprehensive linguistic research, appears to cover a spectrum of styles.

The difference between the two is also one of ideology versus practice, of ideal versus real. The fuṣḥaa, even if in its perceptions and usage it is a variety of fuzzy contours (Kaye 1972; Parkinson 1991) and is rarely[2] used in the real world in its prescribed form, is the variety of preeminent cultural importance [Suleiman, "Folk Linguistics"].

Sociolinguistics, a subdiscipline of linguistics of relatively recent provenance closely related to the older dialectology, shows the degree to which ideal and real can differ in the realm of spoken Arabic. The careful microdocumentation of speech communities consistently has shown (studies from the Arabian–Persian Gulf, Saudi Arabia, Jordan, Damascus, Bethlehem, Cairo, Casablanca, and northeast Nigeria) that features of spoken colloquial varieties are what drive language change [Al-Wer, "Sociolinguistics"]. Moreover, when Arabic meets other languages bilingually, it is again the colloquial that always forms the basic matrix of contact [Davies et al., "Codeswitching"]. Even in mixed colloquial–fuṣḥaa exchanges such as on media talk shows, the colloquial can have a dominant role.

The vibrant co-existence of quite differentiated varieties, a situation hardly unique to Arabic, nonetheless takes on a special, perhaps unique status in the world's languages, precisely because each variety, beyond its linguistic profile, embodies a different history, a different symbolism, a different legitimization. While these differences are of central interest to students of linguistics, they extend beyond the academic lecture hall to the real world of language teaching and language policy. To which variety, for instance, should a program of second-language teaching be tailored, or, if the varieties have different cognitive profiles, what are the implications for L1 teaching? These are questions best not answered by policy fiat. Indeed, the experience of Arabic in post-9/11 America represents probably the sorriest example ever of huge resources expended for developing language teaching programs, largely divorced from the fundamental research on

---

[2] The crucial adverb *rarely* should be understood as follows. Arabic is spoken by, conservatively, 300 million individuals. Each individual, probably conservatively, speaks for two hours per day, at 10,000 words per hour (slightly low probably), giving 6 trillion words of Arabic per day. The only forums where a normative, spoken Standard Arabic is used are certain media broadcasts (e.g., the excellent news channels al-ʕArabiyya or al-Jaziyra, national and commercial channels mainly for information-orientated topics such as news and documentaries) and in various official meetings, including some but hardly all educational formats (see Mejdell 2006; also [Holes, "Orality"]). Of the 300 million speakers, only a tiny minority of them are engaged at any one time in a function prescribing the use of Standard Arabic. Otherwise, for most individuals nearly always, and for all at some time, the basis of everyday speech is a colloquial variant.

the language being taught that would make for a more rational and efficient teaching program [Ryding, "Acquisition"]. Research from across the spectrum of linguistics is implicated in any academicization of Arabic teaching, whether as an L1 or L2.

1.1.4 Grammar

Arabic is thus a language of rare breadth and extension in the world, a language like perhaps no other in the degree to which it embodies the culture and politics of its speakers. It is, however, a language, and it has been studied from a number of classical grammatical perspectives. Even here Arabic has structural features that set it apart from many, sometimes most, of the world's languages.

The phenomenon of emphasis (pharyngealization) of consonants is a hallmark of the language and has engendered numerous studies both in phonetics [Embarki, "Phonetics"] and in phonology [Hellmuth, "Phonology"]. What is emblematic of Arabic, however, hardly exhausts the interest of Arabic for linguistics. As Hellmuth points out, for instance, stress in Arabic has been of central interest in phonological theory.

In morphology, an ongoing debate surrounding Arabic and many other Semitic languages is the status of the consonantal root as a morphemic element. As Ratcliffe ["Morphology"] points out, the Arabic grammatical tradition itself viewed the stem, not the root, as the basis of morphology, and arguments from within contemporary morphological theory have been developed for this as well. But equally, psycholinguistic studies on the basis of carefully constructed experiments have interpreted the consonantal root as having a crucial role in morphological processing [Boudelaa, "Psycholinguistics"].

Besides the Arabic grammatical tradition itself (1.1.2), there are two further prominent approaches to Arabic grammar. The older one is the philological tradition [Edzard, "Philology"], with which the study of Arabic grammar in the West began. Besides its general interest in Arabic grammar, this tradition incorporates cultural issues and has been present at the interface of Arabic texts of all genres and language. The other is more recent and is based on the precepts of theoretical grammar, particularly syntactic theory in the generative tradition, which endeavors to locate what is specifically Arabic within a broader program of universal grammar [Benmamoun and Choueri, "Syntax"].[3] All of the formal grammatical domains feed into the growing domain of computational linguistics and into the broader field of natural language processing [Ditters, "Computational"].

Finally, the classical lexicographical tradition has its counterpart in contemporary lexicography, a field increasingly drawing the vast online publishing industry in Arabic

---

[3] Chapter 6 is a double chapter; the original intent was to have two separate chapters, one on the standard language and the other on dialects. Individual circumstances required conflating the two into one.

for its sources [Buckwalter and Parkinson, "Modern Lexicography"]. Here again one experiences the special challenges confronting the Arabic lexicographer, for instance, whether to lemmatize according to root or stem, how to sublemmatize parts of speech, and whether to lump polysemously or to differentiate identical forms.

The articles in this handbook describe a language that, when looked at in its totality, is of rare thematic linguistic differentiation.

1.2 Scope and Choice of Chapter Topics

Proposition 1 encapsulates an ideal. The handbook is intended to reflect the full breadth of research on Arabic linguistics in the West. Realistically, this implies that it includes only chapters on topics judged to have a critical mass of background research. The reader will therefore miss domains that might be expected in a linguistics handbook. Asymmetries will be noticeable. There is a chapter on L2 acquisition but none on L1 acquisition, a chapter on sociolinguistics, but none on oral discourse, a number of chapters on grammar but none on semantics. The gaps are regrettable but unavoidable so long as the focus of the chapters is on the domains of Arabic linguistics that do indeed enjoy a fairly broad and deep coverage rather than on Arabic-flavored general linguistics, as it were.[4]

The chapters themselves reflect domains of research with great disparities of detail. In some cases the chapter is able to cover nearly all of the published research on a given domain, for instance, the chapter on Pidgins and Creoles and even, surprisingly (see remark at end of 1.1.3), work on L2 Arabic language acquisition. In others the breadth of available material has meant that authors could summarize only broad lines of research, illustrating the topic in greater detail with selected examples. Arabic language contact, particularly as reflected in loanwords, for instance, has a very large literature; the research on Arabic dialects is immense, and the research on the Arabic grammatical tradition is large. As far as Western research goes, these disparities to some degree reflect the relative age of the subdomain. In general, codeswitching, psycholinguistics, sociolinguistics, and pidgin and creole linguistics, for instance, are barely 30 or 40 years old as independent specializations of linguistics. Research on Arabic dialects, on the other hand, was already well established in the 19th century. This does not, however, imply

---

[4] For instance, the justifiably well-regarded *Encyclopedia of Arabic Language and Linguistics* has a chapter on "Cohesion" (Khalil 2006) with nine non-Arabic items in the bibliography and ten on Arabic. Unfortunately, this breakdown realistically reflects the dearth of material on spoken Arabic discourse, for instance, only one book-length work, an edited volume (Owens and Elgibali 2010), which is too little in the editor's view to merit a separate chapter here. The article preceding Khalil's on "Coherence," a central topic equally in literary and spoken texts, treats the subject only as it is reflected in the Classical literary tradition (Faiq 2006). The limitation is regrettable but does reflect the unbalanced state of the art in this domain.

that any domain of Arabic linguistics has been exhaustively treated. As Behnstedt and Woidich point out ["Dialectology"], many dialects, for instance, are poorly described, and the integration of dialectology and sociolinguistics, an essential element of sociolinguistics in the West, has seen only modest progress in the case of Arabic, while historical dialectology, a part of the general field of Arabic historical linguistics, is meager at best.

Gaps should certainly be seen as a challenge to open up wider avenues of research.

1.3   The Real World of Research on Arabic: A Critical Look

Given the current state of research on Arabicist it may be asked: if Proposition 1 is correct, does the linguistic research match the inherent interest of the language?

Here I would answer with only a very conditional "yes." On one hand, as noted in the previous section, there are areas of research with a large literature and well-established research tradition. On the other hand, there are topics central to the study of any language with only modest research traditions in Arabic. Studies on spoken Arabic discourse are rare (see note 4), while more recent domains of linguistics such as psycholinguistics, sociolinguistics, or the study of spoken Arabic pragmatics, though growing, are still in their incipient stages.

Ultimately, however, the study of a language must be more than the sum of its parts. It will be suggested here that, as far as Arabic goes, a holistic linguistic tradition remains an as yet unrealized desideratum. In the past and currently, a number of factors militate against this development. Four factors can be identified.

1.3.1 Arabic Is Large

The first is simply the immensity of the field itself. Arabic presents prima facie anything but a unified domain of inquiry. Consider, for instance, the two basic media that Arabic linguistics works with: the written and the spoken word, the former of which is associated with the Classical and Standard language and the latter with the dialects. These two media are in important respects of a different nature. The written domain is a learned domain, one that itself continues a heritage dating back to the 2nd/8th century, whose standard and norms have been long established. While one might be able to change certain aspects of the Standard language, such as the idiomatic domain ([Newman, Kossmann]), one cannot change its morphology or syntax. The spoken domain, on the other hand, is beholden to contemporary methods of descriptive and field linguistics, associated with, inter alia, corpus collection and language documentation, work with expert consultants, and instrumental phonetics of the spoken language. Norms, such as there are in this domain, emerge from the individual research studies undertaken in it.

Experience, moreover, has shown that in the Western tradition these two domains exist largely in parallel universes, with scholars linked to one or the other but not both. Those concerned with the written language, for instance, to the extent that they move outside the field of the linguistics of the written varieties, gravitate toward the other literary domains of Arabic such as Arabic literature, law, and medical texts. Many such individual cases could be cited, but quite typical in this respect is Carl Brockelmann, whose *Grundriss der semitischen Sprachen* (1908, 1913) remains a standard reference work. After publishing this work, he went on to write another well-regarded book, *Geschichte der islamischen Völker* (1943) (*History of Islamic Peoples*) . Brockelmann never studied a spoken variety of Arabic, and his *Grundriss*, while a work of compendious scholarship, is marked by a decided antipathy toward theoretical issues in historical and contact linguistics (Owens 2009: 43), precisely two areas where Arabic is particularly implicated, as discussed already.[5]

Those working in the realm of the spoken language, on the other hand, are faced initially with a plethora of challenges, for instance, which aspect of language to concentrate on or which varieties of Arabic to try to delineate. Finding a format to integrate these in turn with the Classical or Standard varieties may imply defining variables that are central to neither tradition.

Edzard ["Philology"] correctly notes that there is in principle no contradiction between a philological (written) orientation and a "theoretical" linguistic one; experience has nonetheless shown that relatively few scholars not only work in both domains but also, more importantly, attempt a synthesis of the two.

1.3.2 Stovepiping

The problem is at once abetted and exacerbated by the stovepiping characteristic of contemporary academia. Whereas 30 years ago one could claim to be a linguist, today it is more likely that one will be a sociolinguist, psycholinguist, or general or specialized syntactician. Certainly these developments follow their own internal logic, as methods and theoretical perspectives have become more specialized during this period. At the same time, in this there is the danger that the academic apparatus defines the language rather than the language being served by the apparatus.

To take an example from sociolinguistics, one can ask how many studies are needed to define the social status of the "qaf" variable. On one hand, the fact that there have been fruitful studies on this variable means that it provides a necessary and interesting comparative breadth; on the other hand, certainly many other variables, some of broad

---

[5] Indeed, it is striking that while comparative Semitic and comparative Indo-European literature both came of age in the same era, the 19th century, and to a large degree in the same region—Central Europe—the theoretical contribution of the former to the development of general principles of historical linguistics was negligible whereas that of the latter was essential.

comparative potential and others of particular local interest, await treatment. Added to this, embedding the findings on a comparative basis in the vast Arabic world is a challenge that has received relatively little attention from Arabic sociolinguists.[6] Beyond this is the ever-present danger of calling the game over as soon as a sociolinguistic phenomenon has been studied from within a particular theoretical perspective, as often as not one initially defined from outside of the Arabic-speaking world. Al-Wer's perspective in ["Sociolinguistics"] is better; she shows that ultimately constructs need to be interpreted within a context that does justice to the particularities of a given part of the Arabic world, illustrating her point with the interpretation of the ostensibly universal or at least very general "education" variable as a proxy for other, community-immanent variables.

1.3.3 Clash of Traditions

Complementing the two previously defined issues is that academic and cultural traditions provide ready barriers for synthetic perspectives. Within the West, for instance, Carter (1988: 207) attempts to dissociate Arabic linguistics from Arab linguistics. "… 'Arabic linguistics'…. detaches the language entirely from its environment so that it becomes a pure abstraction." On the other hand, Arab linguistics, the legitimate study of the Arabic language, is "… the vast and continuing output of traditional works, both editions of texts and secondary sources, which remain wholly within the historical norms of Islamic scholarship" (ibid.). In Carter's terms, a handbook of Arabic linguistics that has at its core questions about the Arabic language, however defined, is suspect from the start.

To be fair, one of Carter's objections to an Arabic linguistics deserves attention. "Solving" a problem in Arabic within a general linguistic theory runs the danger of importing an issue, a technique of inquiry, a focus on a grammatical construction whose ultimate interest is dictated from outside of Arabic and whose "solution" offers little to those interested in the complex structure of Arabic. At the same time, however, as noted already, trivial an observation though it is, Arabic is simply a language, so linguistic approaches will want to understand it within general theories of language. Moreover, as argued in Sections 1.1.1 and 1.1.2, Arabic itself has unique geographical, social, historical, and cultural properties that have, as it were, pushed the language in directions hardly encountered elsewhere. Linguistic theory can hardly avoid it, even if, in practice, non-Arabicist linguists often do so (see, e.g., criticisms in [Tosco and Manfredi, "Creoles"] or Ryding ["Acquisition"] on the barriers confronting researchers

---

[6] For instance, despite relatively well-documented accounts of "qaf" variation covering thirty years of research in the Arabic world from the Gulf to Morocco (e.g., Sallam 1980; Holes 1987; Haeri 1996; Amara 2005; Hachimi 2007), no studies have synthesized these accounts with a view toward defining the extent to which a common social dynamic lies behind "qaf" usage. It is, for instance, no sociolinguistic accident that the "qaf" variable is of such marginal interest in Nigerian Arabic, a distinctly minority language in northeast Nigeria, that it was not included as a variable in Owens (1998).

of second-language acquisition due simply to lack of language knowledge). It is easy to formulate a solution to this problem: practitioners need to be as well versed in Arabic in all its linguistic ramifications as they are in the methodologies and theories of linguistics. Nonetheless, its implementation implies a commitment of both individual and institutional time and intellectual resources, which are not necessarily easy to come by.

Perhaps more pernicious than the delegitimization of a linguistic approach to Arabic is Mahdi's (1984: 37) admonition to study dialects to be rid of its debilitating influence on the Standard (fuṣḥaa).[7] This perhaps well-intentioned perspective derives most directly from a normative 19th-century tradition (see [Newman, "Nahḍa"], which attempts to lay the blame for the ill learning of the Standard language on the use of dialects and can justify the study of dialects only against a possible benefit for the Standard. Such a perspective is not uncommon in the Arabic world.[8] Leaving aside the cultural and political issues inherent in this position [Suleiman, "Folk Linguistics"], adopting this perspective would necessarily mean excluding Chapters 10, 12, 13, 14, 15, and 22 from this volume while requiring severe reductions in most others, since the dialect is nothing less than the mother tongue. It is not so much an approach foreign to general linguistic inquiry as it is a rejection of the scientific and empirical study of the world, defining in narrow political-cultural terms the goals of research on one of the most ineffable and undefined domains of human experience: language.

1.4 ATTITUDES

The reader may be confused at this point. On one hand, Proposition 1 claims that Arabic is, for the linguist, an intellectual challenge like no other. On the other, this challenge is often met by traditions, theories, academic structures, and attitudes that at best ensure a fragmented understanding of the language and at worst succeed in a holistic characterization of "Arabic" only at the expense of defining whole domains of language experience into nonexistence.

It can be suggested, without exaggerating the professional and even ideological differences that accrue in the study of Arabic, that the only approach that does justice to Proposition 1 is one grounded on radically open-minded empiricism.

---

[7] Mahdi speaks of the sicknesses of the dialects, which require treatment (الامراض التي يجب علاجها). The passage in fact comes in the Introduction to a well-edited edition of *1001 Nights*, which left the original "Middle Arabic" style intact rather than classicizing out its authenticity, as is the current custom (e.g., the version on arabicorpus).

Another popular approach is the regulation of language use by legal fiat. Munṣif al-Marzuqi, who writes an occasional column for Jezira Net, for instance, would (article of Nov. 6, 2011) criminalize the use of what he terms "Creole" Arabic, by which he intends, in the parlance of contemporary linguistics, a codeswitched variety of Arabic (*tajriym istiʕmaal luɣat al-kriyuwl*).

[8] For instance, generally speaking, "Arabic" in Arabic departments in the Arabic world stop with the classical language.

For precedence, one need go no further than the medieval Arabic grammatical tradition itself. The following quote from the mid-4th-/10th-century Zajjaji in one of the earliest works of metareflection in the Arabic tradition. In Chapter 5 (al-baab al-xaamis), he reflects on the nature of linguistic causes. After identifying three types of linguistic causes (pedagogical, analogical, theoretical-speculative; see Versteegh 1995: 89), he approvingly summarizes the approach to language study attributed to al-Xalil ibn Aḥmad, the polymath contemporary of and teacher of Sibawaih (see [Sara, "Classical Lexicography"]). In the passage, Xalil is said to have likened the scholar trying to ascertain the nature of (Arabic/language) to one trying to understand a house construction:

في ذلك رجل مثل حكيم دخل داراً محكمة البناء؛ عجيبة النظم والأقسام؛ وقد صحت عنده حكمة بانيها ،
بالخبر الصادق او بالبراهين الواضحة والحجج اللائحة ، فكلما وقف هذا الرجل في الدار على شئ منها
قال: إنما فعل هذا هكذا لعلة كذا وكذا ، ولسبب كذا وكذا. سنحت له وخطرت بباله محتملة لذلك ، فجائز ان
يكون الحكيم الباني للدار فعل ذلك للعلة التي ذكرها هذا الذي دخل الدار ، وجائز ان يكون فعله لغير تلك
العلة ، إلا أن ذلك مما ذكره هذا الرجل محتمل ان يكون علة لذلك. فإن سنح لغيري علة لما علتله من النحو هو أَلْيَق
مما ذكرته بالمعلول فليأت بها.
ص ٦٦

If a wise person were to enter a house of sound structure, of marvelous form and proportion, whose builder's wisdom appeared correct to him according to reliable information, the unmistakable lines of proof and clear arguments. And each time the man stopped and pondered a part of the house, he said, "[the builder] did it this way for such and such a reason and such and such a cause." That is what occurred to him and appeared reasonable to him. Now it might be that the builder did build it for the reason the man inspecting the house thought, but it is equally possible he did it for another reason, even if the inspector's reason might be correct. If a different grammatical reason should occur to another person than myself, which is more appropriate than my explanation, so let him suggest it. (p. 66)

With this passage, there are obviously interpretative issues that go beyond an introductory chapter. In particular, the passage is enticingly ambiguous as to what a "more appropriate" explanation might be. The history of the Arabic tradition itself shows that an explanation in the 5th/11th century might be more nuanced than one in the 3rd/9th and that one in the 6th/12th century might add further elements [Larcher, "ALT II"], not to mention the classic grammar-internal differences of the Basrans and Kufans (Sibawaih vs. Farraʾ; Owens 1990). It would be a grave mistake, however, to stop with the classical tradition. The recent history of linguistics is marked not only by the continual reappraisal of classic linguistic ideas and traditional issues but also by new theoretical, methodological, and, increasingly, technical advances, many described in this volume, that promise to transform, expand, and enrich the very idea of grammatical explanation to such an extent that a genius such as Xalil, if he were alive today, would be envious.

Xalil's metaphor unmistakably sets a basic ground rule for linguistic research, namely, that no possible explanatory aspect be excluded on a priori grounds. Since explanations are, ultimately, explanations of linguistic substance, facts, observations, summaries of

data, measurements, and reinterpretations of previous explanations, Xalil's approach implies setting no preconditions as to what comes under the purview of Arabic linguistics.[9]

It is in this spirit that the current handbook should be read; it is a reference work that brings together different approaches and scholarly traditions, an invitation to the reader to explore the multifaceted world of Arabic linguistics. The articles in this volume expertly explore the nature of the house of Arabic from many angles. Many argue for specific points of view, others give descriptions of synoptic breadth, while others provide exhaustive overviews of the state of the art. The parts may or may not come together to describe a common structure; they do provide blueprints for a better understanding of it.

NOTE TO REFERENCES

Chapters 9 "Issues in Arabic Computational Linguistics" and 13 "Dialects and Dialectology" have very comprehensive bibliographies. They are, however, too large to be included in their entirety in the print version of the handbook. Rather than edit away this very valuable resource, it was decided to include the complete bibliographies to these two chapters in the online version of the handbook while including a selected bibliography in the print version.

APPENDIX

This Appendix gives basic background information about Arabic as well as a brief discussion of the transcription and transliteration conventions used in this book.

A.1 Maps

The bulk of the native Arabic-speaking population lives within countries with majority Arabic-speaking populations. Sizable non-Arabic minorities include Berbers (Amazigh), with large minorities in Algeria and Libya and up to half the population of Morocco, where in fact Arabic shares its status as official language with Berber. Other minorities are speakers of the various South Arabian languages in Yemen (and a small population in Oman) and Kurds and Aramaic speakers in Iraq and to a lesser degree Syria. Even after the South Sudan, which has few native Arabic speakers, recently split off from the North, the Sudan has a large and diverse linguistic minority population. Finally, Mauretania has a not insignificant

---

[9] An extreme though in today's world by no means uncommon situation is when Arabic needs to be studied in tandem with other languages in the domain of codeswitching [Bentahila et al., "Codeswitching"; also Kossmann, "Borrowing"; Newman, "Nahḍa"].

non–Arabic-speaking population (Wolof, Fulfulde) in the south of the country. Map 1.1 shows countries with majority Arabic-speaking populations. It can be noted that although the main lingua franca of South Sudan, Juba Arabic, historically derives from Arabic, by linguistic measures it is a different language [Tosco and Manfredi, "Creoles"] and therefore is not included on Map 1.1.

Maps 1.2 and 1.3 illustrate the lack of complete isomorphy between political status of a language and the native language of its inhabitants. The Arab League (الجامعة العربية, Map 1.2) comes close to being composed entirely of countries with Arabic as a majority language. There are only two exceptions: Somalia, where the native language of the vast majority of the population is Somali, a Lowland East Cushitic language genetically very distantly related to Arabic; and the Comoro Islands, whose native Bantu language is closely related to Swahili.

Besides being the official language of all countries in the Arab League, Arabic is also the official language of Eritrea (majority native language Tigrinya; Hailemariam 2002: 75), a country with a tiny population of Arabic native speakers. In addition it is, along with French, an official language in Chad, which does have a sizable native Arabic-speaking minority. In these two countries, Arabic attained official status under quite different circumstances and at different times. In Eritrea, for instance, it was during the brief British rule from 1941 to 1952 that Arabic was introduced as the official language, a status it has maintained until today, whereas in Chad Arabic was adopted as an official language well after independence (1960) in the 1990s, and only after considerable debate (de Pommerol 1997).

Finally, Map 1.4 shows that for the most part Arabic-speaking minorities live on the political borders of majority Arabic-speaking countries. Even the exceptions in this regard, the tiny Arabic-speaking populations of Uzbekistan, Afghanistan, and Khorasan in eastern Iran were, at the time of their settlement in the 2nd/8th century, a part of a continuous migration of Arabs into Central Asia. It can be noted that, while from the perspective of genetic linguistics Maltese can be considered a variety of Arabic (Owens 2010), on a sociopolitical basis and as an official language of the European Union it is an independent language.

A.2 Genetic Affiliation of Arabic

While a definitive classification of Arabic within a Stammbaum representation may be impossible [Retsö, "Arabic"], within traditional genetic models the following two models are the most widely discussed (based on Faber 1997: 5, 6):

(Afro-Asiatic)
Cushitic
Omotic
Chadic
Ancient Egyptian
Berber
Semitic:
   Variant 1
      East Semitic: Akkadian, Eblaitic
      West Semitic
         Northwest Semitic
            Canaanite: Hebrew, Phoenician, Moabite
            Aramaic

    South Semitic
      Arabic
      Southeast Semitic
         Modern South Arabian: Jibbali, Mehri, Harsŭsi, Soqotri
         Ethio-Saebean
         OSA: Sabean, Qatabanian, Hadramauti, Minean
         Ethiopian Semitic
Variant 2
  East Semitic
    Akkadian
    Eblaite
  West Semitic
    Central Semitic
      Arabic
      Northwest Semitic
         Ugaritic
         Canaanite: Hebrew, Phoenician, Moabite, Ammonite, El-Amarna
         Aramaic
         DeirAlla
    South Semitic
      Eastern
         Soqotri
         Mehri, Harsŭsi, Jibbāli
      Western
         Old South Arabian
         Ethiopian Semitic
            North Ethiopic: Geʾez, Tigre, Tigrinya
            South Ethiopia
               Transverse SE
               Amharic, Argobba
               Harari, East Gurage
            Outer SE
               n group: Gafat, Soddo, Goggot
               tt group
                  Muher
                  West Gurage

A.3 Transcription and Transliteration Conventions

The representation of Arabic in Latin script is beholden to different conventions. Rather than try to force standardization in this volume, the various systems used are taken over intact in different chapters. Having said this, the editor is strongly biased toward the use of the International Phonetic Alphabet (IPA), or modified IPA symbols, for representing any spoken text. Nothing, moreover, speaks against using it for transliterated written texts, though here other traditions have developed different conventions.

Ultimately, moreover, justification can be asked of each set of conventions. For instance, representing a long "i" as [ī, i:, ii, or iy] implies different phonological interpretations of the nature of vowel length. It can be noted that IPA conventions themselves should hardly be

Uzbekistan

Ankara

Turkey

Teheran

Kabul

Cyprus

Malta

Lebanon Syria Iraq

Beirut Damascus

Israel Amman Baghdad Iran Afghanistan

Jerusalem

Palestine Jordan Kuwait

Algiers Tunis

Rabat Tunisia

Morocco Tripoli

Algeria Libya Egypt

El-Aaiu'n

W. Sahara

Mauritania

Nouakchott Mali

Dakar Senegal

Bamako

Niamey

Niger Chad

Nigeria

Abuja

Cameroon

Yaounde

Cairo

Kuwait City

Bahrain Qatar

Manama Doha Abu Dhabi

Riyadh UAE

Saudi-Arabia Muscat

Oman

Yemen

Sana'a

Khartoum Eritrea

Asmara

Sudan Djibouti

Djibouti

Ethiopia

South Sudan

Juba Addis Ababa

Uganda Kenya Somalia

Kampala Mogadishu

Nairobi

Comoros

MAP 1.1 Countries with Arabic as a majority language.

MAP 1.2 The Arab league.

MAP 1.3 Arabic as official language.

MAP 1.4 Arabic as minority language.

regarded as sacrosanct. The multiexponential phenomenon of "emphasis," for instance, is now represented by C + ^{ʕ}, such as t^ʕ, that is, C + pharyngealization. As the two relevant articles in this volume make clear, however ([Embarki, "Phonetics"; Hellmuth, "Phonology"]), pharyngealization (tongue retraction toward pharynx, pharyngeal constriction) is but one gesture defining the phenomenon and is not necessarily the most prominent one.[10] Equally relevant would be, for instance, a symbol based on the articulatory metaphor developed in the Arabic tradition of likening the flattened tongue body to a plate or pot cover (iṭbaaq, muṭbaq).

In any case, the multiplicity of transcription/transliteration conventions means that the reader's indulgence is needed for the treatment of proper Arabic names, where the same person will appear in difference orthographic guises, according to the conventions of the chapter, Ibn Jinni, Ibn Jinnī, Ibn Ǧinnī, Ibn Ǧinni. Would that he could comment on the matter.

References

Amara, Mohammad. 2005. Language, migration and urbanization: The case of Bethlehem. *Linguistics* 43: 883–901.

Baalbaki, Ramzi. 2008. *The legacy of the Kitaab*. Leiden: Brill.

Brockelmann, Carl. 1908–1913. *Grundriss der vergleichenden Grammatik der semitischen Sprachen*. 2 vols. Berlin: Reuther and Reichard.

——. 1977² (1943). *Geschichte der islamischen Völker*. Hildesheim: Olms.

Carter, Michael. 1988. Arab linguistics and Arabic linguistics. *Zeitschrift für Geschichte der arabisch-islamischen Wissenschaften* 4: 205–218.

de Pommerol, Patrice. 1997. *L'arabe tchadien: émergence d'une langue véhiculaire*. Paris: Karthala.

de Swaan, Abram. 1998. A political sociology of the world language system (1): The dynamics of language spread. *Language Problems and Language Planning* 22: 63–75.

——. 2001. *Words of the world*. Cambridge, UK: Polity Press.

Faber, Alice. 1997. Genetic subgrouping of the Semitic languages. In *The Semitic languages*, ed. Robert Hetzron, 3–15. London: Routledge.

Faiq, Said. 2006. Coherence. In *Encyclopedia of Arabic Language and Linguistics* , vol. I, ed. Kees Versteegh, Associate Editors: Mushira Eid, Alaa Elgibali, Manfred Woidich, Andrzej Zaborski, 427–430. Leiden, Netherlands: Brill.

Ferguson, Charles. 1959. Diglossia. *Word* 15: 325–340.

Hachimi, Atiqa. 2007. Becoming Casablancan: Fessis in Casablanca as a case study. In *Arabic in the city: Issues in language variation and change*, ed. Catherine Miller, Enam Al-Wer, Dominique Caubet, and Janet Watson, 97–122. London: Routledge.

Haeri, Niloofar. 1996. *The sociolinguistic market in Cairo: Gender, class, and education*. London: Kegan Paul International.

Hailemariam, Chefna. 2002. *Language and education in Eritrea*. Amsterdam: Aksant.

Holes, Clive. 1987. *Language in a modernising Arab state: The case of Bahrain*. London: Kegan Paul International.

Kaye, Alan. 1972. Remarks on diglossia in Arabic: Well-defined vs. ill-defined. *Linguistics* 81: 32–48.

---

[10] For instance, Embarki ["Phonetics"], summarizing Al-Ani (1970), identifies four traits of consonantal "emphasis," only one of which involves pharyngeal space.

Khalil, Esam. 2006. Cohesion. In *Encyclopedia of Arabic Language and Linguistics,* vol. I, ed. Kees Versteegh, Associate Editors: Mushira Eid, Alaa Elgibali, Manfred Woidich, Andrzej Zaborski, 430–433. Leiden, Netherlands: Brill.

Lipiński, Edward. 2000. *The Aramaeans, their ancient history, culture, religion*. Leuven: Peeters.

Mahdi, Muhsin (ed.). 1984. *Alf Layla wa Layla*. Leiden: Brill.

Al-Marzuqi, Munṣif. 2011. ʔayy luɣa sayatakallam al-ʕArab al-qarn al-muqbil? *Jazira Net*, November 6, 2011.

Mejdell, Gunvor. 2006. *Mixed styles in spoken Arabic in Egypt*. Leiden: Brill.

Owens, Jonathan. 1990. *Early Arabic grammatical theory: Heterogeneity and consolidation*. Amsterdam: Benjamins.

——. 1995. Language in the graphics mode: Arabic among the Kanuri of Nigeria. *Language Sciences* 17: 181–199.

——. 1998. *Neighborhood and ancestry: Variation in the spoken Arabic of Maiduguri (Nigeria)*. Amsterdam: John Benjamins.

——. 2000. Introduction. In *Arabic as a minority language,* ed. J. Owens, 1–43. Berlin: Mouton de Gruyter.

——. 2009. *A linguistic history of Arabic*. Oxford: Oxford University Press.

——. 2010. Review article: "What is a language?: Review of Bernard Comrie, Ray Fabri, Elizabeth Hume, Manwel Mifsud, Thomas Stolz & Martine Vanhove (eds.), Introducing Maltese Linguistics. Selected papers from the 1st International Conference on Maltese Linguistics, Bremen, 18–20 October 2007." *Journal of Language Contact* (Varia): 103–118.

Owens J., and Alaa Elgibali (eds.). 2010. *Information structure in spoken Arabic*. London: Routledge.

Parkinson, Dilworth. 1991. Searching for modern Fuṣḥā: Real life formal Arabic. *Al-'Arabiyya* 24: 31–64.

Procházka, Stephan. 2006. Arabic. In *Encyclopedia of language and linguistics,* vol. I, ed. Keith Brown, 423–431. Oxford: Elsevier.

Rabin, Chaim. 1951. *Ancient West Arabian*. London: Taylor's University Press.

Sallam, A. 1980. Phonological variation in educated spoken Arabic. *BSOAS* 43: 77–100.

Thomason, Sarah, and Terrence Kaufman. 1988. *Language contact, Creolization, and genetic linguistics*. Berkeley: University of California Press.

van Coetsam, Frans. 2000. *A general and unified theory of the transmission process in language contact*. Heidelberg: Winter.

Versteegh, Kees. 1984. *Pidginization and Creolization: The case of Arabic*. Amsterdam: Benjamins.

——. 1995. *The explanation of linguistic causes*. Amsterdam: Benjamins.

http://arabicorpus.byu.edu/index.php

http://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers#30_to_50_million_native_speakers "List of languages by number of native speakers."

PHONETICS

MOHAMED EMBARKI

2.1 INTRODUCTION

PHONETICS is a linguistic field that studies speech in terms of production, transmission, and reception. To simplify, as Lodge (2009: 2) says, the three domains of speech study the speaker (production), the hearer (reception), and what takes place between the two (transmission). To this purpose, phoneticians use methods derived from the science of physiology for production, from physics for transmission, and from psychology for reception. Thus, as Ladefoged and Johnson note (2010: 2), there are different types of phoneticians.

However, all phoneticians are conscious that the sounds they describe from these three perspectives are utilized to encode linguistic information. The same object, the sounds of words, has a concrete continuous material face and an abstract, cognitive, and categorical face. The first is phonetic; the second is phonological. Hence, even without mentioning it explicitly, phonetic descriptions have one foot in the domain of phonetics and another in phonology. The border between the two disciplines is very narrow, and the interface is multifarious (see Scobbie 2007: 17–52 for a detailed description of interfaces and overlaps between phonetics and phonology). The phonetic–phonology interface is triadic according to Kingston (2007: 401). First, phonetics defines the distinctive features. Second, it explains the phonological patterns. Third, it implements phonological representations.

At the incipience of the phonetic discipline in the arabophone area (as early as the 2nd/8th century), the boundary between phonetics and phonology for the early Arab grammarians was extremely difficult to delimit (see Chapter 3). Their descriptions simultaneously included articulatory, acoustic-auditory, and phonological criteria.

The first section of this chapter will deal with the principal phonetic descriptions of the Arab system produced by the early Arab grammarians of the classical period

(2nd/8th to 5th/11th century). The second part will present the consonant and vowel systems of Modern Arabic, while the third part will deal with the contribution of experimental phonetics to the specificities of the consonant and vowel Arabic systems focusing in particular on (1) pharyngeal consonants, (2) pharyngealized consonants, (3) temporal aspects (vocalic quantity and gemination), and (4) consonant and vowel variation.

As for prosodic structure, which I will not deal with here, one may refer to the reference works published on Arabic (Al-Ani 1970: 89–95; Watson 2002: 79–121; Canepari 2005: 327–329; Ryding 2005: 35–39; see also [Hellmuth, "Phonology"]).

2.2 The Arabic Phonetic System

As Ladefoged (2003, quoted in Chelliah and de Reuse 2011: 251) quite justly remarks, there is "nothing more ephemeral than the sounds of a language. The sounds will live only as long as the language is spoken. When the sounds are those of elderly speakers whose children belong to another world, then soon those sounds will be gone forever. All that can remain are whatever records we have been able to archive." How to collect and interpret phonetic data is dealt with in detail in Bowern (2008: 63–72) and Chelliah and de Reuse (2011: 251–278), both of which recommend using all types of evidence; written data have a big importance here.

Phoneticians and phonologists working with Arabic have at their disposal an abundance of resources bequeathed by the early Arab grammarians. Reference to these works will allow us to compare in this chapter the early phonetic descriptions with those of modern researchers and to verify if more recent data based on the use of sophisticated techniques and instrumentations validate or invalidate the early descriptions.

2.2.1 Classical Arabic

The early Arab grammarians whose work has reached us are not numerous; there are no more than 350 names in the classical period listed by Al-Suyuṭi (963/1556–1009/1601) in his work *al-Muzhir*. From the phonetic descriptions that we know from this period, the consonant system of Classical Arabic included either 28 or 29 consonants, the number varying according to whether a phonetic value is given to the first letter of the alphabet (ʾalif (ا)) or not. The consonantal phonemes were described as ʾuṣūl [usˤuːl] (primary), among which 25 were nonvocalic (صحاح) [ṣiḥaaħ], described according to an *articulatory region* (حيز) [ħajjiz] and an *aperture* (مدرجة) [madraža, "degree"], while 4 were vocalic (lit. "*hollow*" أجوف) [ʔajwaf] because they are characterized by a relatively unhindered exhalation of air (هواءية) [hawaaʔijja] [Sara, "Classical Lexicography"].

Table 2.1 Classification of classical Arabic consonants according to Al-Khalīl (d. 786) in *Kitāb al-'Ayn* (The book of the 'Ayn)

Phonemes	Plain									Hollow
Region	1	2	3	4	5	6	7	8	9	
Letter-sound	ع ح ه	خ غ	ق ك	ش ج ض	ص س ز	ط د ت	ظ ث ذ	ر ل ن	ف ب م	ي و ء ا
Description	guttural	guttural	uvular	arched	apical	alveolar	interdental	apical	labial	vocalic
	حلقية	حلقية	لهوية	شجرية	أسلية	نطعية	لثوية	ذولقية	شفوية	هواءية
IPA	ʕ ħ h	x ɣ	q k	ž ʃ dˤ	sˤ s z	tˤ d t	đˤ θ đ	r l n	f b m	a: ʔ w j

The first region, the most backed articulatory area,[1] is composed of three consonants: "'ayn" (ع) /ʕ/, "ḥā" (ح) /ħ/, and "hā'" (ه) /h/. The second region is composed of two consonants, "khā'" (خ) /x/ and "ğayn" (غ) /ɣ/, which are qualified as *guttural* (حلقية) [ḥalqijja]. The third region consists of two *uvulars* (لهوية) [lahawijja]: "qāf" /q/ (ق) and "kāf" (ك) /k/. The fourth region consists of three *arched* consonants (شجرية) [šažrijja]: "ğīm" (ج) /ž/; "šīn" (ش) /š/; and "ḍād" (ض) /dˤ/. The fifth region contains the *apicals* (أسلية) [ʔasalijja] "ṣād" (ص) /sˤ/; "sīn" (س) /s/; and "zāy" (ز) /z/. The sixth region has three alveolars (نطعية) [nitˤʕijja]: "ṭā'" (ط) /tˤ/; "dāl" (د) /d/; and "tā'" (ت) /t/. The seventh region is composed of *interdentals* (لثوية) [liθawijja]: "ẓāʔ" (ظ) /đˤ/; "thā'" (ث) /θ/; and "dhāl" (ذ) /đ/. The eighth region consists of three *apical* consonants (ذولقية) [đawlaqijja]: "rā'" (ر) /r/; "lām" (ل) /l/; and "nūn" (ن) /n/. The ninth region consists of the labials (شفوية) [šafawijja] "fā'" (ف) /f/; "bā'" (ب) /b/; and "mīm" (م) /m/. The two consonants "'alif" (ا) /a:/-"hamza" (ء) /ʔ/ (these two last consonants are considered to be independent phonemes by Al-Ḫalīl, but not so by his successors), "wāw" (و) /w/, and "yā'" (ي) /j/ are *vocalic* (هواءية) [hawa:ʔijja], and thus they don't have a precise articulation point in the oral cavity (Roman 1977). Table 2.1 synthetizes these different articulations (the International Phonetic Alphabet [IPA] symbol corresponds, however, to the articulation of the sound–letter in Modern Arabic).

Besides the 28 or 29 normative phonemes of "the language of Arabs" (i.e., 'Al-Arabiyya), the treatises of the early Arab grammarians gave considerable attention to variation. Owens (2001: 422) qualifies Sībawayhi's *al-Kitāb* (177/793) as a grammatical work that has institutionalized the variation the most. For example, Sībawayhi describes 29 phonemes as "primary" (*'uṣūl*), to which are added six articulations "secondary" (*furū'*), referred to as "good" (*mustaḥsana*)—among which can be cited the light "nūn" (ن) (خفيفة) [xafi:fa] (/n/ of assimilation), the "šīn" (ش) /ʃ/ pronounced as "ğīm" (ج) /ž/, the "ṣād" (ص) /ṣ/ pronounced as "zāy" (ز) /z/—and seven articulations referred to as "bad" (*ğayr mustaḥsana*)—among which are the "ğīm" (ج) /ž/ pronounced either as "kāf" (ك) /k/ or as "šīn" (ش) /š/, the "ṭā'" (ط) /tˤ/ and the "ṣād" (ص) /sˤ/ pronounced,

---

[1] The articulatory descriptions of the Arabic grammarians always conventionally started from the back of the articulatory tract and worked their way forward to the lips.

respectively, as "tā'''" (ت) /t/ and "sīn" (س) /s/, as well as the ḍād (ض) /dˤ/ and the "ḍāʔ" (ظ) /ðˤ/, which merge in "thāʔ" (ث) /θ/.

Ibn Jinnī's work *al-Khaṣa:ʔiṣ* (1002) was the first to describe in detail the Classical Arabic vowel system, including the clear distinction among the three short vowels /i, u, a/ and the three long vowels /i: u: a:/. He was also the first grammarian to rid himself of the phonetic spelling trap to describe them.

The Classical Arabic phonological system clearly reveals the membership of the language in a Semitic family whose principal characteristic is a reduced vowel system limited to three cardinal qualities (with length opposition) and a rich consonantal system often exceeding 29 consonants. A second characteristic is the organization of the consonants according to the morphophonological constraints of "root" and "pattern" [Ratcliffe, "Morphology"; Hellmuth, "Phonology"]. A third characteristic is the triadic organization of certain consonantal oppositions (Watson 2002: 2–3). The triad is composed of three consonants sharing several properties and the most frequent pattern of one voiceless consonant, one voiced consonant, and a corresponding pharyngealized one, whether voiced or not.

For Classical Arabic, Watson (2002: 3) gives a stop (also termed "plosive") triad composed of /t, d, tˤ/, a sibilant triad composed of /s, z, sˤ/, and a fricative triad composed of /θ, ð, ðˤ/. It should be noted that the consonants that constitute a triad cannot figure in the same etymon of the word (Boudelaa and Marslen-Wilson 2001; Watson 2002), a point discussed in some detail in this volume [Hellmuth, "Phonology"]. This organization in triads allows for a better understanding of the complex relations between consonants as well as their evolution in Modern Arabic. The pharyngealized stop consonant of Modern Arabic /dˤ/ is inserted in a palatal triad. The discussion of the evolution of the pronunciation of the uvular stop /q/ allows for a better understanding of the nature of the relation that it maintained and that it continues to maintain with /k/ on one hand and /ɣ/ on the other (Jakobson 1957; Bonnot 1976; Roman 1981).

2.2.2 Modern Standard Arabic

2.2.2.1 *The Consonant System*

Modern Arabic has 28 consonants (summarized in Table 2.2). The consonants on the left of the column are voiceless; the ones on the right are voiced.

The system is based on a set of basic contrasts. The differences with the consonant system of Classical Arabic described by the early Arab grammarians seem to be minimal. However, the articulation of some segments seems to have evolved from Classical Arabic to Modern Arabic. We can cite, inter alia, the "ğīm" (ج) /ž/, the ḍād (ض) /dˤ/, the "šīn" (ش) /ʃ/, and "sīn" (س) /s/ (Beeston 1962; Murtonen 1966; Al-Wer 2003; Embarki forthcoming).

Modern Arabic is characterized by a rich consonantal system that places it slightly over the average of 22.8 consonants derived from the University of California–Los

Table 2.2 Consonantal system of Modern Standard Arabic

	Bilabial	Labio-dental	Dental	Dento-alveolar	Post-alveolar	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	b		t, d				k	q		ʔ
Nasal	m			n						
Trill			r							
Fricative		f	θ ð	s z	š ž			x ɣ	ħ ʕ	h
Approximant						j	w			
Lateral approximant				l						
Pharyngealized plosive			tˤ dˤ							
Pharyngealized fricative			ðˤ	sˤ						

Angeles (UCLA) phonological segment inventory database (UPSID$_{317}$)[2] (Maddieson 1984). As Newman (2005: 185) notes, Arabic is distinguished by the presence of certain consonants (up to ten) that are not common to other UPSID-based languages—such as the pharyngealized stops and fricatives.

Arabic consonants are opposed to one another in terms of different articulatory manners and places of articulation. Arabic has a large number of places of articulation: (1) five different places of articulation for stops against an average of three in other UPSID-defined languages; and (2) seven different places of articulation for fricatives against an average of four in other languages. Eight pairs of obstruent consonants contrast in terms of voicing, and there is an overall of 15 voiced consonants and 13 unvoiced. The effects of voicing on consonants have been well described in the literature (Port et al. 1980; Mitleb 1984a). Among Arabic obstruents, the duration of the voiceless is longer than that of the voiced cognate. This temporal pattern is reversed for the contiguous vowel: the duration of the vowel before a voiceless segment is short, whereas that of the vowel accompanying a voiced consonant is long. Mitleb's (1984a) study, carried out on eight speakers of Jordanian origin, confirms this temporal pattern.

The frequency of phonemes reveals several functional aspects of language. For instance, it is linked to the linguistic change and to the representation of the mental lexicon (Bybee 2003: 11–12; [Hellmuth, "Phonology"]). It appears that the most frequent words are pronounced differently from the less frequent words and that they are more subject to reduction phenomena (Gordon 2007: 73). A brief exploration of the Arabic lexicon in speech situations shows that the frequency of anterior consonants

Table 2.3 Frequency of 26 consonants in modern Arabic

Sound	l	m	n	r	t	b	ʕ	d	f	s	h	ʔ	q
%	11.77	6.18	5.14	4.66	4.49	3.35	3.34	3.11	2.56	2.53	2.50	2.06	2.13
Sound	k	ħ	ž	sˤ	tˤ	š	x	đ	dˤ	z	θ	ɣ	dˤ
%	1.85	1.79	1.35	0.96	0.95	0.91	0.80	0.67	0.65	0.64	0.53	0.36	0.20

Source: Newman (2005: 191).

(labials, dentals, and dento-alveolars) is important. Indeed, 9 of the most frequent 10 consonants are articulated at the front of the vocal tract (see Table 2.3). On one hand, this contrasts strongly with the common perception of Arabic as a "guttural language," and, on the other hand, it explains the coarticulatory patterns of the language (a better motor control of the front of the vocal tract).

2.2.2.2 *The Vowel System*

Modern Arabic has one of the most elementary types of vowel systems, composed of three cardinal vowels /i u a/, which are common to a very large majority of natural languages. A reduced vocalic system correlates with a richer consonantal system (Flemming 2001).

The frequency in the lexicon of these three vowels is unequal: the vowel /a/ has a frequency that slightly exceeds 60%; /i/ just below 25%; and /u/ slightly below 15% (Newman 2005: 205). This order probably follows a general tendency in the languages of the world.

The three cardinal vowel qualities are doubled in number in Modern Arabic by a contrastive lengthening opposing the short vowels to the long vowels (Cantineau 1960; Watson 2002: 22–23). All in all, Modern Arabic has six vowel /i iː u uː a aː/. These vowels contrast essentially via two parameters: (1) the height of the body of the tongue or high vs. low, and (2) the front-back position of the tongue or front vs. central vs. back. The vowels /i/ and /iː/ are high front vowels, /u/ and /uː/ are high back, while /a/ and /aː/ are low central vowels. Lip rounding is not contrastive in Arabic; only high back vowels are produced with slight lip rounding. As explained in Section 2.3.4, these vowels show strong variation, determined by the linguistic context, prosodic position, and geographical origin of the speakers.

2.3   THE CONTRIBUTION OF EXPERIMENTAL PHONETICS

As Heselwood and Hassan (2011) indicate in the introduction of their collective work *Instrumental Studies in Arabic Phonetics*, the early Arab grammarians gave very detailed phonetic descriptions of Arabic sounds. From the end of the 12th to the beginning of

the 13th century, they supposedly produced the first representation of the vocal tract with the main consonantal articulations (ibid.). This representation is distinguished by its modernity because it shows precisely both place of consonantal articulation and the articulators. Figure 2.1 provides a diagram of the vocal tract borrowed from Bakalla (1982) and quoted by Heselwood and Hassan.

Although phonetics was not yet a structured discipline and tools for observation did not exist, the early Arab grammarians still gave us ample, precise indications about the articulatory characteristics of consonants as well as on their acoustic and perceptive properties (Bonnot 1976; Heselwood and Hassan 2011). Thanks to the most modern equipment, researchers today can compare their own observations with those of the early Arab grammarians.

This, however, is not without disagreement: for example, Sībawayhi distinguished between the opposing *mažhu:ra* lit. "made loud" and *mahmu:sa* "whispered," which many modern phoneticians and phonologists translate with voiced versus voiceless, with the exception of Jakobson (1957), who uses *lenis* versus *fortis* to describe this contrast. Instead of doing a chronological presentation of the main research in Arabic experimental phonetics, I have chosen to present the most important conclusions of the works, which pertain to (1) pharyngeal consonants, (2) pharyngealized consonants, (3) length, and (4) consonantal and vocalic variation.
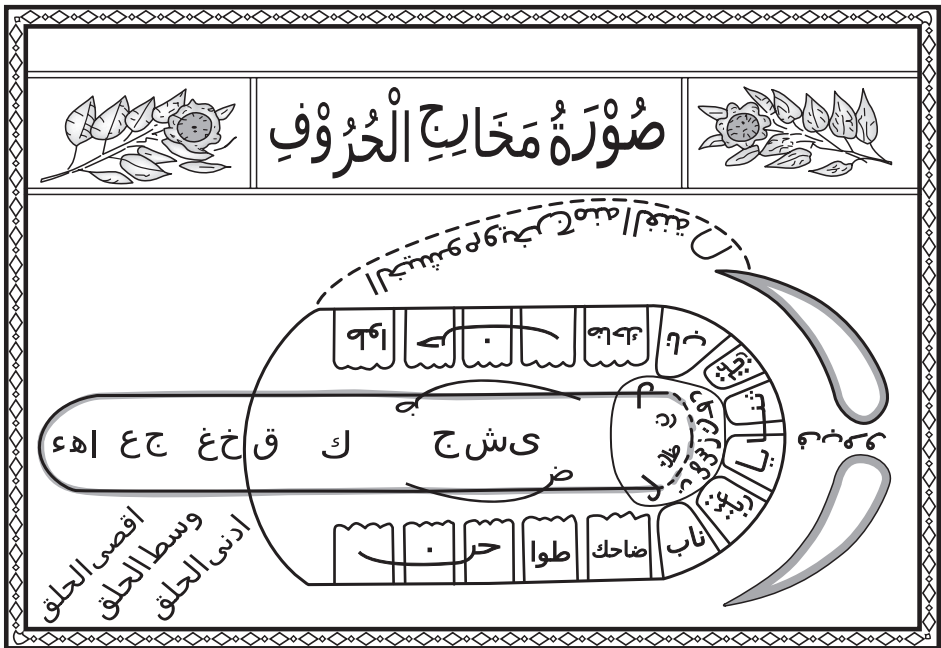
صُوْرَةُ مَخَارِجِ الْحُرُوْفِ

FIGURE 2.1 Vocal tract diagram titled *Ṣūrat makhārij al-ḥurūf* (Picture of the points of articulation of the letters) from *Miftāḥ al-ʿUlūm* (The key to the sciences) by Al-Sakkāki. Dotted line indicates the nasal passage with a nostril above the lip (from Bakalla 1982: 87, quoted by Heselwood and Hassan 2011: 7).

2.3.1 Pharyngeal Consonants

Modern phoneticians qualify the two guttural consonants (حلقية) "'ayn" (ع) /ʕ/ and "hā'" (ح) /ħ/ in Al-Khalīl as pharyngeal. However, not all researchers agree that they belong phonologically to a so-called natural class (see Zawaydeh 2003 for an extensive review of that point). Yet Sībawayhi does not use the same terms to describe these two consonants. The sound /ħ/ is described as fricative (raxw), while the sound /ʕ/, which is situated between the stop (ʃadi:d) and fricative, is produced with [tardi:d], according to Sībawayhi's terminology. Ghali (1983) and, following him, Hassan (2011) use the quality (taraddudijja) "frequentative" to designate the consonant /ʕ/.

Al-Ani (1970) is considered the first experimental work in Arabic phonetics. His research is based on cineradiographic data (x-ray), which give accurate images of the surface of the vocal tract, lips, tongue, uvula, and pharyngeal movements combined with acoustic data to describe the consonants and vowels of Modern Arabic. Relying on the productions of four Iraqi native speakers, Al-Ani (ibid., 59–60) confirms Sībawayhi's description of /ħ/ as a fricative voiceless consonant. If he accepts the pharyngeal place of the consonant /ʕ/, he describes it, however, as a voiceless stop in all positions (initial, medial, final) whether singleton or geminated (ibid., 62–63).

If we carefully examine the mid-sagittal sections from the cineradiographic films (Al-Ani 1970 72–74) and the spectrograms (ibid., 65–71), we can observe that /ʕ/ presents a constriction lower in the pharynx with the body of the tongue in more retracted position compared with /ħ/ (see Figure 2.2). On the acoustic level, /ʕ/ does not have the profile of a stop or even that of a fricative as is /ħ/; one can, however, see that it is clearly voiced.

The stop articulatory manner of /ʕ/ described by Al-Ani was not often followed by other phoneticians. Using the same techniques (x-ray and acoustic measurements), Ghali (1983: 440) chooses the feature frequentative for /ʕ/, which Sībawayhi also uses, and assigns it the "trill" articulatory manner. Besides / ʕ /, Ghali (ibid., 441) classifies four further consonants in the trill category: the alveolar /r/; the two uvulars /x/; the /ɣ/; and the glottal /ʔ/.

In the last two decades, more sophisticated technologies such as ultra-fast imaging have been used. Some researchers successfully applied these techniques to perfect our knowledge of Arabic consonants. Zawaydeh (2003) uses the endoscopic technique to visualize articulatory adjustments during the production of these two consonants in Jordanian Arabic.

The results indicate that, during the production of /ħ/ and /ʕ/ as well as during the production of pharyngealized consonants, the distance between the epiglottis and the pharyngeal wall is reduced (Zawaydeh 2003: 287). These results are similar to those obtained by Ghazeli (1977), who uses cineradiography to study pharyngealized consonants. In a recent study of Iraqi Arabic, Hassan et al. (2011) employs ultra-fast laryngoscopy (an imaging technique using endoscopy), combined with electroglottography (EGG). This technique captures vocal fold vibrations by positioning two electrodes on the neck on both sides of the thyroid cartilage.
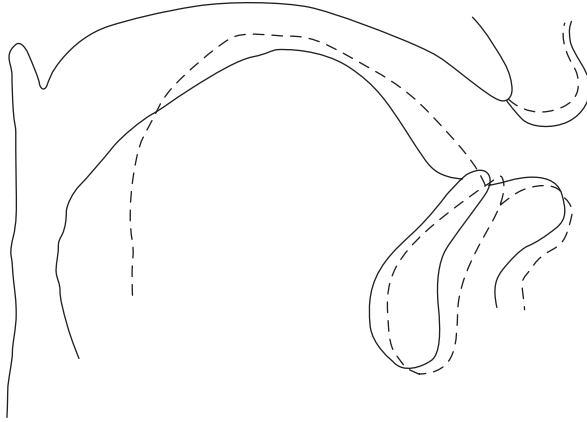
FIGURE 2.2 X-ray tracing of the articulation of /ħ/ (dotted line) and /ʕ/ (plain) in the context of /Ci/ (from Al-Ani 1970: 72).

These physiological data were combined with acoustic data to complete our knowledge of the features of /ħ/ and /ʕ/. While confirming their pharyngeal place of articulation, Hassan et al. (2011: 834) confirm that these two consonants are pronounced by Iraqi speakers as aryepiglotto-epiglottal fricatives, transcribed as voiceless /ʜ/ and voiced /ʢ/, and are considered as variants of /ħ/ and /ʕ/, respectively. Heselwood (2007: 5), relying on Laufer (1996: 114), indicates that /ʕ/ in the production of 21 speakers from 11 different Arab countries is never pronounced as a fricative consonant; it does not inherit this characteristic except through the fact that it is phonologically paired with /ħ/, which is a real fricative.

On the basis of articulatory and acoustic data, Heselwood (2007: 9–28) describes /ʕ/ as a "tight approximant," which he proposes to represent using the following symbol /ʕ̞/. On the basis of acoustic and articulatory data from Moroccan speakers, Yeou and Maeda (2011: 155) conclude that /ħ/ and /ʕ/ are real approximants, since, unlike certain fricatives such as /s/, the two pharyngeals have a larger articulatory constriction and the turbulence is present only for the voiceless consonant.

2.3.2 The Pharyngealized Consonants

Arabic has a specific phonological contrast that opposes plain dental or dento-alveolar consonants to their pharyngealized cognates. Modern Arabic has four pharyngealized consonants /tˤ dˤ ðˤ sˤ/; some modern Arabic dialects have slightly more, while others have less.

Ferguson (1956) shows that allophonic phraryngealized variations exist in Modern Arabic for the /l/. I exclude the consonant /q/ from this correlation and adopt Bonnot's (1976) point of view, who dedicates a long chapter to the relation between the two stops /k/ and /q/ and concluded, based on articulatory and acoustic data, the absence of the pharyngealized feature during the production of the consonant /q/.

The pharyngealized consonants /tˤ dˤ ðˤ sˤ/ existed in Classical Arabic with presumably a slightly more backed place of articulation and a different articulation manner for some of them (see Al-Wer 2003: 28–29 for the evolution of /dˤ/; see Roman 1981 for the evolution of the emphatic among the guttural consonants). These consonants were often designated by a plurality of Arab terms such as "istiʕla:ʔ," "tafxi:m," "'itˤba:q," or "'iħsˤa:r," which modern linguists translated by "emphatic" (see Bonnot 1976, esp. chapter dedicated to emphasis, 84–118).

The main dental or dento-alveolar articulation of these consonants is not a major point of disagreement among researchers, but the same cannot be said of their pharyngealized secondary articulation. According to Ladefoged and Maddieson (1996: 365–366), the place of constriction of the secondary pharyngeal articulation is formed midway between the uvula and the epiglottis. The sagittal sections presented by Al-Ani (1970: 57–58; see Figure 2.3) show that the back of the tongue has a rather flat position and that its root has a more backed position for the pharyngealized consonant compared with its non-pharyngealized counterpart.

The narrowing at the origin of the constriction seems to be produced in the median region of the oropharynx. The acoustic data deal more with the effects of the adjacent vowel showing that the secondary pharyngealized articulation leads to a rise of the frequency of the first formant, F1, and a lowering of the frequency of the second formant, F2 (Al-Ani 1979: 44–56).

Shahin (1997) interprets this acoustic pattern in phonological terms. The observed rise of F1 associated with pharyngealized consonants is shared by other guttural consonants (the glottals, pharyngeal, and uvulars); the author regroups them in a class called *pharyngealization harmony*. The lowering of F2, however, concerns only pharyngealized consonants, and Shahin proposes regrouping them in a different class called *uvularization harmony*.

Ghazeli's (1977) study using the same instrumentation as Al-Ani (1970) shows that the main characteristic of Arabic pharyngealized consonants on the articulatory level is a retraction of the root of the tongue and a flattening of its posterior part in the shape of a plateau, a tightening of the pharyngeal cavity above the epiglottis, and a slight labial protrusion. Based on the cineradiographic data of a Saudi speaker, Bonnot (1976: 369) determines that the constriction of a pharyngealized consonant goes from the uvula region up to the deepest level of the pharynx. Compared with its non-pharyngealized counterpart, a pharyngealized consonant is distinguished by a more backed place of articulation and a superior articulation strength as well as by a slight increase in its length and a shortening of the adjacent vowel (ibid., 472–473). These data are confirmed in Elgendy's (2001) study on pharyngealization.

Several chapters of Hassan and Heselwood (2011) examine the articulatory and acoustic properties of pharyngealized consonants using modern techniques such as nasoendoscopy, videofluoroscopy, electromagnetic midsagittal articulometry (EMA), and ultrasound. In addition to the retraction of the tongue body and the flattening of its posterior part, these data show that pharyngealized consonants are different from their non-pharyngealized counterparts in the volume of buccal and pharyngeal resonance
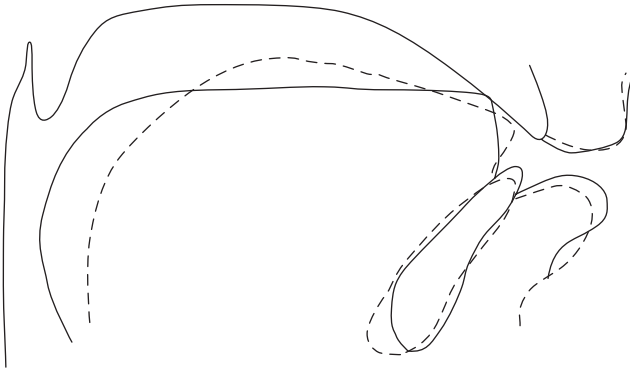
FIGURE 2.3 X-ray tracing of the articulation of /t/ (dotted line) and /tˤ/ (plain) in the context of /Ci/ (from Al-Ani 1970: 57).

cavities; the adjustments of the body of the tongue and its root; the height of the lower jaw; and the different positions of the hyoid bone, the epiglottis, the aryepiglottic cords, and the larynx.

According to Al-Ani (1970), acoustic data focus the effects of pharyngealized consonants on their phonetic environment. These effects manifest themselves through an important modification of the first two formants of the vowels, shown by a significant increase of F1 and a substantial decrease of F2. Jongman et al. (2011: 89) report significant effects on the third formant, F3, in the production of 12 Jordanian speakers. The frequency of F3 increases significantly when the vowel is placed next to a pharyngealized consonant.

Bonnot (1976: 451) shows that F1 and F2 are very close in a pharyngealized consonantal environment compared with a non-pharyngealized environment. Embarki et al. (2011b) analyzes the influence of pharyngealization on Jordanian, Kuwaiti, Moroccan, and Yemeni speakers by comparing $V_1C^ˤV_2$ sequences, where C is /tˤ dˤ ḏˤ sˤ/, with similar V_1CV_2 sequences containing the non-pharyngealized cognates /t d ḏ s/. The frequency measures of the first two formants F1 and F2 as well as the distance F2 – F1 (Fv) taken at three different landmarks of the vowel (*onset, midset,* and *offset*) confirm the frequency differences indicated in the literature, that is, the increase of F1, lowering of F2, and closeness of the two formants (ibid., 146).

The study also shows that the values of Fv in the environment of a non-pharyngealized consonant are on average greater by 348 Hz than the corresponding pharyngealized one. Also, the differences of Fv between the two contexts (non-pharyngealized–pharyngealized) are stronger at the onset of V2 than at the offset of V1 (Embarki et al. 2011b: 147). Embarki et al. conclude that the influence exercised by a pharyngealized consonant on the vocal environment is stronger than its non-pharyngealized cognate. Jongman et al. (2011: 88–89) show that the effects of the pharyngealized consonant in initial or final position of a word significantly impact the three first formants of the adjacent vowel, influencing the vowel in a constant way from

the onset to the offset via the midset. While confirming the effects of pharyngealization on the two first formants of the adjacent vowels, Ghazeli (1981: 275) shows that the direction of the pharyngeal coarticulation has a more regressive or persistent (carry-over) (left–right) nature than a progressive or anticipatory (right–left) one.

Ali and Daniloff (1972) use cinefluographic data in Iraqi Arabic to highlight this characteristic: a left–right effect being more important than a right–left effect both in the magnitude of the retraction gesture of the tongue and in the number of segments affected by the spread of pharyngealization. However, in a study based on articulatory data (EMA) of a Tunisian speaker, Embarki et al. (2011a: 210) indicate that the effect of a pharyngealized consonant in C_2 position (medial position) starts with the first vowel of the word (V1) and continues above the stationary part of the vowel of the second syllable (V2), indicating that the pharyngealization has carryover as well as anticipatory effects. The acoustic data of the study of Jongman et al. (2011: 91) show that the pharyngealized consonant placed in the final position of a word has significant effects on the non-pharyngealized consonant placed at initial position. On the other hand, the same pharyngealized consonant placed in initial position of a word has no significant effect on the final consonant. This shows that the anticipatory effects (right–left) of pharyngealization are more important than the carryover effects (left–right).

Another interesting piece of data in Ghazeli's (1981) work is the spread of pharyngealization in relation to morphemic boundaries. Ghazeli confirms that the spread of phrayngealization effects stops at the boundaries of the word (ibid., 275). These effects do not seem to cross from one word to the other.

The coarticulatory effects of the pharyngealized consonant on adjacent vowels were measured, among other ways, by a linear regression (i.e., the locus equation; see Lindblom 1963) quantifying the coarticulation degree between the consonant and the vowel between extremes: 0 for a null coarticulation; 1 for a maximal coarticulation. Yeou (1997) shows that the value of the slope of a pharyngealized consonant in Modern Arabic produced by Moroccan speakers is weaker than the value of the slope of its non-pharyngealized counterpart. This same pattern is confirmed in Embarki's et al. (2011a) study of the production of pharyngealized consonants in Modern Arabic and Arabic dialect of 16 native speakers from four different countries, Jordan, Kuwait, Morocco, and Yemen.

2.3.3 Duration

Duration is a phonetic parameter that is specific to all linguistic units, consonants, vowels, syllables, words, and sentences. As Coates (1980: 4) says, "Time is vital in the understanding of phonological processes and processing. *A fortiori*, it is vital in phonological representation too."

Traditionally, the rubric *duration* deals with contrastive length that pertains not only to vowels in a large number of natural languages (Ladefoged and Maddieson 1996: 320)

but also to consonants, although to a lesser degree. Arabic is among the languages that use quantity contrast both for consonants and vowels. In addition to these two categories, I will include voice onset time (VOT), which is also mainly a temporal phenomenon. However, I will not talk about the effects of consonantal voicing on the duration of both the consonant and the vowel. In this respect I refer the reader to the works of Port et al. (1980) and Mitleb (1984a).

2.3.3.1 *Vowel Length*

As for vowels, most languages using vowel quantity have a duration opposition between two vowel categories, the short and the long ones. Ladefoged and Maddieson (1996: 320–321) explain that quantity can oppose three or even four vowel categories in some languages. The ratio of duration between vowels (the duration of the long vowel divided by the duration of the short vowel) varies enormously among languages. Some languages use a low ratio, for example 1.3, while others use a significantly longer ratio, such as 3.2 (Lehiste 1970).

In Arabic, studies show that the quantity ratios between vowels vary a great deal. Al-Ani (1970: 75) shows that the relative duration of short isolated vowels is 100 to 150 ms, while with the long ones it is 225 to 350 ms, which makes the ratio long to short more than two to one. Port et al. (1980) present a ratio of 2.6 for Egyptian, Iraqi, and Kuwaiti speakers. Mitleb's (1984b: 231) study on Jordanian Arabic showed that the Arabic long vowel is 65% longer compared with its short counterpart, a 1.5 ratio. Belkaid (1984) presents a ratio slightly greater than 2 for speakers of Tunisian origin. Studying three speakers of different dialectal origin, Abou Haidar (1991) presents varying ratios, but an average of around 2.6. Alioua (1992) finds a mean ratio of 2 for three Moroccan speakers. Jomaa (1994) proposes an intermediate ratio of 2.4 for several dialects. These ratios are, nonetheless, less than those of Modern Arabic and are between 1.3 and 2, with higher relationships in eastern dialects and lower ones in dialects from the Maghreb (Jomaa 1994).

The contrastive vowel length is conveyed essentially through duration (Lehiste 1970). In Al-Ani's (1970: 22–25) study, it seems that quantity contrast is accompanied, in an insignificant way, by vowel quality (see Section 2.2.2.2). The length ratio between short and long vowels is affected by other linguistic parameters. For example, Mitleb (1984b) shows differences that are inherent to the nature of the adjacent consonant, whether it is singleton or geminated (see the previous discussion). De Jong and Zawaydeh (2002: 319) show that long, stressed vowels in Arabic were lengthened 120% by native Jordanian speakers in contrast with their short counterparts. Canepari (2005: 319) indicates that, in a unstressed position, long vowels in Arabic are realized like semi-long vowels.

However, the sensitivity is limited by the theory of acoustic invariance, which is based on the hypothesis that invariable acoustic properties correspond to a segment or to phonetic features, independently of context, speaker, and language (Lahiri et al. 1984; Pickett et al. 1999). Thus, according to Zawaydeh and de Jong (1999), contrastive vowel length is maintained in Arabic fast speech. Port et al. (1980) and Mitleb (1984b: 233)

indicate that the length domain in Arabic is determined at the segmental level, while in other languages it is determined at the syllable level such as in Swedish or at the word level such as in English.

2.3.3.2 *Gemination*

Consonant length is treated in languages in terms of gemination. Duration and gemination refer to different aspects of articulation. Quantity is a matter of length, while gemination applies to the repetition of the same articulation. The question of whether Arabic consonants are really geminated or simply long has been discussed by researchers.

In his study of pharyngealization, Bonnot (1976: 225) uses cineradiographic data to prove that the closure release of the geminated /tˤ/ occurs only at the final occlusion, which leads him to conclude that gemination in Arabic is not present with stops in a two-phase articulation but rather in one single phase (ibid., 450). According to Bonnot, the most important criterion is an increase in duration, and the so-called geminated consonants are in reality long consonants.

Languages such as Arabic, which combine both vowel and consonant quantity, are less numerous. In Modern Arabic, the distribution of the 28 consonants is completely regular, with each consonant occupying three positions: initial, medial, and final. All consonants can be singleton or geminated (Kaye 2009: 563). Contrary to the majority of languages where stops are geminated preferably in the medial position of the word (Ladfoged and Maddieson 1996: 92–93), dentals and dento-alveolars in Arabic can be geminated in initial position as well, as with all so-called solar consonants.

Gemination is thus phonological in Arabic, and it is highly contrastive in distinctions of morphological nature (Watson 2002; see, under the morphology subsection of the chapter of Arabic, Kaye 2009: 572–574; [Ratcliffe, "Morphology"]). Al-Ani (1970: 75–77) shows that the duration of geminated consonants increases until it reaches twice the duration of its singleton counterpart. This ratio between a consonant and its geminated counterpart varies slightly in the literature to the point that sometimes overlaps are noticed between the lowest average durations for a geminated consonant and the highest durations for its singleton counterpart.

Bonnot (1976) indicates that the geminate pharyngealized stop /tˤtˤ/ is distinguished from its singleton counterpart /tˤ/ essentially through the duration of the complete closure, which is longer for the geminated one than for its singleton counterpart, with overlapping zones. Al-Ani (1970: 33) indicates that the duration of the geminated consonant is twice that of its singleton counterpart. In Arabic, the geminated consonant can be preceded by a short vowel (V) or a long vowel (V:); the temporal pattern can also be globally affected without reducing a long vowel to the point of confusing it with a phonologically distinct short vowel (Hassan 2003: 46). Khattab (2007: 156) shows that the geminate–singleton ratio in Lebanese is higher when the preceding vowel is short (2.5) compared with a long vowel context (2.09).

Hassan (2003) indicates that the temporal pattern is different when a long vowel is followed by a singleton consonant (V:C) compared with the short vowel context followed by a geminate (VC). Mitleb (1984b) finds the same distributional pattern.

Bonnot (1976: 235) uncovers a difference in the closure duration of the consonant when it is preceded by a short vowel or a long vowel. In the first case, the closure can have a longer duration of up to 50 ms compared with the second case.

Basing his findings on electromyography (EMG) data for Estonian and English, Lehiste et al. (1973: 146–147) indicate that singleton and geminated consonants are different through the duration of the closure and the amplitude of the peaks. Al-Tamimi and Khattab (2011: 214–215) show that the differences between singleton and geminated consonants of Lebanese speakers included, in addition to the duration, other acoustic parameters such as F0, intensity, and the degree of voicing of the consonant. Bonnot (1976) notices articulatory differences illustrated by the lowering of the lower maxillary. Indeed, the lower maxillary is lowered less during the production of the singleton consonant /tˤ:/ (ibid., 255, 346). Bonnot also indicates that the position of the tongue varies for a singleton pharyngealized consonant and its geminated counterpart. Unlike the front part of the tongue, which does not show differences, the geminated consonant causes a tightening of the posterior part of the back of the tongue (ibid., 371). Other results show that the geminated stop is distinguished from its singleton counterpart through the VOT duration.

2.3.3.3 *VOT*

VOT is defined as the temporal difference between the release of the complete closure and the onset of quasi-periodical vibrations of the vocal folds. This parameter applies only to stop consonants. It is described as positive when the first voiced periodical resonance starts immediately after the release of the consonant, as is the case of voiceless stops. It is described as negative when the vibrations of the vocal folds begin before the closure release, as is the case with voiced stops. Lisker and Abramson's (1964) classic study, based on the examination of stop consonants in 11 languages, showed that this temporal interval, which is the VOT, allowed for the distinction among three categories of stops in those languages: (1) voiceless unaspirated stops, with a positive VOT between 0 and 25 ms, or *short lag*; (2) voiceless aspirated stops with a positive VOT of 60 to 100 ms, or *long lag*; and (3) voiced stops, with vibrations beginning before the closure release. Some languages use the three patterns to oppose stops, while other languages use a binary opposition only between the two patterns.

Lisker and Abramson (1964) show that the VOT duration varied according to the place of articulation of the consonant: longer for the velars, shorter for the labials, and intermediate for dentals. Cho and Ladefoged (1999: 213) mention six criteria as the origin of the VOT variation, including the cavity volume in front of and behind the constriction, the movement of the articulators, and the contact zone between the articulators.

Al-Ani (1970: 76) indicates that Arabic has a negative VOT for voiced consonants; the duration of this prevoicing varies between 50 and 300 ms according to the position of the consonant (initial, medial, or final) and its nature (singleton or geminated). On the other hand, Arabic has a positive VOT for voiceless consonants that varies between 20 and 40 ms for unaspirated and 35 and 60 for aspirated variants.

In their study of Lebanese Arabic, Yeni Komshian et al. (1977: 38) indicate that stops are characterized by a binary VOT, a long prevoicing, or negative VOT for the voiced consonants /b d dˁ/ varying between 40 and 80 ms and a short interval or positive VOT for the voiceless consonants /t tˁ k q/ between 15 and 35 ms. This study did not examine the VOT of the glottal consonant /ʔ/. Al-Ani (1970: 60–62) describes the latter with a short VOT of 15 to 20 ms.

Al-Ani's (1970) study shows VOT differences according to pharyngealization contrast. The VOT of /t/ is longer than that of its pharyngealized /tˁ/ counterpart, by 40–45 ms for the first and only by 20–30 ms for the second (Al-Ani 1970: 44–45). Yeni-Komshian's et al. (1977: 42) results show differences between pharyngealized consonants /tˁ dˁ/ and corresponding non-pharyngealized /t d/ presented in the form of overlapping zones of 0 to 30 ms. The VOT of voiceless consonants /t tˁ/ appears to be different; it is clearly shorter for the pharyngealized consonant (ibid., 40). Ghazeli (1977) confirms this distribution: the VOT of /tˁ/ is positive although very short (15 ms) compared with the double (30 ms) for the non-pharyngealized consonant /t/. Zeroual et al. (2007: 400) also show that the voiceless pharyngealized stop /tˁ/ has a positive VOT, 14 ms shorter than its non-pharyngealized counterpart /t/ (48 ms).

In Arabic phonology, the question of whether there is a phraryngealized relation that links the consonants /k/ and /q/ is amply discussed (see Section 2.2.1). VOT seems to be one of the elements taken into consideration. Al-Ani (1970: 32) found the same pattern in Iraqi speakers: a longer VOT for /k/ between 35 and 44 ms; and a shorter VOT for /q/ varying between 20 and 26 ms. On the other hand, Yeni-Komshian et al. (1977: 42) presented averages of positive VOT that seem similar for the two consonants: 25 to 30 ms for /k/; and 25 to 35 ms for /q/. This is probably because Lebanese speakers utter a /k/ that is close to a uvular consonant.

The dominant VOT pattern for /k/ and /q/ could be explained by the idea according to which the relation linking these two consonants is of the same nature as the one linking /t/ and /tˁ/—that is, a pharyngealized relation that materializes, inter alia, through a long VOT for /t k/ opposed to a short VOT for /tˁ q/. Lisker and Abramson (1964) indicated that the duration of the VOT varies according to the place of articulation of the consonant: longer for velar consonants; shorter for labial consonants; medial for dentals. However, this is not the case here, and, despite a more backed place of constriction than that of /k/, /q/ inherits a shorter VOT. The explanation is given in part in Cho and Ladefoged (1999: 213), which explains the VOT variation in terms of the volume of the cavity in front and behind the constriction, the movements of the articulators, and the contact zone between the articulators. Basing his data on cineradiographic data, Bonnot (1976: 440) gives details on the articulation of the two consonants and on the contact zones between articulators; these details explain the long release of /k/ and nearly simultaneous release of /q/.

Besides the variation according to the place of articulation of the consonant, Yeni-Komshian et al. (1977: 43) show that the duration of the VOT with Lebanese speakers varies according to the adjacent vowel: the VOT is longer with front vowels.

2.3.4 Consonant and Vowel Variation

Variation concerns all segmental units of Arabic—consonants as well as vowels. The best-known phenomenon for consonants is that of assimilation. Kaye (2009: 564) indicates several consonant assimilation cases. The assimilation in Arabic concerns all consonants and can be partial or total. The hamza, the glottal stop /ʔ/, is considered by some as a consonant that gets completely assimilated by the solar adjacent consonant when it is at a word initial position (Canepari 2005: 325). The Arabic linguistic tradition, on the other hand, considers this purely graphic hamza as a latent consonant.

Assimilation as a phonetic phenomenon was well studied by early grammarians, who precisely described the assimilation of /n/ in [ŋ], [nˤ], or [m] before /q, k, ʃ, j, s, z, sˤ, d, t, tˤ, dˤ θ, ð, ðˤ, f/. They underlined the dependency of the nasal expansion on the place of articulation, thus showing that the guttural (stops and pharyngeal fricatives /ħ/, /ʕ/, and glottals /ʔ/ /h/) blocked this assimilation (Bakalla 1983). Consonant assimilation was also explored in its phonological dimension (Abu Salim 1988).

As presented already, early Arab grammarians emphasized the allophonic variants of consonants (see Sībawayhi's description of secondary articulations, mustaḥsana and ġayr mustaḥsana). Embarki et al. (2011a) explain that the differences for locus equations of pharyngealized consonants between Modern Arabic and Dialectal Arabic and among the four countries used in the study (Jordan, Kuwait, Morocco, and Yemen) were due to a weakening of the pharyngealization gesture. Indeed, pharyngealized consonants tend to be articulated like their non-pharyngealized counterparts with very few retraction effects in the back of the tongue; this tendency is very clear in the realization of the consonant /sˤ/ (ibid., 204).

As explained in Section 2.2.2.2, the Arabic vowel system consists of three cardinal qualities that contrast in terms of length: /i u a/ versus /iː uː aː/. This configuration is absolute and does not consider the allophonic realization of phonemes, which is slightly richer. Early Arab grammarians such as Sībawayhi, described these variations, such as the precision of the imala phenomenon (cf. Sara 2007; [Sara, "Classical Lexicography"]). Kaye (2009: 565) explains that variation affects Arabic short vowels more than long ones. He lists a total of 16 different allophones for the six basic phonemes; Al-Ani (1970: 23–24) lists 17 allophonic realizations. Canepari (2005) illustrates on a diagram the principal allophonic realizations of six vowels in Modern Arabic (see Figure 2.4).

These allophonic realizations essentially depend on the phonetic context (the nature of the adjacent consonant) and the prosodic nature (stressed vs. unstressed syllable). The aforementioned studies, which are specific to pharyngealization, show without exception that in a pharyngealized context the frequency of F1 increases and the frequency of F2 decreases noticeably. The phonetic contiguity of certain consonants pushes the cardinal vowels toward less peripheral frequencies. Embarki et al. (2006) show that the formants of three short cardinal vowels of Modern Arabic presented