Introduction

Karin Christina Ryding and David Wilmsen

i.1 Overview

Arabic linguistics is a field that has both expanded and shifted over the last fifty years. The coming to the fore of Arabic sociolinguistics, variation theory, corpus linguistics, language acquisition, intercultural pragmatics, and Arabic media studies has enlarged the nature of research topics, strategies, and results so that both spoken and written forms of Arabic have come to be examined from multiple perspectives. Moreover, the development of social media and discussion platforms has had a profound effect on the interface of spoken and written language that has yielded new forms of Arabic discourse. This handbook brings together articles on a range of traditional and contemporary topics from a wide spectrum of research interests. We hope that the integration of new and traditional will represent both the broadened horizon for Arabic linguistic analysis and new congruence within this disciplinary area.

In discussing the field of Arabic studies, Jonathan Owens has referred to 'the immensity of the field itself' (2013: 9). Arabic linguistics distinguishes itself not only by being both broad and at the same time single-language focused, but also by being profoundly characterized by variation and language change built upon the architecture of a written tradition that has remained relatively stable for over a millennium. Instead of fragmentation, however, there is a new unity to Arabic linguistic research. That unity is complex and emergent, but nonetheless vital and groundbreaking in its depth and in its reach. Arabic is one language, old and new, tradition-bound yet transformative, linking past and future in myriad ways. As Arabic linguistics research develops and extends beyond the classical tradition, it transforms into an interdisciplinary field both linked and divided by the rapid spread and valorization of vernacular studies and the interplay between spoken and written discourse.

Whereas theoretical linguistics has been primarily concerned with language in itself, the field of Arabic linguistics provides a range of language forms and functions from formal to informal, classical to contemporary, from written to spoken, from individual to nation, that have vastly different research traditions and different sets of problems to identify and address. Most languages have similar issues, but for Arabic, the centrality of the fact of diglossia has long distanced traditional indigenous scholarship from vernacular studies and has resulted in politically, culturally, and socially fraught attitudes towards spoken discourse. This is now changing. A new synthesis is bringing these different strands of research together.

i.2 Arabic and Linguistics Research

Indeed, so vast is the study of Arabic that it would seem impossible for individual researchers to encompass it in its entirety. Perhaps this is a reason why Arabic, as an eminently accessible language, often nonetheless escapes the notice of the larger field of linguistics. That is not to say that no mention is ever made of Arabic in current linguistics literature; it occasionally is mentioned, but not in proportion to its size or importance, with upwards of 300 million native speakers, as the fifth largest language on the planet.[1] Even when mention of Arabic is made, the facts of Arabic are often presented incorrectly. Consider, for example, these recent works in the generativist or language typology traditions. Roberts and Roussou (2003: 31–3) in discussing *wh*-interrogative movement, a topic of paramount interest to generativists, state that those of Iraqi Arabic remain *in situ* with occasional *wh*-fronting:

(1) Iraqi Arabic

a. *mona* *šāf-at* *meno*
name see.IPFV-3FSG Q
'Whom did Mona see?'

b. *meno* *šāf-at* *mona*
Q see.IPFV-3FSG name
'Whom did Mona see?'

Quite to the contrary, the formation in (1b) is the normal unmarked sequence, with (1a) being marked or even ill-formed.

In a work on the typology of predicate possession, Stassen (2009) states of Cairene Arabic that 'the dative marking [*li*] on the possessor has been replaced by a preposition with the basic meaning "at"' (2009: 323), when this is clearly not the case:

[1] In the Roberts and Rousseau work discussed immediately below, Arabic is given eight references, two for what the authors call 'Classical' Arabic, and three references each for the spoken Arabic of Iraq and Morocco. By contrast, Greek, with between 15 and 20 million native speakers – equal to the population of Cairo – enjoys a full 155 references.

(2) Cairene Arabic[2]

a. *mā* *la-k-š* *daʕwā* *bi-ya*
NEG DAT-PRO.2MSG-NEG invitation PREP-PRO.1SG
'You have nothing to do with me'

b. *kull* *wāḥid* *wa* *l-uh* *rāʔy-uh*
every one CONJ DAT-PRO.3MSG view-PRO.2MSG
'Everyone has his own opinion'

Finally, the *World Atlas of Language Structures Online* (Dryer and Haspelmath 2013), drawing upon an obscure work on Palestinian Arabic (Shahin 1995: 30–1), types Palestinian Arabic as having no 2nd-person imperative, as opposed to northern Levantine and Moroccan Arabic, the only other two Arabic varieties – aside from Maltese – on the map charting the morphological imperative in world languages (70A). In point of fact, all varieties of Arabic form their imperatives from the 2nd-person imperfective verb in much the same manner all across the Arabophone world:

(3) Palestinian Arabic

a. *šū* *b-ti-nsā* *yā* *zalame*
EXCLAM HAB-2SG-forget.IPFV VOC man
'How you forget, man!'

b. *insā* *yā* *zalame*
forget.IMP VOC man
'Forget [it], man!'

Similarly, in the Atlas, Palestinian Arabic and Maltese are classed as possessing two different types of prohibitives, when Maltese generally and Palestinian often form their prohibitives in the same manner, with postpositive *-š* alone:

(4) a. Palestinian Arabic

ti-nsā-š *inn-u* *zalame* *mitl-ak*
2-forget.PROH-NEG COMP-PRO.2MSG man PREP-PRO.2MSG
'Forget not that he [is] a man like you'

b. Maltese

Tinsiex li Ġesù wkoll ibati mil-lonliness, bħalek u bħali
ti-nsī-š *li* *ġēsu* *wkoll* *i-bati* ...
2-forget.PROH-NEG REL Jesus ADV 3-suffer.IPFV
b-ħal-ek
PREP-condition-PRO.2MSG
'Forget not that Jesus also suffers ... like you'

[2] See, also, the examples in Woidich (2006: 335).

These sources are not entirely to be blamed for misrepresenting Arabic, especially inasmuch as they have referenced few and obscure works about the language. But for that very reason they can be questioned for their choice of references. Roberts and Roussou only list two, both from the 1980s, in their nineteen pages of references. Likewise, Stassen in an even larger reference section of thirty-two pages. To be sure, Stassen lists only one or two references each for most of the 420 languages in his sample. All the more reason for exercising care in the choice of representative works. For its part, the WALS must perforce list more for its mapping of the features of twenty Arabic varieties (twenty-one with written Arabic, but listing that variety achieves the perverse result of its being listed both everywhere on the map of the Arabophone world and nowhere). So, too, do most of its references (twenty-six of thirty-eight) come from the 1980s or earlier; the remainder are from the 1990s, with one from 2000. True, three of those are from the excellent series of indispensable reference grammars published by Georgetown University Press in the 1960s. Nevertheless, by skewing the sample to works from the mid twentieth century – many of them obscure, narrowly focused, and hard to obtain – the compilers of the WALS miss many up-to-date important standard reference works about Arabic dialects, a sterling example of which being Woidich's (2006) authoritative reference grammar of Cairene Arabic.

It would be unrealistic to expect such linguistic surveys as these three recent works to provide comprehensive listings of all pertinent references to the languages that they address, but it is not too much to ask that those works that they do reference be accessible general references. On the contrary, it is the place of a handbook to guide its readers to the wide range of research and reference works available in any of the subfields that it treats, and the contributors to this volume shoulder that responsibility admirably, such that any interested comparative linguist consulting the subsections of this handbook should, by following the references therein, be able to find accurate representation of the facts and any matters of uncertainty or dispute that Arabic can contribute to their reviews of phenomena in which they may be interested.

We have asked contributors to address whenever possible what Arabic specifically has to offer to linguistic theory. This cannot always be done overtly, some matters being of primary concern to the internal discussions in the field of Arabic linguistics proper. Nevertheless, contributions to each of the subsections should guide general linguists toward the involvement of Arabic in matters of crucial concern to linguistics as a whole.

i.3 In This Volume

This handbook offers a wide range of articles on contemporary Arabic linguistics, from theoretical studies of Arabic morphosyntax to studies of

Arabic use in online communities. It connects traditional research areas with new questions, new problems, and new approaches to language use, language learning, theoretical analysis, computational linguistics, and vernacular studies. We have gathered articles from specialists in a number of different areas, reflecting new perspectives on Arabic linguistics as well as updates on traditional areas of investigation. In terms of content, the handbook is divided into six thematic areas: Arabic applied linguistics, Arabic variation and sociolinguistics, theoretical and descriptive studies, Arabic computational and corpus linguistics, Arabic linguistics and new media studies, and Arabic linguistics in literature and translation. Here we provide an outline of chapter topics, along with a more detailed summary of each chapter.

i.4 Outline of Topics

In the first part, which addresses matters of applied linguistics, Mohammad Alhawary begins with a broad overview of the field as it applies to Arabic and as matters peculiar to Arabic intersect with the broader field of linguistics. This is followed by Hussein Elkhafaifi's minute examination of matters of particular concern to official language planners in the Arabophone world, having essentially to do with the matter of Arabic diglossia and widespread bilingualism in Arab countries. Karen Froud and Reem Khamis-Dakwar provide a broad overview of language acquisition studies in Arabic, a subfield of Arabic linguistics in which they are central figures, language acquisition itself being a matter of central concern in linguistics. Michael Raish provides a state-of-the-art study of Arabic testing and assessment, an area of rapidly growing concern as learners need professional assessment and documentation of their oral as well as written skills. He examines both end-of-training and formative assessments in task-based and performance-based formats, including computer-assisted language testing, study-abroad testing, the C-test, and self-assessment. This is followed by Emma Trentman's practical guide to issues involved in teaching Arabic as a foreign language (TAFL) in study-abroad settings, language study abroad attracting much interest in the field of applied linguistics. Kassem Wahba's chapter completes the applied linguistics section, addressing the implications of diglossia in Arabic and the teaching of Arabic as a foreign language, that being how to teach students the phonetics of a language with a multitude of dialects representing wide phonological variation between them.

The sociolinguistics part begins with two contributions that pertain to wider concerns in the field before narrowing the frame to matters internal to Arabic. Samira Farwaneh discusses the issue of language complexity, a theme central to Arabic, inasmuch as the Arabic of writing is widely and inaccurately perceived to be more complex than the dialects of Arabic. At

the same time, the question of the relative equality of complexity in languages is a matter often taken for granted in linguistics, only lately beginning to be challenged. As far as challenging widely held precepts is concerned, Uri Horesh provides an overview of sociolinguistic approaches to Arabic, some of those, of course, being of purely parochial interest to the practice of Arabic sociolinguistics, but others conforming to and at times challenging the basic tenets of sociolinguistics as a whole. From there, things begin to focus on matters internal to Arabic. Gunvor Mejdell returns again to matters of Arabic diglossia, addressing its manifestations in speech production in the Arabophone world, a concern that she has spent her entire career examining. Stephan Procházka then gives a thorough-going overview of the study of Arabic dialects across the width and breadth of the Arabophone world. The sociolinguistics section ends with David Wilmsen's examination of a peripheral dialect of Arabic, Maltese, showing how that language variety is instrumental in charting the unwritten history of the modern dialects of Arabic, at the same time showing how it conforms nicely to the findings of historical linguistic studies of isolated language varieties.

The part on theoretical and descriptive studies fulfils the promise of the handbook more easily than others because its contributors are as a group more involved in matters current in the study of human language. Mohssen Esseesy places Arabic firmly at the centre of grammaticalization theory, a matter to which he has been devoting himself for the better part of two decades. Terrence Potter looks at the onomastics of Arabic, a field not often addressed in Arabic linguistics itself, but of wide implication for historical linguistics. Similarly, Khaled Rifaat addresses matters of intonation in Arabic, something that has been stoutly ignored since the earliest medieval descriptions of Arabic and still gains little purchase in discussions of Arabic linguistics, but which is, nonetheless, of great interest to linguistic studies of prosody in language. The final three chapters bring Arabic into morphosyntax, with co-editor Karin Ryding's contribution on case in Arabic, Usama Soltan's demonstration of the applicability of Arabic morphosyntax to the enterprise of generative grammar, and Janet Watson's discussion of the inflectional and derivational morphology of Arabic. All three of these chapters come from recognized authorities in their respective approaches to morphosyntax.

With that, the handbook turns to matters both applied and theoretical in addressing the computational linguistic and corpus linguistic dimensions of Arabic. First Nizar Habash, a central figure in Arabic computational linguistics, provides an overview of the field. His overview is followed by Mark Van Mol, another figure central to the field, whose contribution dissects the supreme complexity and difficulty of constructing Arabic corpora. The part ends with a contribution from Mai Zaki, Dana Abdulrahim, and the other co-editor of the *Handbook*, three of the few linguists to use Arabic corpora in answering questions pertinent to both

theoretical and applied concerns in the field of linguistics in general and Arabic linguistics in particular.

Remaining in the field of electronic language resources, a part on new media studies presents contributions from two researchers who utilize the media. Amy Johnson addresses localization and internationalization, specifically in the social media platform Twitter and its approach to and effect on language use in the Arab world. Francesco Sinatora examines social activism on Facebook. Neither contribution can escape the parochiality of the wider themes that have concerned Arabic linguistics for half a century, but in revisiting them they interrogate their basic tenets.

The final part, too, is more concerned with applied matters internal to Arabic than with issues of interest to a broader linguistics readership. Eva Håland looks at the tension between the Arabic of formal writing and the informal (read 'spoken') features that occasionally appear in it – a matter of central concern to the study of diglossia in Arabic from its beginning – and that are encroaching with increasing regularity on formal writing. Mai Zaki applies parallel corpora to the translation of deixis in Arabic novels into English, finding, unsurprisingly, that the deictic categories of the two languages do not overlap seamlessly.

i.5 Final Remarks

It should not be surprising that the intensive study of any language, and thus a handbook of the linguistics of that study, often turns inward to address matters of parochial concern. There are probably as many definitions of linguistics and the core concerns of linguistics as there are practising linguists. Nevertheless, it is possible to outline two broad approaches to linguistic analysis: either linguistics is interested in explaining the human capacity for language, in which case it attempts broad generalizations, often missing the details of daily language use; or it is interested in those details, only subsequently attempting to fit its observations into a theory of language. Another way of stating this is that linguistics is either interested in the inherent mechanisms common to all human languages and how those shape language in general, or it is interested in how language users themselves shape their languages in particular. Arabic linguistics in its diverse manifestations acknowledges both, but at the same time, it is captivated by the entrancing details of the language itself. Often this has resulted in the field of Arabic linguistics speaking largely to itself. For that reason, perhaps Arabic linguistics as a field is itself to blame for its low profile in the larger field of linguistics and for the misrepresentations that linguistics makes of Arabic. Nevertheless, the very complexity of the subject matter perhaps contributes both to linguists' superficial dabbling in the sea that is Arabic, making barely a ripple in the surface of the waters, and the bemused indulgence of those dilettantes by the

mariners engaged in the navigation of those waters. One of those seasoned mariners, Clive Holes, has summed up the matter best:

> Arabic, and in particular the Arabic dialects, remain 'on the outside looking in' when it comes to their profile in general linguistics journals. In part, the blame for this low visibility can be laid at the door of Arabic linguists themselves for their failure to use Arabic data to test and critique current theoretical models. Many of them would counter, however, that current linguistic model-building is simply too theory-driven, too abstract, and too distant from their interests to justify the expenditure of time.
>
> (Holes 2018: 28)

The contributions of this handbook do show how Arabic can contribute to linguistic theorizing and model-building; but they also show where those involved in the study of Arabic are preoccupied with forming the groundwork for describing the language in its vast but intricate whole.

References

Dryer, M. S. and Haspelmath, M. (eds.) (2013). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info; last accessed 8 November 2020.)

Holes, C. (2018). Introduction. In C. Holes, ed., *Arabic Historical Dialectology: Linguistic and Sociolinguistic Approaches*. Oxford: Oxford University Press, 1–28.

Owens, J. (2013). A house of sound structure, of marvelous form and proportion: An introduction. In J. Owens, ed., *The Oxford Handbook of Arabic Linguistics*. Oxford: Oxford University Press, 1–22.

Roberts, I. and Roussou, A. (2003). *Syntactic Change: A Minimalist Approach to Grammaticalization*. Cambridge: Cambridge University Press.

Shahin, K. N. (1995). *Rural Palestinian Arabic* (*Abu Shusha dialect*). Munich: Lincom Europa.

Stassen, L. (2009). *Predicative Possession*. Oxford: Oxford University Press.

Woidich, M. (2006). *Das Kairenisch-Arabische: Eine Grammatik*. Wiesbaden: Harrassowitz.

Part I

Arabic Applied Linguistics

1

Arabic Applied Linguistics

Mohammad T. Alhawary

1.1 Introduction

As the term 'applied' implies, broadly speaking the definition 'applied linguistics' encompasses every field where language is *applied*. The study or application of language with respect to its use and users is far and wide and covers fields such as language learning, language testing and assessment, language processing and cognition, discourse and conversation analysis, language and law, language policy and planning, bilingualism and multilingualism, language contact, language variation and change, dialectology, language disorders and pathology, translation and interpretation, and natural language processing, to name but a few. In addition, many areas in applied linguistics draw on theoretical approaches, some of which are language related (i.e., across the different language domains including syntax, semantics, phonology, morphology, and pragmatics) and some are from other disciplines, such as psychology, anthropology, sociology, communication, and cognitive neuroscience. Indeed, the fields which fall within the confines of applied linguistics and the topics which are treated have become so many that at least two encyclopedias have so far been published in English (Chapelle 2013; Berns 2010). The present chapter focuses on applied linguistics as it relates to second language acquisition (SLA), second language pedagogy, and second language proficiency testing. The first two sub-disciplines were central to the preoccupation of applied linguists since the field's early inception in the twentieth century (especially during the earliest formative period of second language acquisition in the 1940s–60s) and continued for the greater part of the twentieth century, for English as a second/foreign language in particular. For Arabic, second language acquisition and second language pedagogy, as well as testing, have remained to the present the central topics of Arabic applied linguistics in the USA and in the Arabic-speaking world. Other areas in theoretical linguistics and sociolinguistics are also discussed,

since they have a direct bearing on Arabic second language acquisition and second language pedagogy with respect to both their development and practice. Other related areas and sub-disciplines of Arabic applied linguistics are covered in other chapters. The remainder of the chapter is organized as follows: Section 1.2 provides an overview of the different approaches investigated within Arabic SLA, Section 1.3 deals with Arabic proficiency testing and assessment, in particular the status of testing–SLA interface; and Section 1.4 discusses Arabic second language pedagogy with respect to the ongoing debate of addressing diglossia and teaching Arabic variation. The chapter concludes with challenges and future directions.

1.2 Arabic SLA: Past and Present Approaches

Arabic SLA kept abreast, progressively more so, with the development of the general SLA field and its different approaches and models, from contrastive and error analysis to the present (for a detailed survey, see Alhawary 2009a). Thus, Arabic SLA studies started (a decade after the emergence of error analysis) with the appearance of a few error analysis studies (e.g., Al-Ani 1972–1973; Rammuny 1976), where the approach of SLA analysis relied on analysing L2 learners' competence-related errors. Despite the insights such studies (based on cross-sectional written composition data) lent about types of errors made by English-speaking learners of Arabic (e.g., orthographic/phonological, lexical, structural, and stylistic), one of the main limitations of error analysis studies in general is that they did not provide the bigger picture of what and how much the learners got right versus what and how much they got wrong.

Error analysis limitations led to the abandonment of the approach and the adoption of no fewer than seven approaches since then. One such approach sought to analyse the overall performance of the learner (i.e., the big picture) or to identify emerging patterns or developmental sequences in acquiring a given form or a set of forms (see Larsen-Freeman and Long 1991). Within this approach, Al-Buainain (1986) examined the acquisition of negation and interrogation by British English-speaking learners of Arabic (based on cross-sectional written translation and manipulation data). Al-Buainain's findings indicated that, in acquiring the target forms, participants progressed along intermediate stages from simple and deviant interlanguage forms to ones that are more acceptable and target-like. Relatedly, Alhawary (2009a) examined negation constructions and relevant mood endings (based on longitudinal production data). The findings showed the emergence of a certain order of the constructions and such order was impacted by input effects, such as amount and exposure timing.

A second post-error analysis approach considered other phenomena not previously examined, such as communication strategies (e.g., generalizations,

avoidance, circumlocution, message adjustment, prefabricated patterns, and codeswitching). Although initially thought to be peculiar to L2 learners coping with their limited knowledge of the L2, communication strategies were later problematized as not being dissimilar to adjustments made by L1 speakers to accommodate real-time processing and communication constraints (e.g., Tarone 1980; Bialystok 1990). At least one study, Fakhri (1984), examined communication strategies in Arabic L2 acquisition (based on production data from an L1 English speaker). In addition to observing similar communication strategies used by the participant, Fakhri found the strategies were constrained by the specific narrative genre components (i.e., the orientation, episodic, and evaluation parts) in that certain strategies were used more often during one part than other parts.

A third post-error analysis approach is the application of SLA to two major cognitive processing models. The first is that of speech processing constraints adopted by Processability Theory (PT). The theory seems attractive, as it provides a detailed framework and a strong committed stance towards foreign language pedagogy through its learnability and teachability claims. In particular, it provides a cognitive account for L2 grammatical development and posits that L2 learners create L2 language-specific processing resources or prerequisites along an implicational hierarchy of stages which cannot be skipped; i.e., structures which are from a higher stage cannot be learned before those from a lower one. This learnability prediction lends itself to another prediction to do with teachability: that instruction can be beneficial if it focuses on structures from the next stage (e.g., Pienemann 1998, 2005; Pienemann and Keßler 2011; Baten et al. 2015). However, based on longitudinal and cross-sectional production data, Arabic SLA studies produced mixed evidence on a number of morphosyntactic structures. Such studies either found evidence contrary to the hypothesized implicational processing hierarchy (e.g., Nielsen 1997; Alhawary 2003, 2009a, 2009b, 2019) or claimed evidence in support of it (e.g., Mansouri 2000; Al Shatter 2008) (for a review of such studies, see Alhawary 2009a, 2019). The second cognitive model which was tested by Arabic L2 studies is the Competition Model developed by Bates and MacWhinney (1987). Arabic L2 studies on sentence processing within the Competition Model produced mixed evidence. Abu Radwan (2002) found that English-speaking learners of Arabic initially relied on Arabic L2 cues of case markings, followed by gender and animacy and did not exhibit L1 transfer, though the study did not investigate word order.[1] By contrast, Al-Thawahrih (2018) found English L1 learners of Arabic initially exhibited L1 transfer by relying on L1 word-order cues and started utilizing L2 verbal agreement cues in processing Arabic L2 sentences and assigning agency with more exposure to the language.

[1] For a detailed review of the study and input factors that may explain the findings, see Alhawary 2009a: 39–41.