



languages

The Classification of Arabic Dialects

Traditional Approaches, New Proposals,
and Methodological Problems

Edited by

Simone Bettega and Roberta Morano

Printed Edition of the Special Issue Published in *Languages*

**The Classification of Arabic Dialects:
Traditional Approaches, New
Proposals, and Methodological
Problems**

The Classification of Arabic Dialects: Traditional Approaches, New Proposals, and Methodological Problems

Editors

Simone Bettega

Roberta Morano

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade • Manchester • Tokyo • Cluj • Tianjin



Editors

Simone Bettega
University of Turin
Italy

Roberta Morano
University of Leeds
UK

Editorial Office

MDPI
St. Alban-Anlage 66
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Languages* (ISSN 2226-471X) (available at: https://www.mdpi.com/journal/languages/special_issues/arabic_dialects).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. <i>Journal Name</i> Year , <i>Volume Number</i> , Page Range.
--

ISBN 978-3-0365-6139-4 (Hbk)

ISBN 978-3-0365-6140-0 (PDF)

© 2022 by the authors. Articles in this book are Open Access and distributed under the Creative Commons Attribution (CC BY) license, which allows users to download, copy and build upon published articles, as long as the author and publisher are properly credited, which ensures maximum dissemination and a wider impact of our publications.

The book as a whole is distributed by MDPI under the terms and conditions of the Creative Commons license CC BY-NC-ND.

Contents

About the Editors	vii
Simone Bettega and Roberta Morano	
Preface	
Reprinted from: <i>Languages</i> 2022, 7, 58, doi:10.3390/languages7010058	1
Antonella Torzullo	
The Emergence of a Mixed Type Dialect: The Example of the Dialect of the Bani ʿAbbād Tribe (Jordan)	
Reprinted from: <i>Languages</i> 2022, 7, 9, doi:10.3390/languages7010009	5
Bruno Herin, Igor Younes, Enam Al-Wer and Youssef Al-Sirour	
The Classification of Bedouin Arabic: Insights from Northern Jordan	
Reprinted from: <i>Languages</i> 2022, 7, 1, doi:10.3390/languages7010001	25
Bettina Leitner	
New Perspectives on the Urban–Rural Dichotomy and Dialect Contact in the Arabic <i>galāt</i> Dialects in Iraq and South-West Iran	
Reprinted from: <i>Languages</i> 2021, 6, 198, doi:10.3390/languages6040198	35
Felipe Benjamin Francisco	
The Southern Moroccan Dialects and the Hilāli Category	
Reprinted from: <i>Languages</i> 2021, 6, 192, doi:10.3390/languages6040192	57
Navdeep Sokhey	
Auditory and Acoustic Evidence for Palatalization of the Nasal Consonant in Cairene Arabic	
Reprinted from: <i>Languages</i> 2021, 6, 190, doi:10.3390/languages6040190	67
Ana Iriarte Díez	
A Typological Analysis of Cognate Infinitives in Lebanese Arabic Based on Comparative Semitic Evidence	
Reprinted from: <i>Languages</i> 2021, 6, 183, doi:10.3390/languages6040183	83
Adam Benkato and Christophe Pereira	
An Innovative Copula in Maghrebi Arabic and Its Dialectological Repercussions: The Case of Copular <i>yabda</i>	
Reprinted from: <i>Languages</i> 2021, 6, 178, doi:10.3390/languages6040178	103
Ahmad Al-Jallad	
Connecting the Lines between Old (Epigraphic) Arabic and the Modern Vernaculars	
Reprinted from: <i>Languages</i> 2021, 6, 173, doi:10.3390/languages6040173	117
Alexander Magidow	
The Old and the New: Considerations in Arabic Historical Dialectology	
Reprinted from: <i>Languages</i> 2021, 6, 163, doi:10.3390/languages6040163	129
Phillip W. Stokes	
A Historical Reconstruction of Some Pronominal Suffixes in Modern Dialectal Arabic	
Reprinted from: <i>Languages</i> 2021, 6, 147, doi:10.3390/languages6030147	157
Stefano Manfredi and Caroline Roset	
Towards a Dialect History of the Baggara Belt	
Reprinted from: <i>Languages</i> 2021, 6, 146, doi:10.3390/languages6030146	175

Cristina La Rosa

Mahdia Dialect: An Urban Vernacular in the Tunisian Sahel Context

Reprinted from: *Languages* 2021, 6, 145, doi:10.3390/languages6030145 193

Islam Youssef

Contrastive Feature Typologies of Arabic Consonant Reflexes

Reprinted from: *Languages* 2021, 6, 141, doi:10.3390/languages6030141 217

Mike Turner

Definiteness Systems and Dialect Classification

Reprinted from: *Languages* 2021, 6, 128, doi:10.3390/languages6030128 235

Thomas A. Leddy-Cecere

Interrogating the Egypto-Sudanic Arabic Connection

Reprinted from: *Languages* 2021, 6, 123, doi:10.3390/languages6030123 257

About the Editors

Roberta Morano

Roberta Morano completed a Ph.D. in linguistics at the University of Leeds, UK, in 2019, specializing in the Arabic varieties spoken in northern Oman. Her research interests focus on language documentation and ecolinguistics, especially the relationship between the environment and its representation in language. She has published several articles on the morphosyntax and sociolinguistics of Omani dialects. Her most recent publication is “Diachronic variation in the Omani Arabic vernacular of the al-‘Awabi district: from Carl Reinhardt (1894) to the present day” (2022).

Simone Bettega

Simone Bettega obtained his Ph.D. in linguistics at the University of Turin, Italy, in 2016. His research focuses on the documentation and typological classification of spoken varieties of Arabic, particularly those of southeastern Arabia. He has published articles and monographs on several aspects of the syntax of spoken Arabic. His most recent publication is the volume “Gender and Number Agreement in Arabic”, which was co-authored by Luca D’Anna.

Editorial

Preface

Simone Bettega ^{1,*} and Roberta Morano ²

¹ Department of Foreign Languages and Literatures and Modern Cultures, University of Turin, 10124 Torino, TO, Italy

² Department of Linguistics and Phonetics, School of Languages, Cultures and Societies, University of Leeds, Leeds LS2 9JT, UK; r.morano@leeds.ac.uk

* Correspondence: simone.bettega@unito.it

To seek for knowledge is to strive for systematization. All scientific disciplines, from physics to biology, from philosophy to medicine, have always been haunted by the question of classification. That categorization constitutes such a fundamental prerequisite to the progress of science is made all the more problematic by the fact that reality is, ultimately, a slippery thing. The world that we explore through our senses is a unified whole, continuous and undivided; it comes to us as an uninterrupted stream of perception and experience, and in the midst of this flow, it is hard to decide where something ends, and something new begins.

No one, arguably, knows this better than a linguist. Languages are, by definition, classificatory systems. They are all shared by a community of speakers who have all agreed on the fact that the world is to be segmented in a certain way. Leaves rest on branches; branches stem from trees; trees congregate to form woods, or sometimes forests. The act of speaking allows us to put order into fractal chaos. Unfortunately, not all communities of speakers, and therefore not all languages, interpret reality in the same way. Some have looser boundaries, others stricter ones. The idea of a forearm is a different thing to different people (different speakers, that is); and where does an elbow belong, exactly? No one better than a linguist understands the frustration that comes with the impossibility of classification, and this is because to classify languages is to classify classifications; even worse, it is to classify self-classifying classifications (and the dangers implicit in recursive categorizations have been well-known at least since the formulation of Russell's famous paradox).

Arabic dialectology is but a minor sub-branch of the general field of linguistics. Yet, its scope is undeniably vast. Varieties of Arabic are legion: they are spoken by hundreds of millions of people, scattered over a territory remarkably larger than geographical Europe, spanning from the Atlantic Ocean to the Persian Gulf, from southern Turkey to the southernmost tip of the Arabian Peninsula, and even making inroads into central Asia, sub-Saharan Africa, and a number of Mediterranean islands. Arabic dialects represent the modern descendants of one of the language families with the longest history of written attestation in the world, and display an amazing variety of forms and structures at the typological level. To put order into so vast a matter is obviously no easy task, and it should come as no surprise that the different classifications currently available to scholars of Arabic dialectology are all somewhat unsatisfactory, and have been subject to heavy critiques over the course of the years. As Owens (2013) elegantly puts it, "If till today simple models for classifying Arabic dialects elude us [. . .], 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".

We are not saying, of course, that classificatory systems for Arabic dialects do not exist: they do, and have been employed to some effect. Critiques to these systems, however, also abound. To make but a couple of examples, one can think of what is probably the most widely employed categorization in the field of Arabic dialectology, the one that distinguishes "Bedouin" varieties from "sedentary" ones. This bipartite subdivision has

Citation: Bettega, Simone, and Roberta Morano. 2022. Preface. *Languages* 7: 58. <https://doi.org/10.3390/languages7010058>

Received: 24 February 2022

Accepted: 28 February 2022

Published: 3 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

been put into question for conflating the synchronic and diachronic dimensions, together with sociolinguistic considerations, not always in a methodologically sound manner (for a critique, see, among many Palva 2006, p. 605; Watson 2011, p. 869; Vicente 2019, p. 109; and also an interesting note in Magidow 2016, p. 93). Even when we focus on systems of categorization that are narrower in scope, because they are concerned with specific areas within the Arab World, or specific subsets of phenomena, we still encounter problems. It is the case, for instance, of the traditional labels employed to classify Maghrebi varieties, which have recently been put into serious question by Taine-Cheikh (2017), Mion (2018), and Benkato (2019).

From all of the above, it should be clear that we are still far from finding unified and satisfactory solutions to many practical and terminological problems that have long been haunting the field of Arabic dialectology. We hope that the present volume can represent yet another, if minor, contribution to the vast collective effort of trying to better assess, organize and understand varieties of spoken Arabic. Obviously, the numerous papers that appear in the following pages differ greatly from one another in terms of scope and focus. This is no doubt because, as we have said, the subject of inquiry is vast, and exists simultaneously at both a local and supra-local, general scale. Both dimensions, we believe, are important, and both are represented among the studies presented here. Clear examples of the former are the article by Herin, Younes, Al-Wer, and Al-Srūr and the one by Torzullo. The two papers tackle similar issues, and put into question some of the categories which have historically been used to classify the dialects of Northern Arabia and the Southern Levant, with a focus on Jordan. In particular, they do so by also paying attention to the recent social developments of the area, which have involved great amounts of linguistic contact and consequent dialect levelling and mixing. Such attention to the socio-linguistic landscape of the area under investigation is also present in Leitner's discussion of the labels "rural", "urban", and "gōlōt" in contemporary Iraqi and Irani Arabic; here, these terms are re-examined in the light of recent phenomena of urbanization and population movement.

The dialectological situation of northern Africa, as noted above, is particularly complex, and several of the contributions that appear in this volume focus on this specific area. We start East, from the Egypto-Sudanic region, whose dialects are the object of Leddy-Cecere's inquiry. By also applying the methods of historical glottometry, Leedy-Cecere calls into question the validity of the traditional classification that claims a relationship to exist between the dialects of modern Egypt and Sudan. Sudanic Arabic is also treated in Manfredi's and Roset's study, whose focus, however, is broader, as it encompasses the whole "Baggara Belt", a strip of land more than 2500 km long that stretches from Sudan to Nigeria and is mostly inhabited by semi-nomadic and Arabic-speaking cattle herders. While examining the internal dialectal composition of Baggara Arabic, the two authors also provide new data for the refinement of the isoglosses commonly adopted for the identification of a West Sudanic dialect subtype. Finally, Sokhey's treatment of the palatalization of /n/ in Cairene Arabic suggests this specific trait to be sociolinguistically salient, and indexical of socioeconomic status, thus warranting further inquiries in the sociolinguistic situation of the Egyptian's capital.

If we move to the Maghreb proper, three of the articles presented here deal specifically with this region. Benkato and Pereira argue for broader inclusion of syntactic isoglosses in the classificatory systems of Arabic dialects, and offer a contribution in this sense by examining the emergence of a verbal copula in some dialects of Tunisia and northwestern Libya, a feature that appears to cut across the established isogloss lines of the area. Francisco proposes a re-examination of the categorization of southern Moroccan dialects in light of new data that have recently become available, questioning the validity of the labels "Bedouin", "Hilāli", and "Maṣqili" when referred to these varieties. Finally, La Rosa offers a preliminary description of the dialect of the Mahdia area in Tunisia, in which Bedouin, rural and urban features seem to be conflated, an observation that could help to better assess the linguistic nature of the Tunisian Sahel.

Issues of classification, of course, do not only arise in relation to the diatopic distribution of linguistic features; diachrony plays an important role as well. In Iriarte Díez's article we find both dimensions being addressed at once: starting from a survey on the role of cognate infinitives in Lebanese Arabic, the author broadens the scope of her analysis by bringing data from the Semitic language family at large to bear. This comparison reveals the Lebanese data to be in line with what is known about other Semitic languages, with the possible (and curious) exception of Classical Arabic, whose descriptions have often adopted a dismissive attitude towards the topic of cognate infinitives. The question of diachrony, and, therefore, of origins and evolution, is more directly addressed by Al-Jallad, who isolates a number of features that appear to characterize both the modern dialects and the ancient pre-Islamic epigraphic inscriptions, to the exclusion of Classical Arabic. Stokes is also concerned with the historical developments of Arabic, when he argues that the vowels which appear before the pronominal suffixes in several modern dialects are actually derived from original case vowels, and subdivides dialects into two main groups depending on how these vowels developed. Magidow's paper, finally, takes issue with the possibility of reconstructing the linguistic history of the Arabic languages by directly relating it to attested population movements and settlement patterns, and proposes the application of a new heuristic approach, based on sociolinguistics and geography, to re-examine the extant categories of Arabic dialectology.

Yet, another approach to the problem of classification is that adopted by both Turner and Youssef, who employ the tools of linguistic typology to try and bring order into the variegated reality of spoken Arabic. Turner uses definiteness as a case study, which he investigates applying the Reference Hierarchy framework, thus showing the importance of using semantic typology as a metric for grouping dialects, rather than relying on the presence of forms alone. Youssef, on the other hand, laments how attempts at classifying Arabic varieties based on consonantal realizations have historically employed a mixture of both linguistic and non-linguistic parameters: as an alternative, he proposes to investigate the phonological nature of the dialects through the use of segmental typology. As with Turner's paper, Youssef also underlies how typological inquiry allows for different possible categorizations of the dialects, which can support, refine, or disprove already existing classificatory models, but also suggest new viable groupings, and provide insights into diachronic processes.

In conclusion, there is not doubt that the systems that have been used to classify Arabic dialect up until this moment are not entirely satisfactory, and can be improved. What remains to be understood is what, of these systems, is there to be saved, what can be safely discarded, and what hitherto unexplored methodologies can be fruitfully applied to the field of inquiry. It seems to us that the articles that make up this volume are all relevant in this sense, and that they contain promising and interesting ideas worth exploring and expanding upon. We can only hope that the readers will share our views, and that in the following pages they will be able to find answers, new questions, and the inspiration to push the boundaries of their research even further.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Benkato, Adam. 2019. From Medieval Tribes to Modern Dialects: On the Afterlives of Colonial Knowledge in Arabic Dialectology. *Philological Encounters* 4: 2–25. [\[CrossRef\]](#)
- Magidow, Alexander. 2016. Diachronic Dialect Classification with Demonstratives. *Al-'Arabiyya* 49: 91–115.
- Mion, Giuliano. 2018. Pré-hilalien, hilalien, zones de transition. Relire quelques classiques aujourd'hui. In *Mediterranean Contaminations. Middle East, North Africa, and Europe in Contact*. Edited by Giuliano Mion. Berlin: Klaus Schwarz, pp. 102–25.
- Owens, Jonathan. 2013. A House of Sound Structure, of Marvelous form and Proportion. In *The Oxford Handbook of Arabic Linguistics*. Edited by Jonathan Owens. Oxford: Oxford University Press.
- Palva, Heikki. 2006. Dialects: Classification. In *Encyclopedia of Arabic Language and Linguistics*. Edited by Kees Veerstegh. Leiden and Boston: Brill, vol. 1, pp. 604–13.

- Taine-Cheikh, Catherine. 2017. La classification des parlers bédouins du Maghreb: Revisiter le classement traditionnel. In *Tunisian and Libyan Arabic Dialects, Common Trends, Recent Developments, Diachronic Aspects*. Edited by Veronika Ritt-Benmimoun. Zaragoza: IEIOP, pp. 15–42.
- Vicente, Ángeles. 2019. Dialect contact and urban dialects. In *The Routledge Handbook of Arabic Sociolinguistics*. Edited by Enam Al-Wer and Uri Horesh. London and New York: Routledge, pp. 106–16.
- Watson, Janet C. E. 2011. Dialects of the Arabian Peninsula. In *The Semitic Languages an International Handbook*. Edited by Stefan Weninger. Berlin and Boston: De Gruyter Mouton, pp. 897–908.

Article

The Emergence of a Mixed Type Dialect: The Example of the Dialect of the Bani ^ʿAbbād Tribe (Jordan)

Antonella Torzullo

Department of Near Eastern Studies, University of Vienna, 1090 Vienna, Austria; antonella.torzullo@univie.ac.at

Abstract: The present article aims at questioning the status of the *šāwi* dialect of the Bani ^ʿAbbād tribe by providing a new analysis of the main distinctive phonological, morphological, and syntactical traits which may hint at dialect mixing. The data provided by the field research, based on a functional framework that relies on descriptive linguistics and a typological approach, show that this dialect is deeply affected by a koineizing tendency due to increasing contacts with the populations of the neighboring areas (especially ^ʿAmmān and Salt) which, in turn, leads to the gradual loss of its authentic features. Finally, this paper discusses whether the dialect of the Bani ^ʿAbbād should still be considered as belonging to the *yigūl* group (recently renamed Central Bedouin *yǧūlu*) of the Syro-Mesopotamian sheep-raising tribes or if a new typology of mixed type dialects should eventually be adopted for the dialects displaying important markers of both Bedouin and sedentary types.

Keywords: spoken Arabic varieties; dialect classification; Jordanian Arabic; Arabic dialectology; Arabic linguistics

Citation: Torzullo, Antonella. 2022. The Emergence of a Mixed Type Dialect: The Example of the Dialect of the Bani ^ʿAbbād Tribe (Jordan). *Languages* 7: 9. <https://doi.org/10.3390/languages7010009>

Academic Editors: Simone Bettega and Roberta Morano

Received: 19 October 2021

Accepted: 18 December 2021

Published: 5 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Traditional Classification of Jordanian Dialects in the Light of Recent Developments

The Hashemite Kingdom of Jordan is characterized by a considerable diversity of regional dialects.

Nevertheless, in some cases the typological classifications of the Arabic dialects spoken in this country prove to be problematic, because there seem to be a considerable number of transitional and mixed type dialects.

As stated by Sawaie (2011, p. 499): “records of the linguistic situation in Trans-Jordan in the early part of the 20th century are not available. Consequently, it is hard to state with certainty which dialects dominated then”.

The only available source on the Arabic dialects spoken in this area in this period is Bergsträsser's (1915) *Sprachatlas von Syrien und Palästina*.

In 1936 Cantineau drew a classification of several nomadic dialects in his *Études sur quelques parlers de nomades arabes d'Orient*. In the first part of his study, he distinguished the dialects of the camel-rearing tribes from the small-cattle ones and divided the latter into two groups: the atrochaic and the trochaic dialects (Cantineau 1936, p. 114). However, after extending the scope of his research to some other tribes and completing some data he had previously collected, in the second part of his study Cantineau (1937, p. 110) classified the dialects into four groups (which he labelled as a *classement rationnel*): group A, B, C and Bc.

Further investigations were reported in 1963 when for the first time R. Cleveland typologically classified the Jordanian dialects through phonological, morphological, and syntactic characteristics and drew “the most general outlines of the situation illustrated by a very limited number of dialectal characteristics” (Cleveland 1963, p. 56).

In his work, he divided the Arabic dialects spoken in the area into four groups:

(1) *Yigūl*–(2) *Bəǧūl*–(3) *Bəkūl*–(4) *Bəʿūl*.

The groups were named according to the pronunciation of “*he says*” since this feature “indicates both an important phonetic and morphological characteristic” (Cleveland 1963, p. 171).

From 1969, Palva dedicated a considerable number of publications to the linguistic situation of Jordan and in 1984 he set a more complete classification of the Jordanian dialects in his “A General Classification for the Arabic Dialects Spoken in Palestine and Transjordan”, where he added some new criteria of analysis with regard to those applied by Cleveland, namely the reflex of the sequences *CvCaCv-* and *-aXC* (where *X* is a guttural), the gender distinction in the 2nd and 3rd persons plural in personal pronouns and verbs, the use of the adverbs *here* and *now*, and the occurrence of the compound negation *mā . . . š*.

The new classification he outlined divides the Jordanian dialects into three main groups:

1. The urban Palestinian dialects (*bi²ūl*), which typologically belong to the Levantine urban dialects;
2. The rural dialects, which are in turn grouped into:
 - a. the Galilean dialects (*biqūl*);
 - b. the central Palestinian dialects (*bikūl*), more conservative than the Galilean dialects;
 - c. the south Palestinian dialects (*bigūl*), closely related to the previous ones, they have some features that show a greater Bedouin (especially Negev) influence;
 - d. the north and central Transjordanian dialects (*bigūl*), closely related to the Horan dialects;
 - e. the south Transjordanian dialects (*bigūl*), that are influenced by the Hijazi Bedouin dialects of Arabia Petraea and represent a mixed dialect type;
3. The Bedouin dialects, which are in turn divided into:
 - a. The dialects of Negev (*bigūl*), which show some features typical of the sedentary dialects (namely the *b*-imperfect). The dialects of this group typologically belong to the Sinai type, and they exhibit some similarities with the Bedouin dialects of Arabia Petraea;
 - b. The dialects of Arabia Petraea (*yigūl*), they display some affinities with the Hijazi dialects;
 - c. The dialects of the Syro-Mesopotamian sheep-rearing tribes (*yigūl*), spoken in Transjordan, “(they) belong to the same type as the rest of the dialects of the sheep-rearing tribes in the Syrian and Mesopotamian peripheries of the Syrian Desert” (Palva 1984, p. 372).
 - d. The dialects of the North Arabian Bedouin type, *yigūl*, spoken in Transjordan by the Sirhān, the Bani Šaxar and the Bani Xālid.

Herin et al. (2022) have recently implemented Cantineau, Cleveland and Palva’s groupings with new elements and obtained the following categories¹:

- (I) Sedentary *bigūlu*: Mu²ābi and Balgāwi-Ḥōrāni;
- (II) Southern Bedouin *yigūlu* (Ḥwētāt, Bdūl, Zawāyda, etc . . .);
- (III) Southern Bedouin *bigūlu* (mostly Nagab and Sinai);
- (IV) Central Bedouin *yigūlu* (ʕAğārma, ʕAdwān, ʕAbābīd, etc . . .);
- (V) Northern Bedouin *yigūlūn*: ʕNizi, Šammari, Bc (Misāʕid), Šāwi, Ca (Bū ʕId et ʕIdīn in Lebanon, so far unattested in Jordan)

Their “taxonomy” has the merit to account for the changes caused by recent dialect contacts between speakers of Bedouin sub-groups varieties but does not take into consideration the effects of the increasing dialect mixing between the Bedouin and the sedentary population, which contributed to blur the linguistic boundaries set by this dichotomy. In particular, some Bedouin tribes such as the Bani ʕAbbād (or ʕAbābīd), who live in the vicinity of the big urban centers, start to assimilate the speech habits of neighboring communities and lose the authentic *badawī* features that characterize their dialects.

During the last century, most of the Jordanian tribes, except for some living in the southern districts of the country (like the Bdül), abandoned their nomadic lifestyle and settled in some definite areas in semi-sedentary or sedentary conditions.

Many economic and social factors have contributed to intensifying the contacts between Bedouin and sedentary variations and the tendency to dialect levelling or koineization. Among these, the most relevant factors were: the increasing contacts with the inhabitants of the cities of Salt and ʿAmmān and those with the refugees coming from Palestine and Syria, the growing access to education, the proliferation of television series produced in Egypt, Syria, and Lebanon, and in some cases even inter-marriage with sedentary people (Rosenhouse 1984).

This is particularly evident in the dialect of the ʿAbābīd, in central Jordan: their dialect is deeply affected by a levelling process toward sedentary dialects that has left traces in phonology, morphology and syntax that will be discussed in detail in the section *Results*.

In the light of these findings, the aim of this study is to question the status of the *šāwi yigūl* dialect type that is geographically and historically attributed to this vernacular and to point out that the adoption of a more refined classification would better account for mixed dialects which display important markers of both Bedouin and sedentary types.

2. Materials and Methods

This article is based on a corpus consisting of 20,000 words² of transcribed unmonitored interviews³ of both men and women.

The speakers selected for this qualitative analysis belong to the Bani ʿAbbād tribe, are distributed across different ages (the youngest girl was 5 years of old at the time of the interviews and the oldest man was 94 years old) and display various educational and socio-economic backgrounds.

The data resulted from two fieldwork campaigns carried out in July and August 2016 and from January until August 2017.

In order to limit the impact of my presence on the oral productions of the informants, the interviews⁴ were carried out by Jordanians⁵: a boy from Karak, and one belonging to the semi-nomadic tribe of the ʿAğārma, as well as an ʿAbbādi girl.

2.1. The Tribe

The Bani ʿAbbād are a confederation that is divided into two main groups: al-Ġburiyya and al-Ġrumiyya (Peake 1958, p. 166). As noted by Shryock (1997, p. 40), “over time the ʿAbābīd have been internally fragmented and politically weak. [...] This lack of consensus is commonly attributed to the diverse genealogical origins of the tribe’s clans”.

According to Oppenheim (1943, p. 227), the areas most densely populated by the Bani ʿAbbād were Māḥiṣ, Wād is-Sīr, ʿArāg al-ʿAmīr, ʿĀrḍa and al-Ġōr. However, nowadays some branches of the tribe also live in Bader al-Ġadīda, aṣ-Ṣḥabīḥi, aṣ-Šūna aġ-Ġanūbiyya, al-Karāma, Wādi aṣ-Šitā, Yarga, Marġ al-Ḥamām, ʿĪra, and Nāʿūr, in the periphery of ʿAmmān and in part in Salt (Figures 1 and 2).

Most of the land in the Balga governorate belongs to the ʿAbābīd and indeed, a common Jordanian saying recites *ʿAbbād min sīl ʿa-s-sīl* (ʿAbbād from river to river), alluding to the fact that their tribal territory extends from the Zarqa Torrent to the Jordan River.

The exact number of members of the tribe is hard to estimate since there are no official censuses available. Nevertheless, according to the data recorded during my fieldwork the tribe could be composed of approximately 350,000 people.

Shryock (1997, p. 43) reports that until the 1950s most of the ʿAbbādi clans were still living in tents (*byūt aṣ-šaʿr* in Arabic). They used to spend the winter in the Jordan Valley (locally called Ġōr or Aġwār), where they grazed sheep and goats, whereas during the spring and summer months they moved to its crest (the *šifa*), where the temperatures are cooler.

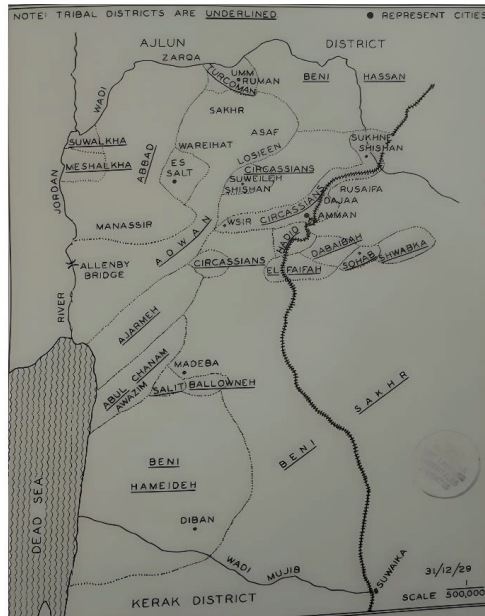


Figure 1. Tribal map of the Balga District, Jordan—Adapted from Peake (1958, p. 253).

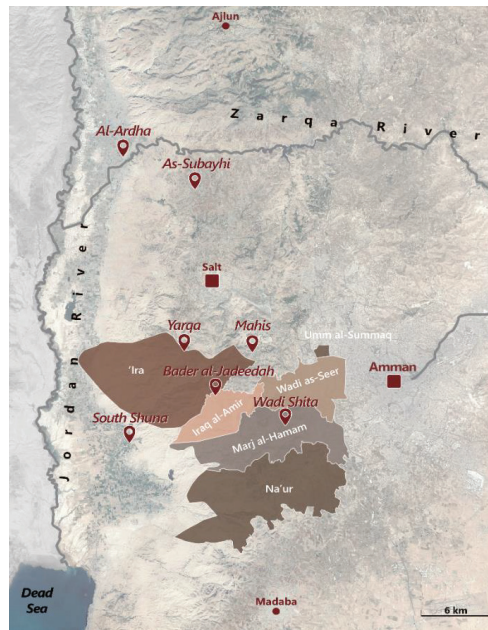


Figure 2. Reconstruction of the Bani 'Abbād *dīra* based on the information gathered during my fieldwork and realized by A. Cristaldi.

Unlike the desert tribes, the Bani 'Abbād had always been engaged in farming and in the late 1960s some members of the tribe started to prefer a more sedentary way of life. However, according to Shryock (1997, p. 46) even “in the midst of this rapid change, the

Balga tribes [have always] consider[ed] themselves fully Bedouin”and their identity is strictly linked with the values of *karāma* (generosity) and *ʿaṣābiyya* (inter-tribal solidarity).

As stated by Sakarna (1999, p. 8) “today, the people of ʿAbbādi tribe are no longer nomads and mostly live in settlements. They work as government employees, military individuals, farmers, and in other kinds of occupations”.

2.2. Functional Framework

The research method applied in this work relies on descriptive linguistics which is “based on the empirical observation of regular patterns in natural speech” (François and Ponsonnet 2013, p. 184).

The present linguistic analysis is built on a corpus of narrative transcribed texts (20,000 words) taken from the recordings of unmonitored speech. After collecting this body of data, it was segmented through the program ELAN, and analyzed in order to “identify the distinctive component of the system and the principles that underline its organization” (François and Ponsonnet 2013, p. 184).

The functional framework used for my analysis combines typological and discourse-based approaches, since this “provides the tools necessary to address questions of the meanings underlying language variation” (Brustad 2000, p. 7).

In order to efficiently describe and explain the phonological and morphological variation of the dialect of the Bani ʿAbbād, the methodology adopted was to synthesize those concepts that are able to most efficiently account for the data recorded during my fieldwork. It especially relied on the doctoral thesis of Herin on the dialect of the city of Salt (which to date constitutes the only Jordanian vernacular exhaustively described), and the work of Haspelmath and Sims (2010), *Understanding Morphology*.

While examining the syntactical structures of the data corpus collected, I referred in particular to *Éléments de syntaxe générale* and *Syntaxe Générale, une introduction typologique, Tome 1 and 2* by Creissels (1995, 2006a, 2006b, respectively), and *The Syntax of Spoken Arabic* by Brustad (2000).

3. Linguistic Analysis

Although in 1999 Sakarna analyzed some phonological aspects of the dialect of the Bani ʿAbbād, in particular those of the As-Sakarna branch, no in-depth study had been carried out on the dialect of the Bani ʿAbbād tribe when in 2016 I started to gather the material for my master’s thesis *Le dialecte des Bani ʿAbbād: Analyse des traits phonologiques, morphologiques et syntaxiques discriminants*⁶ (Torzullo 2018).

According to the information reported in Cantineau (1937), the above-mentioned classifications provided by Cleveland (1963) and Palva (1984), and the description of the dialect of the semi-nomadic al-ʿAğārma tribe (Palva 1976), my initial assumption was that the dialect of the Bani ʿAbbād typologically belonged to the Jordanian *yğül*-group which is part of the so-called *Šāwi* dialects (in French also called *petits-nomades*).

My findings show that this dialect has a number of characteristics belonging to the small-cattle nomads described by Cantineau, such as the gender distinction in the 2nd and 3rd pl. persons, but they point out that it is deeply affected by a koineizing tendency that caused some significant changes in the structure of the vernacular and a loss of some authentic features.

This phenomenon is due to the increasing contact with the sedentary population of the neighboring areas of ʿAmmān and Salt, especially due to economic and educational reasons: most of the members of the tribe have access to higher education, which brings them to study in the capital, where the nearest universities are to be found. Furthermore, many ʿAbbābid work in these two cities and consequently have daily exchanges with people who do not speak their same variety of vernacular, and in some cases do not even understand their original dialect. These circumstances force them to opt for a more common variety, or at least to put aside some of the original traits of their vernacular and adopt some *madani*

(urban) linguistic features, in order to facilitate the communication with the outsiders of the tribe.

If, on the one hand, this convergence to a “common linguistic style” (Giles and Ogay 2007, p. 296) for the sake of intelligibility has improved the effectiveness of their contacts, on the other, this “accommodative code variation” (Giles et al. 1973, p. 179) has led over time to a penetration of sedentary elements in the phonology, morphology, and syntax of their speech and to the disappearance of distinctive *šāwi* marks, mainly among the youngest members of the tribe.

3.1. Contact Induced Changes in the Contemporary Dialect of the Bani ^šAbbād

3.1.1. Phonology and Phonotactics

Reflexes of OA /q/

Old Arabic (OA) /q/ is usually realized as a voiced velar stop /g/ in the dialect of the Bani ^šAbbād. Ex: *gawi* ‘strong’, *garye* ‘village’, *manṭega* ‘region’, *daḡṭga* ‘minute’, *galb* ‘heart’, *giṣṣa* ‘story’.

However, it does not occur as a palatal variant /ǧ/ in the contiguity of front vowels, as it certainly was in the 1960s and 70s in the *yigūl* dialect of the ^šAḡārma tribe who also lives in the Balga District.

In the corpus under analysis there are only four instances of this phenomenon: two of them, ^š*eǧib* ‘young children’ and *ǧiddāmi* ‘in front of me’, were produced by the oldest members of the tribe interviewed, a woman and a man of 90 and 94 years old respectively, and the latter two, *ǧider* ‘cauldron, copper pot’ and *šigǧ* ‘part of the tent reserved to men’, were obtained through elicitation.

The words *ǧiddāmi* and *ǧider* occur along with their variants *geddāmi* and *gider* where the velar is not realized as affricate. This alternation can be considered as an example of “stylistic contrasts *plain colloquial* vs. *koineized colloquial*” (Palva 1976, p. 10).

Reflexes of OA /k/

The affricate /č/ of OA /k/ shows a regressive character within this dialect. It rarely appears in spontaneous speech, and when occurring it is only used by the oldest speakers: *iḥčilha* ‘tell her’, *ṭhači* ‘you (f.) speak’, *čil* ‘measure, weigh’, *čif* ‘how’, *bičila* ‘type of Bedouin dessert’, *čānna* ‘daughter-in-law’, *hēč* ‘thus’, *čalb* ‘dog’, *yčammilen* ‘they (f.) complete’, *čam* ‘how much’, *čilīr* ‘a lot’, *mičān* ‘place’.

Given the number of instances of this variant it is possible to conclude that these lexemes are “vestigial variants” and represent some “fossilised traces of an earlier dialect system when such forms were more general” (Trudgill 1999, p. 321).

Even the morphological contrast between *-k* and *-č*, which allows the opposition of the pronominal suffix of the 2nd m. sing. person and the pronominal suffix of the 2nd f. sing. person, is poorly attested⁷. It occurs in the speech of the older and less-educated speakers, while the youngest and those who work or study in ^šAmmān replace the allomorph *-č* with *-ki*.

The recessive status of the affricate /č/ is particularly evident within a family I interviewed: the use of the affricate /č/ gradually disappears over the three generations of women who live in the same household⁸:

- I. The grandmother regularly uses it in the contiguity of a front vowel, and she uses *čān* even to express the verb ‘to be’.
- II. The mother uses the affricate /č/ especially to contrast the pronominal suffix of the 2nd m. sing. person and the pronominal suffix of the 2nd f. sing. person; she never uses *čān* to express the verb ‘to be’, and the only word where she employs /č/ is *čānna* ‘daughter-in-law’, while reciting a poem.
- III. The daughter never uses the affricate /č/.

This example clearly showcases the shift from a purely Bedouin trait of the dialect to a more mixed type one by sedentarization.

Reflexes of the Old Diphthongs /ay/ and /aw/

The monophthongisation in /ī/ and /ū/ of the old diphthongs /ay/ and /aw/ does not appear as a prominent trait in this dialect.

In *Studies in the Arabic Dialect of the Semi-Nomadic ʿal-ʿAğārma Tribe*, Palva (1976, p. 19) reports: “one of the most striking characteristics distinguishing the dialects of the nomadic type from those of the sedentary type in the Syro-Palestinian dialect area is the fluctuation /ē/-/ī/ and /ō/-/ū/”.

This feature is briefly mentioned by Cantineau in *Les parlers arabes du Ḥōrān* (Cantineau 1946, p. 156), who refers to it as occurring in sporadic cases, while Bettini (2006, p. 30) only accounts for a fluctuation between the realizations /ē/ and /ī/ of /ay/ (which appears as a common trait in the dialects analysed in *Contes Féminins de La Haute Jézireh Syrienne*) and a long vowel /ō/ for the diphthong /aw/.

Behnstedt (1997, pp. 62–63), on the other hand, does not report this fluctuation at all in the Map31 of his *Sprachatlas von Syrien*.

In the corpus under analysis there are only four occurrences of monophthongisation in /ī/ and /ū/ occurring in natural speech: *čīl* ‘measure, weight’, *čif* ‘how’, *đif* ‘guest’ and *hūšāt* ‘fights’, together with a sentence reported by a man describing the Bedouin traditions concerning the drinking of coffee:

- | | | | |
|----------------------------|-----------------|------------------|-------------------|
| (1) <i>Ygūl-lak</i> | <i>awwal</i> | <i>finḡān</i> | <i>la- ǰ- đif</i> |
| to _say.IPFV.3msg-DAT.2msg | first | cup | for-ART-guest |
| <i>tāni</i> | <i>la-s-sif</i> | <i>at- tālīt</i> | <i>la-l- ħif</i> |
| second | for-ART-sword | ART-third | for-ART-injustice |
- One says: the first cup is for the guest, the second to talk about war, and the third to discuss injustice.

In addition to these instances, two examples of a fluctuation between the realizations /ē/ and /ī/ were obtained by elicitation: *līl~lēl* ‘night’ and *zīt~zēt* oil’.

According to the tendencies that appear in the data, the realizations /ē/ and /ō/ of the old diphthongs /ay/ and /aw/ gained ground in the dialect of Bāni ʿAbbād, to the detriment of the most archaic ones. So, it is possible to observe: *bēt* ‘house, tent’, *bēn* ‘between’, *ḡēr* ‘other’, *xēr* ‘good, well’, *xēl* ‘horse’, *ʿēla* ‘family’, *ʿēn* ‘eye’, *tōr* ‘bull’, *yħōšen* ‘they (f.) plough’, *zōḡa* ‘wife’, *xōf* ‘fear’, *fōg* ‘above’, *tōr* ‘cave’.

This state of the affairs suggests that the above-mentioned statement by Palva (1976) concerning the monophthongisation in /ī/ and /ū/ does not currently hold true for this dialect. The few instances of the phonemes /ī/ and /ū/ (</ay/ and /aw/) may thus represent the last traces of an older monophthongisation system where such forms were more widespread, and that in the last decades have been replaced by more sedentary forms.

Cases of Trochaim

According to Cantineau (1936, p. 114), trochaim is a typical trait of some *petits nomades* dialects, namely those of the ʿŌmūr, Šlūt, Bani Xāled and Sirhān.

Dialects characterized by a trochaic rhythm maintain the older Arabic inflectional -a as well as the -a of the feminine morpheme between a long syllable and a pronominal suffix (Palva 1976, p. 25), i.e., CvCC or CṽC + pron. suff. > CvCCA + pron. suff. & CṽCA + pron. suff.

In my data this phenomenon is poorly attested (see Tables 1 and 2), and it only rarely appears in the speech of the oldest speakers:

Table 1. Trochaic occurrences in verbs.

<i>yčammilen</i> ‘they (f.) finish’ 3	<i>yḡammirin</i> ‘they (f.) collect’ 1
<i>ytālifu</i> ‘they (m.) gathered’ 1	<i>rākib-u</i> ‘he mounted it’ 2

Table 2. Trochaic occurrences in nouns and prepositions.

<i>ahl-a-ha</i> ‘her family’ 6	<i>wağh-a-ha</i> ‘her face’ 1	<i>geddām-a-na</i> ‘in front of us’ 1
^ʕ <i>ašān-a-ha</i> ‘because of her’ 1		^ʕ <i>omr-a-ha</i> ‘her age’ 1

Given the number of instances of this trait, it is possible to conclude that nowadays the dialect of the Bani ^ʕAbbād belongs to the atrochaic group even if the vestigial variants found in the corpus suggest that a trochaic rhythm characterized the vernacular at an earlier stage. This change in the rhythm of the syllable is most likely due to the influence of the surrounding sedentary dialects that display an atrochaic pattern.

3.1.2. Morphology

Personal Pronouns

The independent forms of personal pronouns in the dialect of the Bani ^ʕAbbād are illustrated in Table 3:

Table 3. Independent personal pronouns.

1. s.	^ʔ <i>ana</i> ~ ^ʔ <i>ani</i>	1. p.	^ʔ <i>ihna</i> ~ ^ʔ <i>aḥna</i> ~ ^ʔ <i>əḥna</i>
2. m. s.	^ʔ <i>inta</i> ~ ^ʔ <i>inte</i>	2. m. p.	^ʔ <i>intu</i>
2. f. s.	^ʔ <i>inti</i>	2. f. p.	^ʔ <i>intin</i>
3. m. s.	<i>huwwa</i> ~ <i>hū</i>	3. m. p.	<i>humma</i> ~ <i>hummm</i>
3. f. s.	<i>hiyya</i> ~ <i>hī</i>	3. f. p.	<i>hinna</i> ~ <i>hinm</i> ~ <i>hunna</i>

It is possible to observe for the 1 s. an alternation between the forms ^ʔ*ana*~^ʔ*ani*.

According to the data concerning the dialect of the ^ʕAğārma, Palva (1976, p. 27) states: “it is impossible to decide whether [the first variant] is a traditionally genuine form or a loan from the neighbouring dialects [Bani Xaled, Sirḥān, etc.]”.

The second form, ^ʔ*ani*, is regarded by Younes and Herin (2016, p. 4) and Isaksson (1999, p. 59) as characteristic of the *šāwi* tribes, in free variation with the form ^ʔ*āni*, already mentioned by Cantineau (1937, p. 173). Behnstedt (1997, p. 501), B. Herin (2010, p. 46) and Al Tawil (2019, p. 138) report that this form is commonly found in the Ḥōrān (both Syrian and Jordanian) and that it also marginally appears in the dialect of Salt.

As for the dialect of the ^ʕAbābīd, it appears only in sporadic instances:

- in the negative copula *māni* ‘I am not’ (6)⁹;
- in the expression: *māni* ^ʕ*ārif* ‘not knowing’ (1), ^ʔ*ana*, *māni sāmi*^ʕ ‘not listening’ (1), ^ʔ*ani ma ḥibbo al-gatū* ‘I don’t like the cake’ (1).

The 1 p. also exhibits two variants: ^ʔ*aḥna*~^ʔ*əḥna* (49) and ^ʔ*ihna* (20). The form *hənnna* attested in the neighboring tribe of the ^ʕAğārma and phonetically associated with the *gahawa* syndrome is not attested. Palva (1976, p. 26) affirms that the form ^ʔ*əḥna* is “a stylistic variant” of *hənnna*, and that their respective use depends on the form found in the tribe to which the ^ʕAğārma address their speech.

It is possible to suppose that *hənnna* may also have been employed in the dialect of the Bani ^ʕAbbād but that due to the contact with the adjacent sedentary dialects it disappeared, while the borrowing ^ʔ*ihna* penetrated in the vernacular and started to become more popular, especially among younger speakers.

Regarding the bound pronouns, it is worth mentioning that due to the recessive character of -*č*, the original *šāwi* morphological contrast between the pronominal suffix of the 2nd m. sing. person -*k* and the pronominal suffix of the 2nd f. sing. person -*č*, is not consistent.

The occurrences of the feminine bound pronouns’ forms -*eč* and -*č* are in total 12. The only instance that occurs in free speech is *xuwayāteč* ‘your (f.) friends (m.)’ (2), while the

other words containing this bound pronoun are to be found in optative sentences, in a few words occurring in some lines of poems and in some traditional quotes: *salāma tsallmeč* ‘(may you) live in good health’ (1), (*Allah*) *yis^šid^š unreč* ‘may God make you (f.) happy for all your (f.) life’ (1), *salānteč* ‘I wish you (f.) good health’ (1), *š^šammeč* ‘your (f.) paternal uncle (1)’, *yxāfeč* ‘you (f.) are afraid’ (3), *la-hāleč* ‘alone (f.)’ (1), *š^šineč* ‘your (f.) eye’ (1), *š^šahebbeč* ‘I love you (f.)’ (1).

Thus, even if as noted by Palva (1976, p. 47), “there is [still] a (bedouinizing) tendency which is actualized in certain speech situations associated with traditional culture”, in this case religious formulae, poetry and citations of famous quotes, the allomorph -č is nowadays being replaced by the sedentary form -ki.

Interrogative Pronouns

The interrogative pronouns occurring in the data collected are reported in Table 4:

Table 4. Interrogative pronouns.

Who ... ?	<i>mīn; min</i>
What ... ?	<i>šū~šu; wēš; əš~[?]ēš</i>
Which ... ?	<i>[?]ayy(a)</i>
How ... ?	<i>čif; šlōn; kef~kif</i>
Where ... ?	<i>wēn</i>
When ... ?	<i>mata; mita; [?]ēmta; waymat</i>
Why ... ?	<i>lēš; lwēš</i>
How much ... ?	<i>gaddēš~geddēš; čam; kam</i>

The traditional Bedouin forms (in bold in Table 4) are very marginal in the dialect of the Abābid: *min* (3), *wēš* (1), *čif* (1), *lwēš* (6), and they are to be found only in the recordings of the older speakers.

The variants *mita; [?]ēmta; waymat* for ‘when’ and *čam* for ‘how much’ were obtained only through elicitations and were never used by the speakers in natural speech.

As for *šlōn*, its use seems to be limited to the question *šlōn-ak/ič?* ‘how are you’ and it does not occur in other positions.

Sedentary forms (which are to be considered as loans from the dialects of ^šAmmān and Salt) are the most attested ones in the corpus, and in many cases, they have already completely replaced their genuine nomadic counterparts.

Verbs C1=[?]

The weak verbs C1=[?] *axač-yāxuč* ‘to take’ and *akal-yākul* ‘to eat’ are as a rule reinterpreted as III *w/y* verbs in *šāwi* dialects (Younes and Herin 2016, p. 10) and realized as *kala* (ou *čala*) et *xača*.

In respect to this trait, Cantineau (1936, p. 87) writes: “cette conjugaison caractérise d’une façon remarquable les parlers de nomades moutonniers et les oppose aux parlers de sédentaires syro-palestiniens, qui ont toujours [?]*akal*, [?]*axač*”.

However, as illustrated in Tables 5 and 6, the dialect of the Bani ^šAbbād shows the coexistence of two patterns for these two verbs, in both perfective and imperfective: one typically Bedouin (*xača-kala/yāxuč-yākul*) and one sedentary ([?]*axač-[?]akal/yōxuč-yōkil*).

Table 5. Inflection of perfective.

1. sg. <i>xadēt~[?]axaḍt</i>	<i>kalēt~[?]akalt</i>	1. pl. <i>xadēna~[?]axaḍna</i>	<i>kalēna~[?]akalna</i>
2. m. sg. <i>xadēt~[?]axaḍet</i>	<i>kalēt~[?]akalet</i>	2. m. pl. <i>xadētu~[?]axaḍtu</i>	<i>kalētu~[?]akaltu</i>
2. f. sg. <i>xadēti~[?]axaḍti</i>	<i>kalēti~[?]akalti</i>	2. f. pl. <i>xadēten~[?]axaḍten</i>	<i>kalēten~[?]akalten</i>
3. m. sg. <i>xaḍa~[?]axaḍ</i>	<i>kala~[?]akal</i>	3. m. pl. <i>xaḍu~[?]axaḍu</i>	<i>kalu~[?]akalu</i>
3. f. sg. <i>xaḍat~[?]axaḍat</i>	<i>kalat~[?]akalat</i>	3. f. pl. <i>xaḍēn~[?]axaḍēn</i>	<i>kalēn~[?]akalen</i>

Table 6. Inflection of imperfective.

1. sg. <i>āxuḍ~ōxuḍ</i>	<i>ākil~ōkil</i>	1. pl. <i>nāxuḍ~nōxuḍ</i>	<i>nākul~nōkil</i>
2. m. sg. <i>tāxuḍ~tōxuḍ</i>	<i>tākul~tōkil</i>	2. m. pl. <i>tāxdu~tōxdu</i>	<i>tāklū~tōklu</i>
2. f. sg. <i>tāxdi~tōxdi</i>	<i>tākli~tōkli</i>	2. f. pl. <i>tāxḍin/en~tōxḍin</i>	<i>tāklin/en~tōklin</i>
3. m. sg. <i>yāxuḍ~yōxuḍ</i>	<i>yākul~yōkil</i>	3. m. pl. <i>yāxdu~yōxdu</i>	<i>yāklū~yōklu</i>
3. f. sg. <i>tāxuḍ~tōxuḍ</i>	<i>tākul~tōkil</i>	3. f. pl. <i>yāxḍin/en~yōxḍin</i>	<i>yāklin/en~yōklin</i>

The forms reported by Cantineau (1936, p. 87) were fully confirmed by elicitation (even with some minor differences), while the instances contained in the recorded corpus are mixed:

- (2) [?]*alwēh ani min [?]arab* I wish I was of the people
Hadla u la min xaḍa garāyibha of Hadla and not of those who took her relatives
La (a)hutṭ[?] ana min ad- dahab ganṭār so I [could] put a quintal of gold
W aṭla[?] min al-lōm [?]axṭobha and get rid of the blame and betroth her
- (3) *rāḥ [?]and aš-šēx ṭalabha, iṣi ḥabbha al-muhimm, banāt garāybo w an-nās illi yaḡrabūlo mā ḥabbu*
inno yḥibbha la[?] anno ṭāri u mrattab ḥabbu inno yōxiḍ minhum u mā xaḍa minhum.

He went to the sheikh to ask for her hand, what is important is that he loved her, the girls of his relatives and the people who were related to him, did not like [the fact] that he loved her, because he was rich and wealthy, [so] they wanted him to take one of them, [but] he did not choose one of them.

- (4) *axaḍu* *humma* *al-ḡanūb* *kāmel*
to_take.PFV.3mpl they.m ART-south complete
They took over the whole South.
- (5) *ḡabel* *al-[?]eḡib* *bass* *yākol* *xobez*
before ART-children but to_eat. IPFV.3msg bread
min *[?]a- ṣ- ṣayḡān* *mā[?]a* *šāy*
from on-ART-ṣayḡān with tea
Before the children only ate bread made on the *ṣāḡ¹⁰* with tea.
- (6) *al-aḡnām* *yōklu* *min-ha* *aḡtar* *xuḍār* *mā* *fi*
ART-sheep to_eat.IPFV.3mpl of.3fsg most vegetables NEG there were
They ate mostly sheep (meat), there were no vegetables.

The traditional paradigms still represent the majority of occurrences in the texts (52.5%), however, it is possible to observe a growing use of the sedentary forms (47.5%).

It is interesting to notice that the morphological variation between the two forms occurs only in natural speech but not when reporting some highly traditional tales or poems.

In this respect, Henkin (2010, p. 219) reports: “of all the oral registers, it is vernacular that reacts most significantly to dialectal and demographic variables [. . .]. In contrast, the traditional registers of oral literature, including oral narrative and oral poetry [. . .] are less affected by everyday communicational needs”.

The use of one variety over the other depends on sociolinguistic factors, namely age and level of education. The use of sedentary paradigms is increasing, and they are replacing

the traditional forms mostly in the speech of the youngest generations, who have greater contact with the inhabitants of the biggest urban centers.

Prepositions

The preposition ^ʕ*ugub* ‘after’ represents a vestigial variant in the dialect of the Bani ^ʕAbbād, who nowadays use the form *ba^ʕd*.

According to Herin (2010, p. 132) “ dans le parler traditionnel, *ba^ʕd*- signifie ‘encore’ et non ‘après’ qui est une influence des parlers urbains (^ʕAmmān et Palestine)”.

In the corpus under analysis, ^ʕ*ugub* occurs only twice, in the speech of the oldest people interviewed, i.e., a woman and a man of 90 and 94 years old, respectively:

- | | | | | |
|-----|--|---------------------|-------------------------------|--------------------------|
| (7) | <i>hād</i> | <i>faddān</i> | <i>ad-darrāsāt</i> | ^ʕ <i>ugub</i> |
| | DEM.msg | pair of oxen | ART-harvester | after |
| | | <i>yudrusu</i> | ^ʕ <i>a-l-ḥamīr</i> | <i>w</i> <i>al-bugar</i> |
| | Before | to_thresh.IPFV.3mpl | on-ART-donkeys | and ART-cows |
| | After one threshed with a pair of oxen, [but] before with some donkeys and cows. | | | |
| | | | | |
| (8) | <i>Mīn</i> | <i>ḥakam</i> | ^ʕ <i>ugubu-hum</i> | <i>al-inglīz</i> |
| | who | to_rule.PFV.3msg | after-3mpl | ART-English |
| | Who ruled after them? The English. | | | |

This preposition was probably more widespread some decades ago when its sedentary counterpart *ba^ʕd* had not penetrated so deeply in the vernacular. The current status of ^ʕ*ugub* shows once again the linguistic changes in progress in this dialect, in the direction of a de-bedouinization of the most conservative and traditional forms, which facilitates communication with those outside the tribe.

Conditional Conjunctions

The most frequently attested conditional conjunction in the data collected is ^ʔ*ida* (39):

- | | | | | | |
|------|--|----------------------------------|----------------------------|--------------------------|---------------------------------|
| (9) | ^ʔ <i>ida</i> | ^ʔ <i>ḡa</i> | <i>dēf</i> | <i>min</i> | <i>barra</i> |
| | if | to_come.PFV.3msg | guest | from | outside |
| | <i>ma</i> | <i>ysawwū-lo</i> | <i>akel</i> | ^ʕ <i>ādi</i> | <i>yīdḥāhū-lo</i> |
| | NEG | to_make.IPFV.3mpl-for-3msg | food | simple | to_slaughter.IPFV.3mpl-for-3msg |
| | | | | | |
| | <i>xartīf</i> | <i>yīdḥāhū-lo</i> | <i>na^ʕḡe</i> | ^ʕ <i>anz</i> | |
| | ram | to_slaughter.IPFV. 3mpl-for-3msg | sheep | goat | |
| | If a guest comes from far away, one does not prepare for him a normal meal, one slaughters for him a ram or a sheep or a goat. | | | | |
| | | | | | |
| (10) | ^ʔ <i>ida</i> | <i>enta</i> | <i>ma</i> | <i>hazzēt</i> | |
| | if | you | NEG | <i>to_shake.PFV.2msg</i> | |
| | <i>al-fīnḡān</i> | <i>yḡall</i> | <i>yṣubbū-l-ek</i> | | |
| | ART-cup | to_keep.IPFV.3msg | to_pour.IPFV.3msg-for-2msg | | |
| | If you don’t shake the cup, one keeps pouring you (coffee). | | | | |

However, this form does not belong to the original ^ʕAbbādi repertoire, but it is a progressive sedentary variant that supplanted the traditional form to introduce conditional clauses, *čān*, which has never been employed by the speakers.

3.1.3. Syntax

Genitive Exponent

According to Younes and Herin (2016, p. 12) the most frequent form of genitive exponent found in most ^ʕ*āwi* dialects is *ḡiyy*.

This form, first reported by Cantineau (1946, p. 204), was also attested by Cleveland (1963, p. 61), who defined it as characteristic of the Jordanian Bedouin dialects.

In the dialect of the ⁵Abābīd there are no instances of this local variant, and the only occurrences of a genitive exponent in natural speech is the more general and widespread Levantine *taba⁵*, which has to be regarded as a borrowing from the adjacent urban dialects:

- | | | | |
|------|---|---|---|
| (11) | <i>kānu</i>
to_be.PVF.3mpl
<i>bi-l-maḡrāfe</i>
in-ART-ladle
They stirred those with the wooden ladle | <i>yħarrkū</i>
to_stir.IPVF.3mpl-3msg
<i>taba⁵at¹¹</i>
<i>taba⁵.fsg</i> | <i>ħaḏōla</i>
DET.pl
<i>al-xašab</i>
ART-wood |
| (12) | <i>al-gider</i>
ART-cauldron
<i>bī-ha</i>
in-3fsg
The cauldron is a big pan where they put the grain of wheat | <i>ṭaṅḡara</i>
pan
<i>al-ḡarīša</i>
ART-grain | <i>kbīra</i>
big
<i>taba⁵at</i>
<i>taba⁵.fsg</i> |
| | | <i>yħuṭtu</i>
to_put.IPVF.3mpl
<i>al-gameḥ</i>
ART-wheat | |

The forms *giyy* and *šiyy*¹² were productive at an earlier stage, as reported by a speaker of ⁵Arāḡ al-²Amīr, but today they are only used in the field of trade:

- (13) *Dār gīti dār taba⁵ti, hal-kalime kānat maḡḡūda, gīti aw šīti, ya⁵ni mulki w ili, ā zayy ħēk ā, al-banāt giyyāti aw ḥalālāti ya⁵ni ili ḥatta⁵an al-ḡanam yḡūlu-lo giyyāti, kānat maḡḡūde ḥassā⁵ bista⁵ malūha at-tuḡḡār⁵ enna, maṭalan yḡūl lak² ana gayyāti 3 alāf² ana gayyāti 10 alāf, ya⁵ni flūsi.*

Dār gīti¹³ or *dār taba⁵ti* ‘my house’, *gīti* or *šīti*, this word existed, it means of mine, mine, like this yes, *al-banāt giyyāti* ‘my daughters’ or *ḥalālāti* ‘my cattle’ for example, one also used to say *al-ḡanam giyyāti* ‘my sheep’, this existed but now businessmen use it, to say my 3 or 10 thousand, to indicate my money.

It is possible to conclude that the progressive variant *taba⁵* superseded the equivalent traditional forms of genitive exponents in the dialect of the Bani ⁵Abbād.

A similar scenario is also attested by Procházka (2018), Younes (2014) and Younes and Herin (2013) for the *Šāwi* Bedouin dialects spoken in Syria¹⁴ and for those spoken in Lebanon by the Abu ⁵Id and the ⁵Atīḡ.

This state of the affairs confirms the analysis of Palva (1982, p. 28), who states that “the genitive exponent belongs to the features in the dialects that are particularly exposed to koineization”.

Negation

The negation in the dialect of the Bani ⁵Abbād has also undergone a process of sedentarization.

It is possible to notice that the negative form *miš* has gained ground in the non-verbal negation at the expense of the genuine negative particles *mā* and *mū* (Palva 1976, p. 42; Al Tawil 2021, p. 23). These last variants represent, respectively, 16.45% and 13.92% of non-verbal negation in the data:

- | | | | | |
|------|--|---|--|--|
| (14) | <i>⁵ašā²ir</i>
tribes
<i>bēn-na</i>
between-1pl
There were no problems between us and the tribes of the South. | <i>al-ḡanūb</i>
ART-South
<i>u</i>
and | <i>lā²</i>
NEG
<i>bēn-hum</i>
between-3mpl | <i>mā</i>
NEG
<i>mašākil</i>
problems |
| (15) | <i>gabel</i>
before | <i>yinsawi</i>
to_make.PASS.3msg | <i>šaḡge</i>
<i>šaḡge¹⁵</i>
NEG | <i>mū</i>
NEG
<i>dabka</i>
dabka |
- Before one [used to] dance the *šaḡge* and not the *dabka*.

In total, 63.29% of this type of negation is constituted by the item *miš* and the remaining 5% by *muš*, which has to be regarded as an older borrowing as it is already attested in Ḥōrān (Cantineau 1946, p. 389):

- (16) *hāy* DEM.3fsg *əl-muṣṭalahāt* ART-terms *hāy* DEM.3fsg *alān* now *ixtafat* to_change.PFV.3fsg
min from *ḡīlna* generation-1pl *miš* NEG *maṭwǧūde* present.3fsg *ya^ʕni* it is to say
 These terms here have now disappeared from our generations, they are no more present.
- (17) *ḏahab* gold *ṭala^ʕ* to_appear.PFV.3msg *zayy* like *hǧār* stones *miš* NEG *ḏahab* gold
 The gold appeared in the form of stones not gold.
- (18) *muš* NEG *al-kull* ART-all *kwayyis* good *muš* NEG *al-kull* ART-all *mirtāḥ* wealthy
 We do not all have a good financial situation; we are not all wealthy

It is also interesting to point out that the negative structure *mā fi-š* began to spread in the speech of the youngest speakers of the tribe¹⁶:

- (19) *mā* NEG *fi-š* there is-NEG *ma^ʕo* with-3msg *mašāri* money
 He has no money.
- (20) *mā* NEG *fi-š* there were-NEG *maḥākim* tribunals *mā* NEG *fi-š* there were-NEG *muḥāmiyyīn* lawyers
 [Before] there were no tribunals nor lawyers.

According to Palva (1976, p. 42) the nominal negation *miš* and the structure *mā fi-šš* are to be considered as K-forms borrowed from the neighbouring sedentary dialects and not as genuine Bedouin negation.

However, the high frequency of the use of *miš* suggests that this negation has already been integrated into the dialect of the Bani ^ʕAbbād and that the other particles, i.e., *mā*, *mū*, which used to be employed in non-verbal negations, are to be regarded as vestigial variants.

b-Imperfect

Cantineau (1936, p. 83) affirms: “Le *b*-préfixe de l’inaccompli, si caractéristique des parlers des sédentaires syro-palestiniens, fait entièrement défaut dans les parlers des nomades”. This statement, however, does not hold true anymore for the dialect of the Bani ^ʕAbbād.

In fact, the use of the morpheme *b*- prefixed to the imperfect is no longer excluded from this Bedouin dialect, and it is possible to find 53 occurrences of this trait in the corpus:

- (21) *humma* they *bistagbilu* to_welcome.IPFV.HAB.3mpl *aḏ-ḏyūf* ART-guests
 They welcome the guests.
- (22) *biḥuṭṭi* to_put.IPFV.HAB.2fsg *al-laḥme* ART-meat *w* and *al-snōbar* ART-pine nuts
 You add the meat and the pine nuts.
- (23) *ayyām-ha* days-DEM *ən-nās* ART-people *mā* NEG *ma^ʕā-ha* with-3fsg
drāsa education *mā* NEG *bta^ʕrif* to_know.IPFV.HAB.3fsg *iši* thing
 In those days people were not educated, they didn’t know anything.

The adoption of the *b-imperfect* to mark habitual actions which take place in the present can be considered as an ongoing process in the dialect of the ^ʕAbābīd, and thus it is not rare to find some speakers using both forms (with and without *b*- prefix) in one and the same sentence:

- (24) *Fa yitğawwaz az-zalama . . . ʔiða kân fi ʕindhum dâr biskin bi-nafs ad-dâr ygullak waladna mâ biṭla^ʕ barra ya^ʕni muḥarram inno yitla^ʕ yiskin xârğ al-manteğa illi humma sâknîn fiha . . . bi-z- zabt̄... hây al-^ʕâda.*

Then the man gets married . . . if they have a house, he lives in the same household, one says: our boy does not go elsewhere. That is to say that it is forbidden (for him) to go live outside the area where they live . . . Exactly . . . this is the custom.

- (25) *zulum yṭlubu al-^ʕarūs bya^ʕṭi-hum*
 men to_ask.IPFV.HAB.3mpl ART-bride to_give.IPFV.HAB.3msg-3mpl
 Some men ask for the hand of the bride, he grants it to them.

Thus, it seems that, at this stage, this syntactic feature is used in free variation with the imperfect without *b-*, like among the Negev and Sinai Bedouins (Palva 1994, p. 462; Blanc 1970; De Jong 2000).

The *b-*imperfect can also be used to mark the progressive aspect of an action. However, in the corpus under analysis there is only one instance of this type of structure:

- (26) *biṭḍakkar waḥda bardō qişşa min al-qişşa*
 to_remember.IPFV.PROG.1sg one.f also story between ART-stories
 I also remember another story [now].

This sedentary trait has penetrated mostly the speech of the younger members of the tribe who have the tendency to conform more to the vernaculars of ʕAmmān and Saṭṭ, and so it reveals “a generational variation amongst the receptiveness of the borrowings” (Younes 2017, p. 136).

Future

The most prominent way in ʕāwi varieties to express future and volition is the use of the pseudo-verbs *widd-* and *rād-y(i)rād* ‘to want’¹⁷ (Younes and Herin 2016, p. 11). The first particle is well attested in the dialect of the Bani ʕAbbād while the verb *rād* is completely absent from it.

It is worth noticing that the pseudo-verb *widd-* occurs together with the sedentary variant *bidd-* all along the corpus:

- (27) *wədd-ak trūḥ ʕ-al-ḥakīm*
 want-2msg to_go.IPFV.2msg to-ART-doctor
 You will go to the doctor.
- (28) *bidd-i ōkil tīn*
 want.1sg to_eat.IPFV.1sg figs
 I will eat figs.

Since the number of instances of these two forms is almost equivalent¹⁸, it is possible to affirm that the Bedouin and sedentary varieties of this pseudo-verb coexist in this vernacular. However, their use remains dichotomous: it is considered as equally eligible by the speakers of the new generation, while the older ones (55 years and older) seem to prefer the genuine Bedouin item over the *madani* one.

4. Discussion

4.1. Bedouin or Sedentary?

In the light of the traits analyzed above, and the common division of the Arabic dialects available, the classification of the dialect of the Bani ʕAbbād is not straightforward.

According to the historical background and geographical position in the Balga district, it should typologically belong to the Bedouin Jordanian *ygūl-*type (part of the broader ʕāwi dialects’ *continuum*) or the recently theorized Central Bedouin *ygūlu* group. However, can it really be unproblematically labelled as such? Is it appropriate to still refer to this vernacular as a Bedouin dialect?

Such classifications which are based on typical, but nonetheless very few features, in fact do not account for all the discrepancies found in this vernacular and do not accurately allow to represent the actual state of the dialect.

As pointed out in the section *Linguistic Analysis*, this dialect diverges in many aspects from the “essential structure” (Palva 1969, pp. 14–15) of the *ygūl* group and it remains inaccurate to identify this dialect of the Jordan Valley with the *šāwi* label.

In fact, in addition to the elements already mentioned throughout the article, it is also possible to observe from Table 6 the lack of the formative /n/ in the 3. m. pl. of the imperfective. This constitutes a noticeable difference (*yāxdu* vs. *yāxḏūn*) with the typical inflexion found in *šāwi* dialects (Younes and Herin 2016, p. 8) and adds supplementary intricacy in cataloguing this vernacular. Furthermore, other genuine *šāwi* characteristics have been replaced by sedentary forms, while some typically *madani* items are being acquired in addition to the original Bedouin ones.

All the features taken into account for this analysis hint at highlighting that this dialect is in a situation of dialect contact and in an ongoing process of linguistic change.

Holes (1995, p. 278) notes that during the last 70 years “Jordan has experienced a long-term drift from the countryside and the desert into its towns and cities”, and massive migrations from neighboring countries, namely Palestine and Syria. As in other Middle Eastern states, “these social and political developments have had, and continue to have profound effects in spoken Jordanian Arabic varieties” (Holes 1995, p. 278), which in some cases tend to overlap more and more.

As Romaine (2010, p. 321) states:

“When groups in contact need to communicate, they have a number of possible choices. One is to use a lingua franca they both share [. . .]. A second option is for one or more parties to learn the other group’s language(s). In cases involving no substantial imbalances of power between the groups, stable multilingualism may result. However, where bilingualism is asymmetrical and the more powerful group imposes its language on a subordinate group, contact often leads to language shift or loss”.

The second scenario applies in this case. The members of the Bani ^ʕAbbād have day-to-day and long-lasting relations with the speakers of sedentary adjacent dialects and in order to facilitate the communication and mutual intelligibility avoid the use of traditional Bedouin features and adopt forms that can be easily understood by outsiders to the tribe¹⁹.

However, the linguistic patterns of accommodation and borrowings are not identical for all the people of the community (Palva 1982).

Traditional Bedouin features are disappearing faster from the speech of the young members of the tribe, and there was already a major tendency for women to make use of sedentary characteristics in the everyday vernacular 50 years ago (Palva 1976), since they appear more feminine and sophisticated than the respective Bedouin ones which relate to rough and tough masculinity (Holes 1995).

The borrowing of sedentary speech-habits does not mean that the Bani ^ʕAbbād “give up their language willingly, but continue transmitting [it], albeit in changed form over time” (Romaine 2010, p. 321).

The members of the tribe are aware of this state of the affairs and in fact one of the older members declares in one of the recordings:

- *Fi al-muṣṭalahāt hāḏi ya^ʕni wladna allī ^ʕomrhum ṭalāṭīn sana u xamsa u ^ʕaṣrīn sana mā kānu ya^ʕrīfūha li^ʕannha iltaḡat min zamān ā, iltaḡat min zamān.*
- *Ṭabb al-luḡa al-^ʕabbādiyya ṣam bitrūḥ?*
- *Mu^ʕḏamha hī trūḥ hassā^ʕ ṣārat madaniyya ma^ʕḏomha hassā^ʕ madaniyya . . . kull iṣi ^ʕindana ya^ʕni al-luḡa al-^ʕabbādiyya mā ḏall ma^ʕāha galil ḡiddan.*
- There are these terms [that] our children whose age is 30 or 25 years old do not know because they changed a while ago, yes, changed [a lot of] time ago.
- Well, so where is the ^ʕAbbādi language going?

- Most of it is disappearing, now it has become madani, most of it has become madani . . . all of our things, I mean there is very little left of the ^ʕAbbādi language.

These changes by sedentarization²⁰ result in a hybrid form of dialect which may be considered as a *dialect compromise* (Holes 1995) built on Bedouin and sedentary forms, which is not simple to define.

4.2. Emergence of a Mixed Dialect Type by Sedentarization

The members of the linguistic community in Jordan do not have equal command of all the varieties in use in the country. Thus, Bedouin speakers often accommodate to the more common speech patterns of the city dwellers for the sake of intelligibility. Over time this convergence materially affects the distinctiveness of a dialect in a situation of language contact (Gumperz 1969, p. 436)

Already in 1992 Palva illustrated the issue of the typological division into Bedouin and sedentary dialect types in Jordan basing his analysis on the vernaculars spoken in the cities of Salt and Karak: “both dialects are labelled as *bəgūl* dialects. [. . .] However, as a matter of fact (they) display many typically Bedouin features, so markedly different from Syro-Palestinian sedentary dialects” (Palva 1992, pp. 53–54).

In this study of the dialect of the Bani ^ʕAbbād the problem is the reverse, i.e., a historically Bedouin-type dialect that shows traits of sedentary dialects, and that thus exhibits important markers of both dialect types.

Due to the new *madani* established features it is not possible to label this dialect only as Bedouin, and according to the retention of some *šāwi* traits it cannot be defined as a sedentary dialect either.

The traditionally dichotomous split of Arab dialects in terms of Bedouin versus sedentary (Versteegh 1984; Rosenhouse 1984, 2006; Cadora 1992), is no longer sufficient to describe and classify those varieties like the Bani ^ʕAbbād dialect, i.e., languages which have experienced relatively high degrees of contact to the extent that change is additive (Trudgill 2010, p. 301).

As pointed out by Lentin (1994) and Watson (2011), the mere use of this bipolar division is inaccurate and gives rise to a number of difficulties especially when classifying dialects in a situation of language contact.

Due to the linguistic developments that occurred in the dialect of the Bani ^ʕAbbād it seemed functional to define it as a mixed type, in order to better underline its particular nature.

However, the emergence of mixed type dialects, such as this one, brings attention to the need to conceptualize new ways of grouping dialects and that further criteria of analysis need to be adopted, especially morphological and syntactic ones.

Funding: This research was partially funded by INALCO, with a grant destined to a fieldwork campaign in July and August 2016.

Institutional Review Board Statement: Ethical review and approval were waived for this study since it guarantees the anonymity of the participants. The study was conducted according to the guidelines of the Declaration of Helsinki. The interviewees and informants have always been regarded as partners who had a word in the (re)use of the material furnished by them. In order to respect all ethical demands usual in linguistic studies, they were always informed about the aims of the research, and before recording them, their consent was asked for publishing their speech in printed sources and online. Furthermore, the interview partners have never faced any problems with the Jordanian authorities because of the cooperation with the author. First, because most interviews and recordings were made by a local person who accompanied the author and, second, because research on spoken Arabic is not a sensitive topic in Jordan.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Only private dataset was analyzed for this study. It stems from the author’s personal fieldwork in Jordan in 2016 and 2017, and it is available on request from the corresponding author and with the permission of the informants.

Acknowledgments: I would like to thank Stephan Procházka, Veronika Ritt-Benmimoun, Francesca Bellino and Bettina Leitner for their valuable remarks and critical comments on the draft versions of this article.

Conflicts of Interest: The author declares no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Notes

- 1 Herin (2019, p. 96) already introduced those within the *yigūl* label (Southern Nomadic *yigūlu*, Central Nomadic *yigūlu*, and Northern Nomadic *yigūlūn*), differentiating between the Jordanian Bedouin varieties that display or not the final *-n* in the imperfective 3. m. pl. form of the verb to “to say”.
- 2 This corpus was originally transcribed in 2017 for my master dissertation *Le dialecte des Bani ʿAbbād Analyse des traits phonologiques, morphologiques et syntaxiques discriminants*.
- 3 Elicitations were used in some specific cases in order to obtain complete paradigms and inquire about specific vocabulary.
- 4 In order to respect all ethical demands usual in linguistic studies, all the present people were informed about my aims, and they accepted to be recorded for research purposes.
- 5 Only the interview recorded in ʿArāg al-ʿAmīr was carried out by Prof. B. Herin, my master supervisor, who at that time was also in Jordan.
- 6 The thesis was written under the supervision of Prof. B. Herin and defended at INALCO, Paris, in 2018.
- 7 This feature will be treated in detail in the section dedicated to the bound pronouns.
- 8 For this example, I refer to the records of three women, members of the same family: the illiterate grandmother of 90 years old, who has always lived in Yarga and who hasn’t had long contacts with Palestinians; the mother of 55 years old, who had access to academic education and who also lives in Yarga; the daughter of 23 years old, a student at The University of Jordan.
- 9 This number indicates the number of occurrences of the described phenomenon in the corpus.
- 10 Heating plate on which the Bedouins bake the bread.
- 11 It is worth mentioning that this genitive exponent is gender variable.
- 12 This form is attested in the speech of the old and conservative speakers in Fhēs as reported by Herin (2010, p. 373).
- 13 Seemingly used with or without article on the head-noun.
- 14 From the manuscript of his lecture for the workshop *Machtverhältnisse und Sprachkontakte in der Syrischen Steppe: Die Beziehungen zwischen Beduinen und Sesshaften im Spiegel der Dialekte*, FU Berlin, 27 June 2018.
- 15 Traditional dance performed by men during weddings.
- 16 In the corpus there are 15 occurrences of *mā fiš*, against 27 instances of *mā fi*.
- 17 This pseudo-verb can be used also to express deontic shade (Younes and Herin 2016, p. 11).
- 18 There are 32 instances of *bidd-* and 31 of *widd-*.
- 19 In this respect, one of the women interviewed told me that during her university years she was talking with other girls of her class, not Bedouin, and she used the expression *bi-s-sāʿ* ‘quickly’. They did not understand her and pointed out how weird was her way of speaking. Then she added “if you speak like we do to other people, they don’t understand, they laugh at you”.
- 20 With this term it is implied that major sedentary linguistic features penetrated in a historical Bedouin dialect.

References

- Al Tawil, Miriam. 2019. *La langue arabe parlée dans le Hōrān*. Master’s thesis, Università degli Studi di Catania, Catania, Italy.
- Al Tawil, Miriam. 2021. *Morpho-Syntactic Features of Bedouin Varieties in Northern Jordan*. Naples: Maydan, Working Paper in Publication.
- Behnstedt, Peter. 1997. *Sprachatlas von Syrien*. I: Kartenband, *Semítica Viva* 17. Wiesbaden: Harrassowitz.
- Bergsträsser, Gotthelf. 1915. *Sprachatlas von Syrien und Palästina*. *Zeitschrift des Deutschen Palästina-Vereins* 38: 169–222.
- Bettini, Lidia. 2006. *Contes Féminins de La Haute Jézireh Syrienne. Matériaux Ethno-Linguistiques d’un Parler Nomade Oriental*. Florence: Dipartimento di Linguistica.
- Blanc, Haim. 1970. *The Arabic Dialect of the Negev Bedouins*. Jerusalem: Academic Press.
- Brustad, Kristen. 2000. *The Syntax of Spoken Arabic: A Comparative Study of Moroccan, Egyptian, Syrian, and Kuwaiti Dialects*. Washington, DC: Georgetown University Press.
- Cadora, Frederic J. 1992. *Bedouin, Village and Urban Arabic: An Ecolinguistic Study*. Leiden: Brill.
- Cantineau, Jean. 1936. *Études Sur Quelques Parlers de Nomades Arabes d’Orient* 1. *Annales de l’Institut d’Études Orientales* 2: 1–118.
- Cantineau, Jean. 1937. *Études Sur Quelques Parlers de Nomades Arabes d’Orient* 2. *Annales de l’Institut d’Études Orientales* 3: 119–237.
- Cantineau, Jean. 1946. *Les Parlers Arabes du Hōrān*. Paris: Klincksieck.
- Cleveland, Ray L. 1963. A Classification of the Arabic Dialects of Jordan. *Bulletin of the American Schools of Oriental Research* 171: 56–63. [[CrossRef](#)]

- Creissels, Denis. 1995. *Éléments de Syntaxe Générale*. Paris: Presses Universitaires de France—PUF.
- Creissels, Denis. 2006a. *Syntaxe Générale, une Introduction Typologique: Tome 1, Catégories et Constructions*. Paris: Hermes Science Publications.
- Creissels, Denis. 2006b. *Syntaxe Générale, une Introduction Typologique: Tome 2, La Phrase*. Paris: HERMES SCIENCE.
- De Jong, Rudolf. 2000. *A Grammar of the Bedouin Dialects of the Northern Sinai Littoral: Bridging the Linguistic Gap between the Eastern and Western Arab World*. Leiden: Brill.
- François, Alexandre, and Maia Ponsonnet. 2013. Descriptive Linguistics. In *Theory in Social and Cultural Anthropology: An Encyclopedia*. Edited by Jon R. McGee and Richard L. Warms. London: SAGE, vol. 1, pp. 184–87. Available online: https://www.academia.edu/16430720/_Fran%C3%A7ois_and_Ponsonnet_Descriptive_linguistics_encyclopedia_entry_ (accessed on 12 June 2021).
- Giles, Howard, Donald M. Taylor, and Richard Bourhis. 1973. Towards a Theory of Interpersonal Accommodation through Language: Some Canadian Data. *Language in Society* 2: 177–92. [CrossRef]
- Gumperz, John. 1969. Communication in Multilingual Societies. In *Cognitive Anthropology*. Edited by S. Tyler. New York: Holt, Rinehart and Winston, pp. 435–49.
- Haspelmath, Martin, and Andrea D. Sims. 2010. *Understanding Morphology*, 2nd ed. Understanding Language Series; London: Hodder Education.
- Henkin, Roni. 2010. *Negev Arabic: Dialectal, Sociolinguistic, and Stylistic Variation*. Wiesbaden: Harrassowitz.
- Herin, Bruno. 2010. Le Parler Arabe de Salt, Jordanie: Phonologie, Morphologie et Éléments de Syntaxe. Ph.D. thesis, Université libre de Bruxelles, Bruxelles, Belgium. Available online: <http://hdl.handle.net/2013/> (accessed on 16 October 2020).
- Herin, Bruno. 2019. Traditional Dialects. In *The Routledge Handbook of Arabic Sociolinguistics*, 1st ed. Edited by Enam Al-Wer and Uri Horesh. Series: Routledge Language Handbooks: Routledge; New York: Routledge, pp. 93–105. [CrossRef]
- Herin, Bruno, Igor Younes, Enam Al-Wer, and Youssef Al-Sirour. 2022. The Classification of Bedouin Arabic: Insights from Northern Jordan. *Languages* 7: 1. [CrossRef]
- Holes, Clive. 1995. Community, Dialect and Urbanization in the Arabic-Speaking Middle East. *Bulletin of the School of Oriental and African Studies* 58: 270–87. [CrossRef]
- Isaksson, Bo. 1999. The Non-Standard First Person Singular Pronoun in the Modern Arabic Dialects. *Zeitschrift für arabische Linguistik* 37: 54–83.
- Lentin, Jérôme. 1994. Classification et Typologie Des Dialectes Du Bilād Al-Šām: Quelques Suggestions Pour Un Réexamen. *Matériaux Arabes et Sudarabiques (MAS-Gellas)* 6: 11–36.
- Giles, Howard, and Tania Ogay. 2007. Communication Accommodation Theory. In *Explaining Communication: Contemporary Theories and Exemplars*. Edited by Bryan B. Whaley and Wendy Samter. Available online: https://www.academia.edu/35366642/Howard_Giles_et_Tania_Ogay_2007_Communication_Accommodation_Theory (accessed on 10 December 2021).
- Oppenheim, Max Freiherr von. 1943. *Die Beduinen 2 Die Beduinenstämme in Palästina, Transjordanien, Sinai, Hedjäs*. Leipzig: Harrassowitz.
- Palva, Heikki. 1969. Balgāwi Arabic. 2. Texts in the dialect of the Ygūl- group. *Studia Orientalia (Helsinki)* 40: 1–15.
- Palva, Heikki. 1976. *Studies in the Arabic Dialect of the Semi-Nomadic ʿAl-ʿAğārma Tribe (al-Balqāʿ District, Jordan)*. Goeteburg: University of Goeteburg.
- Palva, Heikki. 1982. Patterns of Koineization in Modern Colloquial Arabic. *Acta Orientalia (Copenhagen)* 43: 13–32.
- Palva, Heikki. 1984. A General Classification for the Arabic Dialects Spoken in Palestine and Transjordan. *Studia Orientalia* 55: 357–76.
- Palva, Heikki. 1992. Typological Problems in the Classification of Jordanian Dialects: Bedouin or Sedentary? In *The Middle East Viewed from the North. Papers from the First Nordic Conference of Middle Eastern Studies, Uppsala, 26–27 January 1989*. Edited by Bo Utas and Knut S. Vikør. Bergen: Alma Mater Forlag, pp. 53–62.
- Palva, Heikki. 1994. Bedouin and Sedentary Elements in the Dialect of es-Salt. Diachronic notes on the sociolinguistic development. In *Actes des premières journées internationales de dialectologie arabe de Paris*. Edited by Dominique Caubet and Martine Vanhove. Paris: Inalco, pp. 459–69.
- Peake, Frederick. 1958. *A History of Jordan and Its Tribes*. Miami: University of Miami Press.
- Procházka, Stephan. 2018. Manuscript of his lecture for the workshop. In *Machtverhältnisse und Sprachkontakte in der Syrischen Steppe: Die Beziehungen zwischen Beduinen und Sesshaften im Spiegel der Dialekte*. Berlin: FU Berlin, June 27.
- Romaine, Suzanne. 2010. Contact and Language Death. In *The Handbook of Language Contact*. Hoboken and Chichester: John Wiley & Sons, Ltd., pp. 320–39. [CrossRef]
- Rosenhouse, Judith. 1984. *The Bedouin Arabic Dialects. General Problems and a Close Analysis of North Israel Bedouin Dialects*. Wiesbaden: Harrassowitz.
- Rosenhouse, Judith. 2006. Bedouin Arabic. In *Encyclopedia of Arabic Language and Linguistics*. Edited by Kees Versteegh. Leiden: Brill, pp. 259–69.
- Sakarna, Ahmad Khalaf. 1999. Phonological Aspects of 9abady Arabic, a Bedouin Jordanian Dialect. Ph.D. thesis, University of Wisconsin, Madison, WI, USA.
- Sawaie, Mohammed. 2011. Jordan. In *Encyclopaedia of Arabic Language and Linguistics*. Edited by Lutz Edzard and K. De Jong. Leiden: Brill.
- Shryock, Andrew. 1997. *Nationalism and the Genealogical Imagination: Oral History and Textual Authority in Tribal Jordan*, 1st ed. Berkeley: University of California Press.
- Torzullo, Antonella. 2018. Le Dialecte des Bani ʿAbbād: Analyse des Traits Phonologiques, Morphologiques et Syntactiques Discriminants. Master’s thesis, INALCO, Paris, France.

- Trudgill, Peter. 1999. New-Dialect Formation and Dedialectalization: Embryonic and Vestigial Variants. *Journal of English Linguistics* 27: 319–27. [[CrossRef](#)]
- Trudgill, Peter. 2010. Contact and Sociolinguistic Typology. In *The Handbook of Language Contact*. Hoboken and Chichester: John Wiley & Sons, Ltd., pp. 299–319. [[CrossRef](#)]
- Versteegh, Kees. 1984. *Pidginization and Creolization: The Case of Arabic*. Current Issues in Linguistic Theory, 33. Amsterdam: Benjamins.
- Watson, Janet C. E. 2011. Arabic Dialects (General Article). In *The Semitic Languages: An International Handbook*. Edited by Stefan Weninger. Berlin and Boston: De Gruyter Mouton, pp. 851–96.
- Younes, Igor, and Bruno Herin. 2013. Un Parler Bédouin Du Liban Note Sur Le Dialecte Des ‘Atīġ (Wādī Xālid). *Zeitschrift Für Arabische Linguistik* 58: 32–65.
- Younes, Igor, and Bruno Herin. 2016. Šāwi Arabic. In *Encyclopedia of Arabic Language and Linguistics*, online ed. Leiden: Brill.
- Younes, Igor. 2014. Notes Prélminaires Sur Le Parler Bédouin Des Abu ‘Īd (Vallée de La Békaa). *Romano-Arabica* 14: 355–87.
- Younes, Igor. 2017. Dialect Contact in the Beqaa Valley. *Romano-Arabica* 17: 131–40.

Article

The Classification of Bedouin Arabic: Insights from Northern Jordan

Bruno Herin ^{1,*}, Igor Younes ², Enam Al-Wer ³ and Youssef Al-Sirour ⁴¹ Etudes Arabes: INALCO, 75013 Paris, France² Independent Researcher, 1000 Brussels, Belgium; igoryounes@gmail.com³ Department of Language and Linguistics, University of Essex, Colchester CO4 3SQ, UK; enama@essex.ac.uk⁴ Independent Researcher, Umm El-Jimal 25488, Jordan; yousefalsirour@gmail.com

* Correspondence: bruno.herin@inalco.fr

Abstract: The goal of the present paper is to provide a reevaluation of the classification of the Bedouin dialects of Northern Arabia and the Southern Levant, based on published or publicly available data and on first-hand data recently collected amongst some Bedouin tribes in Northern Jordan. We suggest extending previous classifications that identify three types of dialects, namely A (*nizi*), B (*šammari*), and C (*šāwi*). Although intermediary or mixed types combining *šammari* features with *šāwi* features were already noted, our data suggest that further combinations are possible, either because they had so far been unnoticed or because recent levelling and dialect mixing have blurred the boundaries between some of the varieties.

Keywords: Arabic dialectology; classification; Bedouin Arabic; Jordan; Masāʿid

Citation: Herin, Bruno, Igor Younes, Enam Al-Wer, and Youssef Al-Sirour. 2022. The Classification of Bedouin Arabic: Insights from Northern Jordan. *Languages* 7: 1. <https://doi.org/10.3390/languages7010001>

Academic Editors: Simone Bettega and Roberta Morano

Received: 9 August 2021

Accepted: 28 October 2021

Published: 23 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The goal of the present paper is to provide a reevaluation of the classification of the Bedouin dialects of Northern Arabia and the Southern Levant, based on published or publicly available data and on first-hand data recently collected by the authors amongst some Bedouin tribes in Northern Jordan. We suggest extending Cantineau's (1936, 1937) classification that identifies three types: A (*nizi*), B (*šammari*), and C (*šāwi*). Although Cantineau already noted intermediary or mixed types combining *šammari* features with *šāwi* features, our data suggest that further combinations are possible, either because they have so far not been noticed or because recent levelling and dialect mixing have blurred the boundaries between some of the varieties. Foundational surveys include Cleveland (1963) who, much in the same way as Blanc (1964) coined the *gilit*–*qāltu* dichotomy, coined the dialectonyms *biqūl*, *bikūl*, *bigūl*, *bi'ūl* and *yigūl* based on the 3.m.sg. of the imperfective of the verb **qāl* 'he said'. Further developments can be found in Palva (1984). Palva divides the Bedouin dialects of the Southern Levant into four groups, as below:

- The dialects of the Negev Bedouins.
- The dialects of the Arabia Petraea Bedouins such as the *Ḥiwēṭāt*.
- The dialects of the Syro-Mesopotamian sheep-rearing tribes, which corresponds to the *šāwi* type (Cantineau's type C, Younes and Herin 2016).
- The dialects of the North Arabian Bedouins (Cantineau's types A and B).

The problem with the *biqūl*–*yigūl* appellation is that it fails to capture the difference between a major split in Jordan, namely between dialects that exhibit final /n/ in the imperfective endings *-in* and *-in* and those which exhibit *-ī* and *-ū* (Herin 2019). Using the 3.m.pl. of the imperfective of *qāl* would partially solve this problem, which, combined with geography, yields the following classification: Southern *ygūlu*, Central *ygūlu*, and Northern *ygūlūn*. Central *ygūlu* is in many ways identical to the Northern *ygūlūn šāwi C*; the presence or absence of /n/ is the main difference. Only Southern *ygūlu* is an extension of the North-West Arabian type (Palva 2011). Our focus will be the hitherto under-studied

Northern *ygūlūn* type with a special focus on the *Misāʿīd* dialect which exhibits many *šammari* features such as the apophonic passive (*yīdkar* ‘it is remembered’) or a [dʰ] reflex of */ǧ/ (*dʰībal* ‘mountain’), but also *šāwi*-like traits such as the [q] < /ǧ/ (*qēr* ‘other’) and more surprisingly, features that are reminiscent of North-West Arabian such as the resyllabification of *^hinC₁aC₂aC₃a . . . into ^hinC₁C₂ʔC₃a (*in^hḥkumat* ‘it was ruled’). Consequently, the major taxonomies have to be combined to represent the overall picture more accurately. Additionally, sociolinguistic developments which have affected the classification of these dialects, such as dialect contact and koineization, need to be incorporated.

The data on which this paper draws were collected amongst members of the *Misāʿīd* tribe in 2019 in the municipality of *Umm al-Ġimāl* in Northern Jordan, twenty kilometres East of Mafraq. With the help of Youssef Al-Sirour, a permanent resident of *Umm al-Ġimāl* and an immediate member of the community under investigation, we visited local families and recorded two casual conversations. Because of the limited nature of the corpus, the present discussion should be considered provisional until more data are collected. We will first sum up Cantineau’s classification followed by those put forward by Cleveland and Palva. Based on our own observations, we suggest essential amendments to these classifications. We then present the salient features of the dialect, followed by a small sample taken from the recordings. The last part deals with the classification of the present dialect in the light of previous literature. We also highlight some methodological issues regarding data collection, levelling, and short-term accommodation.

2. Cantineau’s Classification

The first scholar to draw a comprehensive classification of the Bedouin dialects of Northern Arabia is Cantineau (1936, 1937). The first distinction relates to the occupational profile of the Bedouins located in this area, whom Cantineau called “*grands nomades*” (‘great nomads’) as opposed to “*petits nomades*” (‘little nomads’). The former designates tribes which mostly rely, at least historically, on camel rearing, and the latter designates tribes which were mostly active in sheep rearing. This bipartite separation was further divided into three broad groups to which he attributed the letters A, B, and C. The A-group designates camel-rearers from the *Niza* confederation. The B-group refers to camel-rearers from the *Sāmmar* confederation, whereas the C-group refers to the sheep-rearing tribes of the Syro-Mesopotamian *bādyā* ‘steppe’. More marginally, Cantineau also talks about three smaller subgroups, the variety of *ar-Rass* in the *Ġašīm* region in the central-northern part of Saudi Arabia, the dialect of *al-Ġōf* located in the far north of Saudi Arabia, and finally the dialects of the oasis of the Syrian desert of *al-Qarītēn*, Palmyra and *Suxne*.

Some features of the A-group (*Niza*) include the affricate [ts] and [dʒ] of etymological /k/ and /g/ (Standard Arabic /q/) in the vicinity of front vowels: *čalbatī* ‘my she-dog’ (< *kalbatī*), *ġiddam* ‘front’ (< *giddām*). Etymological /ǧ/ can be realized [gʲ], [dʲ], and [dʒ]: *did^hād^ha* ‘hen’ (< *daǧāǧa* ‘hen’). The feminine ending -a exhibits no raising except in the vicinity of /i/, /ī/, or /j/ in which case it raises towards [æ]: *lahyā* ‘beard’ (< *lihyā*). Etymological diphthongs /aw/ and /ay/ are not monophthongised although the distance between the two elements is reduced, yielding, respectively and approximately, [o^w] and [eʲ]: *ǧowz* ‘nut’ and *beyt* ‘tent’. An important feature is the so-called *gahawa* syndrome, understood as the insertion of an anaptyctic /a/ vowel between /ǧ/, /x/, /ħ/, /h/, or /ʕ/ and a following consonant of the type Ø → /a/ / aX_C in which X is one of the aforementioned consonants and C is different from X: *ǧ ahr* → *ǧahar* ‘back’. In addition to this, *C₁aC₂aC₃v sequences are resyllabified into C₁C₂ʔC₃v: *xšība* < *xašāba* ‘piece of wood’. The *gahawa* syndrome is also active in the passive participle template *maC₁C₂ūC₃, in which case it also combines with the resyllabification rule: *maḥtūt* → *maḥatūt* → *mḥatūt* ‘put’. Another important distinction introduced by Cantineau is trochaism vs. atrochaism. While these terms refer to a type of meter in Classical Greek poetry, his use of this parameter entails a particular syllabic type. Accordingly, Cantineau separates trochaic from atrochaic varieties. Trochaic varieties have the tendency to favour sequences of Cv/Cv̄ syllables. CvC syllables are tolerated in final position or if followed by Cv or a final CvC/Cv̄C: *ihāšadan*

'they (f.) harvest', *yākalan* 'they (f.) eat', *rāsa-na* 'our head', *nāgat-i* 'my she-camel'. Atrochaic dialects do not restrict sequences of CvC syllables: *ihāšdan* 'they (f.) harvest', *yākalan* ~ *yāčlan* 'they (f.) eat' *rās-na* 'our head', *nāgt-i* 'my she-camel'. The A group is strongly trochaic.

As far as morphology is concerned, these dialects feature the nominal suffix *-n* commonly called 'nunation' in Semitic studies, which essentially marks nouns denoting indefinite specific referents when they are complex NPs consisting of a nominal head and a modifier (Holes 2004). Another salient feature is the pronominal indexes which feature a final /n/ in the prefix conjugation: *t(v)gūlīn* 'you (f.) say', *t(v)gūlūn* 'you (m.pl.) say' and *y(v)gūlūn* 'they (m.pl.) say'. As far as bound pronouns are concerned, a noticeable trait is the allomorph *-ah* of the 3.f.sg. after a final weak root consonant: '*aly-ah* 'on her' and *abw-ah* 'her father'. The 2.m.sg. and 2.f.sg. in those dialects surface as *-k* and *-é* after words ending in a short vowel: *farās-k* 'your (m.) horse'. The 2.m.pl. and 2.f.pl. forms are *-kam* and *-kin* and the 3.m.pl. and 3.f.pl. are *-ham* and *-hin*. Specific independent forms of free pronouns include 1.sg. *āna* and 1.pl. *hīnna*. Another salient feature is the forms of the verbs *axaḏ* 'he took' and *akal* 'he ate', instead of *kala*, *xada*.

As far as group B (*šammari*) is concerned, much of the phonology and morphology is shared with group A. Differences arise in the following features. As noted by Cantineau (1937, p. 130), "*l'imāla de la terminaison féminine est nette et forte, a un tel point qu'elle semble résister au tafxīm d'une consonne précédente*": *gargūre* 'she-lamb', *nāge* 'she-camel'. These dialects are also characterised by the lenition of the feminine plural ending *-āt* in pause in which case it reduces to *-ā'*: *xams ḏbšalā'* 'five onions'. Concerning bound pronouns, *šammari* dialects exhibit *-ak* and *-ic* in the 2.m.sg. and 2.f.sg. with any vowel syncope. In addition to this, the 1.sg. allomorph *-an* surfaces in all positions: *ḡrub-an* 'he hit me' (< *ḡarab-an* → *ḡarāb-an* → *ḡrub-an*). Cantineau also notes the allomorph *-(w)o* after final long *-ā*: *ḡadā-o* 'his lunch'. Our data suggest that this allomorph is selected after any long vowel, whether plain or monophthongised.

Group C dialects, also known as *šāwi* dialects, are spoken by the sheep-rearing tribes of the Syro-Mesopotamian *bādyā* 'steppe' and its fringes. Distinct features include the affricates [tʃ] and [dʒ] as reflexes of /k/ and /g/ in front vowel environments. The reflex of etymological /ǧ/ is always the affricate [dʒ]. A slight raising towards [æ] of final *-a* and *-ā* is heard in non-back and non-velarised contexts: *šinūnā* 'butter milk', *ihnā* 'we'. In terms of phonotactics, *maC₁C₂ūC₃ stems are not susceptible to the *gahawa* syndrome and hence, there is no resyllabification. *Šāwi* dialects are also atrochaic, in that sequences of CvC syllables are not restricted: *yihārban* 'they (f.) escape', *yākalan* 'they (f.) eat'. Specific morphological forms are 1.sg. *āni* 'I' and *ihnā* 'we' for free pronouns and the pairs *-kum/-ēdn* and *-hum/-hḏn*.

3. Cleveland's Classification of the Dialects of Transjordan

Cleveland (1963) is an attempt to classify the dialects spoken in Jordan and Palestine, both sedentary and Bedouin. Cleveland coined new terms using the 3rd person singular of the verb *qāl* 'he said' in the imperfective in order to designate the different dialectal groups. His first cluster, which he calls *yigūl*, refers to all the Bedouin varieties which lack the *b-* prefix of the imperfective. The second group he distinguishes is *bigūl*, by which he refers to the sedentary populations of Jordan, including some locations on the west bank of the Jordan river. His third group is the *bikūl* type, which is characteristic of the sedentary rural populations of central Palestine. Lastly, the *bi'ūl* group incorporates the sedentary urban populations of Palestine, including those which settled more recently in Jordan. Cleveland does not mention a *biqūl* group which would include the Druze dialect of Azraq, Northern Jordan. This dialect is as yet undocumented but research in this community is ongoing and the findings will be published in due course.¹ As we will see below, Cleveland's classification does not capture important differences found amongst the Bedouins. It also fails to capture the divergences amongst the indigenous sedentary dialects of Jordan, which, although all belong to the *bigūl* group, exhibit a sharp division between a southern *mu'ābi* type and a northern-central *balgāwi-hōrāni* type.

4. Palva's Classification

Palva (1984) delves deeper into Cleveland's classification using a larger pool of variables. Palva mentions the urban Palestinian dialects, which correspond to Cleveland's *bi'ūl*. As far as rural dialects are concerned, he distinguishes between Galilean dialects (*biqūl*), central Palestinian dialects (*bikūl*), south Palestinian dialects (*bigūl*), north and central Transjordanian dialects (*bigūl*), and south Transjordanian dialects (*bigūl*). His classification of the Bedouin dialects includes those of the Negev Bedouins (*bigūl*), the dialects of southern Jordan (*yigūl*), the dialects of the Syro-Mesopotamian sheep-rearing tribes (*yigūl*), and lastly the dialects of the North Arabian Bedouins (*yigūl*). Palva's classification distinguishes well between all the subgroups of the sedentary types but lumps together sub-divisions within the Bedouin type that ought to be differentiated. In the dialects of the Syro-Mesopotamian sheep-rearing tribes, no distinction is made between the dialects of the Jordan valley and the *šāwi* type. As regards the dialects of the North Arabian Bedouins, no further distinction is made between Cantineau's A and B groups.

5. Addenda to Cantineau, Cleveland, and Palva

5.1. Younes' Subgrouping of Ca

So far, only tribes which had *šāwi* type dialects had been located and for some of them investigated, thus belonging to Cantineau's C group. These are for example the *N'ēm*, *Lhēb*, and *Bani 'Azz* who, in Lebanon, are mainly located in the Northern and Eastern parts of the country. The dialects spoken by these tribes are all unmistakably of the *šāwi* type, exhibiting features such as the /č/ and /ğ/ reflex of etymological /k/ and /g/, a first or second degree raising of final *-a* and *-ā* to [æ] or [ɛ], atrochaism, absence of the *gahawa* syndrome on the *maC₁C₂ūC₃ template, the pseudo-verb *wōdd* 'want', and the lexeme *θl̥dm* for 'mouth'. In recent fieldwork carried out in the central part of the Bekaa valley by one of the authors of the present study, two new Bedouin tribes were investigated: the *Abu 'Id* and the *'Idīn*. Their presence in that part of the country had been, until then, unnoticed. Indeed, the presence of *Ḥsina* clans, who are a big sub-section of the *'Niza* confederation and to whom the *Abu 'Id* and the *'Idīn* are connected, was already attested in Syria. The *Ḥsina* are to the *'Niza* what the *Ṭayy* are to the *Šammar* in that they are the first clans who migrated northwards into the Syro-Mesopotamian steppe around a millennium ago. This resulted in a prolonged contact with Bedouin tribes who had migrated earlier into the area such as the *Muwāli*, *Ḥadīdīn*, and *N'ēm*—who had dominated the Syro-Mesopotamian steppe. The linguistic outcome of this prolonged contact was convergence towards the *šāwi* type. After investigation, it turned out that the dialect of the *Abu 'Id* and the *'Idīn* exhibited a similar profile, with core *šāwi* features alongside with *'nizi* features. For instance, these dialects exhibit no raising of *-a* and *-ā*, *gahawa* active in the *maC₁C₂ūC₃ template, the verb *yibi* 'he wants', and a more pervasive use of nunation. This state of affairs led us to coin a new term for this type of configuration, using Cantineau's terminology. Consequently, it seemed opportune to use the combination of Ca letters to designate this type of dialects: upper case C for the *šāwi* component and lower case a for the *'nizi* component. Cantineau (1937) already used such a combination of letters for the varieties spoken in the *Gašīm* area in modern-day Saudi Arabia that combine predominantly *šammari* features alongside with *'nizi* features: Ba.

5.2. Herin's *ygūlu* vs. *ygūlūn*

As noted in Herin (2020), one of the shortcomings of Cleveland's *yigūl* type is that it lumps together three sub-types within the Bedouin dialects of Jordan: the dialects of the Jordan valley Bedouins such as the *'Ağārma*, *'Adwān*, and *'Abābīd*, the dialects of Bedouins of northern Jordan such as the *Bani Ṣaxar*, *Sardiyye*, *Sirḥān*, *Al 'Isa*, and *Misā'id*, and finally the Bedouin varieties of Southern Jordan such as the *Ḥwēṭāt*, *Bdūl* and *Zawāyda*. The Jordan valley type differs from Cantineau's C group in that they lack the final /n/ in the imperfective endings *-īn* and *-ūn*, also found in the dialects of the Bedouins of northern Jordan. It appears that it would be more conclusive to use the 3.m.pl. inflexion of the

imperfective of the verb *gāl* to capture some of these differences. The following general classification would arise:

- (I) Sedentary *bigūlu*,
- (II) Southern Nomadic *ygūlu*,
- (III) Central Nomadic *ygūlu*, and
- (IV) Northern Nomadic *ygūlūn*.

6. Features of the *Misā'id* Dialect

In 2019, Bruno Herin, Enam Al-Wer, and Youssef Al-Sirour began fieldwork amongst the *Misā'id* tribe in Umm al-Ġimāl, Northern Jordan. The fieldwork was facilitated by Yūsif, who is a member of the tribe, as noted above. In this exploratory phase of the research, we recorded two forty-minute sessions consisting of casual conversations and narratives. These recordings were subsequently transcribed and analysed. In the remainder of this article, we present our analysis of the salient features of this dialect based on these recordings.

6.1. Phonology

The phonetics of the feminine ending was mostly recorded as the unraised reflex [a]: *šidāda* 'severe, extreme', *šāša* 'piece of fabric/muslin', *mayya* 'water', *waḥda* 'one (f.)'. A first degree raising was recorded in *sāknā* 'dwelling (f.)', *ʿašīrā* 'clan', *ʿuṭmāniyyā* 'Ottoman', *lahd^hā* 'speech, accent'. A second degree raising was also recorded in a handful of items such as *zǧīre* 'small' and *kṭīre* 'much (f.)', and also after an emphatic sound as in *miḥmāše* 'coffee bean roasting pan'. The unraised reflex [a] is typical of 'nizi type (in the Syro-Mesopotamian steppes) whereas the first-degree reflex is equally found in the *šāwi* varieties as in the 'nizi dialects, although it is contextually conditioned (e.g., in front contexts). The second-degree raising found in some items most likely represents short-term accommodation, induced by the presence of speakers of other Jordanian dialects.² It may also be indicative of the course of future developments in the dialect, viz. convergence to koineised Jordanian varieties, especially since the younger members of the tribe have frequent face-to-face contact with speakers of other Jordanian dialects through formal education and in the workplace. The raising heard in *miḥmāše* after a velarized consonant on the other hand, is typical of the *šammari* type. Despite some degree of variation in the realization of the feminine ending in our data, the distribution found amongst the informants overall is consistent with the 'nizi type.

In pause, a slight aspiration occurs after the feminine ending: *ʿašīrā^h#* 'clan', *gībīlā^h#* 'tribe'. This feature is found in both the A 'nizi and B *šammari* groups.

The etymological diphthongs /aw/ and /ay/ are both monophthongised to /ō/ and /ē/, respectively: *fōg* 'above', *yōm* 'day', *hōl* 'around', *dōr* 'turn/point in time', and *bēt* 'tent', *ṭnēn* 'two', *xēl* 'horses'. Diphthongised realisations occurred in *Zbeyd* (tribal patronym), *xēyš* 'jute'. These reflexes are common in the group C *šāwi* dialects. Groups A and B usually have more consistent slight diphthongised reflexes.

As far as the affrication of etymological /k/ and /g/ is concerned, the recorded reflexes all pattern respectively with the *šāwi* type /č/ and /Ġ/: *hīč* 'so', *čimā* 'desert truffle', *čīṭīr* 'much'. Only one instance of /ǧ/ < /g/ was recorded in *ṭīǧ* 'endure'. Other items which were expected to be realised with /Ġ/ were recorded with /g/: *šarg* 'east', *giddām* 'in front'. This, in all likelihood, is a short-term accommodation phenomenon induced by the presence of speakers of standard Jordanian. The same observation can be made about non-affricated reflexes of /k/ in items such as *kān* 'he was', *kiṭīr* 'much' (also recorded with /č/, see above), and *kibīr* 'big' all of which are normally affricated in the vernacular.

Etymological /ǧ/ was recorded /d^y/ in *d^yibal* 'mountain', *d^yaw* 'they (m.) came', and *id^yiban* 'they (f.) brought'. The affricate /Ġ/ was also recorded: *yīǧūn* 'they (m.) come', *ǧawwa* 'inside', *ǧīld* 'skin'. The /d^y/ reflex is common in groups A and B whereas the affricate /ǧ/ is a hallmark of the *šāwi* type. The indigenous reflex is undoubtedly /d^y/. Although a short-term accommodation effect cannot be ruled out, the presence of /ǧ/

could also be due to earlier change within the dialect, as noted by Cantineau in some camel-breeder varieties.

An interesting and somehow unexpected feature that was occasionally recorded is the the *qalqala*, understood to be the uvular realisation of etymological /ğ/: *qēr* ‘other’ (<ğēr), *qāli* ‘expensive’ (<ğāli), *muqsil* ‘washing area’ (<mağsil). To the best of our knowledge, this phenomenon is a hallmark of the Mesopotamian *šāwi* dialects.

Final /t/ in the plural feminine ending *-āt* interestingly drops in pause: *ğuza*# ‘raids’, *šağla*# ‘things’, *Rda’iyya*# (toponym), *halāla*# ‘livestock heads’. This feature, as mentioned above, was already noted as commonly occurring in the B and Bc dialects.

The laryngeal stop /ʔ/ was recorded once as pharyngeal /ʕ/ in *sa’alt* ‘I asked’, which is a salient feature of North-West Arabian. In addition to this, /ʕ/ is often glottalised in pause: *hassā*# [hassa:ʕʔ] ‘now’, *mānī*# [ma:niʕʔ] ‘hindrance’, *bē*# [be:ʕʔ] ‘sale’.

Expectedly, *C₁aC₂aC₃v sequences are resyllabified into C₁C₂vC₃v: *skānaw* (<*sakanaw*) ‘they settled’, *Šrufāt* (tribal patronym < *Šarafāt*). Our corpus also attests the presence of resyllabification in derived templates such as form VII *inC₁aC₂aC₃: *ʾinḏḥkūmat* ‘it was ruled’ (*inḥakamat* → *inḥkamat* → *inḥkūmat* → *inḏḥkūmat*).

As far as the *gahawa* syndrome is concerned, it appears to be present in the dialect. Examples are *nḥāšid* ‘we harvest’ (here combined resyllabification *nāḥšid* → *nāḥšid* → *naḥšid* → *nḥāšid*), *ba’ad* ‘after’. Our data do not attest the presence of the *gahawa* syndrome in *taC₁C₂īC₃ and *maC₁C₂ūC₃ templates, which would suggest that it patterns in this respect with the *šāwi* type. Further data are needed to firmly confirm this observation.

As expected, the article receives primary stress as is normally the case in all of the Bedouin varieties of the area. To the best of our knowledge, only monosyllabic words of the type C₁vC₃ and disyllabic words of the type C₁vC₂v(C₃) can trigger the stress of the definite article. Attested instances in our data are: *ʾāl-muṭar* ‘the rain’, *ʾān-nifal* ‘the clover’, *ʾāl-ʿarab* ‘the Bedouins’. In addition to this and quite unexpectedly, we also encountered a stressed article with a C₁vC₂C₃v word in *ʾaṣ-ṣaḥra* ‘the desert’. Further data are needed to confirm whether stress assignment on the article is licenced in other words of this type and also possibly in other templates, which, as far as we know, would be a novelty.

An unexpected stress-related feature we found in the data is the second syllable stress in the plurals of C₁vC₂vC₃ type as in *nigāt* ‘points’ which also surfaced as *ngat* after high vowel elision in unstressed position. This is a feature found in North-West Arabian (Palva 2011).

6.2. Morphology

In the realm of verbal morphology, it appears that both the allomorphs *-aw* and *-am* in the 3.m.pl in the perfective are found: *winn-o gta’am kassaram min-ʿind giddām al-ḡamal* ‘and there they had cut and broken into pieces (the engravings) in front of the camel’. The *-aw* allomorph was recorded in the following: *ḥḏmaw ba’aḏ-ham ʿāšaw u-tikāṭaraw u-lamma tikāṭaraw, dʷaw ʔḏbitaw hānā* ‘they protected each other, lived and multiplied and when they multiplied they came and settled here’.³ These examples suggest that *-aw* and *-am* allophones are not in complementary distribution, unlike in some *šāwi* tribes along the Middle-Euphrates where one of the allomorphs is used exclusively in pause.

Person prefixes in the imperfective were often recorded with /a/ vowel: *yaṭla* ‘he goes out’, *takbar* ‘it gets bigger’, *yamši* ‘he walks’, *talga* ‘you find’. This is a typical camel-rearing trait not found in the *šāwi* dialects.

Initial glottal stop verbs such as *akal* and *axaḏ* behave similarly to what is found in the B, Bc, and C groups: *kalēt-o* ‘I ate it’, unlike *nizi*-type dialects which have *akalt* and *axaḏt* ‘I ate/have eaten’, ‘I took/have taken’.

As far as derived forms are concerned, the causative Form IV template *aC₁C₂aC₃-yiC₁C₂iC₃ is well attested in our data: *nḏṭḏl-o w-unḏnḏḏl-o* ‘we take it out and dust it’, *yumṭar* ‘it rains’, *yiwšil* ‘he brings’. The presence of this feature is not diagnostic of any sub-group but in the context of dialect contact and levelling, it is a noticeable feature. The imperfective of Form V *taC₁aC₂C₂aC₃ was recorded as *ytiC₁aC₂C₂aC₃* as in *ytidarrab*

'he trains'. Given that *šāwi* dialects are known for having $yiC_1aC_2C_2aC_3$ (*yidarrab*), the presence of this form is another indication of the camel-rearer background of the present dialect. This, in all likelihood, should also happen in form VI * $taC_1āC_2aC_3$ but our data lack instances of any verb of this type.

Another typical camel-rearer feature that is found in our data is what is referred to as the apophonic passive, known to be lost in the *šāwi* varieties. Only two instances were recorded: *yidkar* 'it is remembered' and *timadd* 'it is presented'. The template in the imperfective $yiC_1C_2aC_3$ in which the /i/ vowel contrasts with the /a/ vowel was noted above as a marker of the active forms. Further data are needed to assess the productivity of the apophonic passive in the modern-day form of the dialect.

The pronominal morphology of the dialect appears to be mixed. We recorded the first person free forms *ana* and *iḥna*, which are found in the C-*šāwi* group. Inversely, the bound plural forms *-kam* and *-ham* were found, which are camel-rearer forms. In the feminine plural, only the third person *-hin* is recorded in the data, but no second person. The first person singular bound pronoun surfaced as *-an* after a consonant: *wḥīat-an* 'it hurt me', *tūd^ya'-an* 'it hurts me'. This *-an* form is typical of the B and Bc groups. In the same vein, we recorded the form *-wo* after long vowels, which are also found amongst the B and Bc groups: *alē-wo* 'on him', *ḥinxallī-wo* 'we let him', *šifnā-wo* 'we saw him'. Moreover, an *-ah* allomorph in the 3rd person feminine singular was recorded after final /w/ and /y/ stems: *aly-ah* 'on her', *abw-ah* 'her father', which patterns with both the A and B camel-rearer dialects. After consonants, initial consonant bound pronouns all have initial vowel allomorphs: *bilād-a-na* 'our country', *kill-a-ham* 'all of them'. This, of course, is reminiscent of the trochaic syllable type of the dialect and a distinctive feature of all the A and B camel-rearer varieties.

7. Dialect Sample

We present here a sample of the recordings to enable the reader to capture the nature of the dialect. Because much of the sessions consisted of group conversations in which turns were for the most part quick and uncontrolled, it was difficult to isolate long stretches of monologue. Another problem that quickly surfaced was the presence of several instances of mixed forms, which are due to dialect mixing and perhaps ongoing changes in the dialect itself. As explained earlier, the session involved participants with different dialect backgrounds, which as we quickly realised, prompted the informants to accommodate towards other Jordanian dialects. Nevertheless, the two short excerpts exhibit salient features that can be safely attributed to the local form of speech of the *Misā'id* tribe.

Speaker 1: Bū Ṣāliḥ:

'al-Misā'id ham 'akbar 'ašrā w-al-'ašāyir hādōl d^vivār-na 'aširtēn dōl... kull al-'ašāyir hādī ḥōl ba'ad-ha hān sāknā b-al-manṭaḡa hāy d^virān. 'u-sābigan gabəl an-nās kānāt k^təḡzi^k 'ala ba'ad-ha sābigan gabəl-ma nḥkūmat ha-l-ḥlād ya 'ni [...] 'ala dōr al-'utnāniyyā yimkin ḥakm al-ḥlād hādī kānt an-nās ḥṭōma ba'ad-ha b-al-guwwa. ya 'ni yḡazu ba'ad-ham u-hādōl.. ḥasb guww^t al-'ašrā lli giddāma-ham [...] mā-ni wa 'i kitīr 'ana 'umr-i yimkin 'aktar min-saba'in sinā, ḥass mā sma'ət min-ha-l-ḡdām gabəl. ḡāl^w al-Misā'id mā 'umra-ham 'inno xaḡaw, 'illi yfukkūn ḥāla-ham b-əl-ḡza⁷, yim^dvūn min-ḡazu k^yifukko^k ḥāla-ham, dāyman maṣṣūrīn sibḥānallāh.

The *Misā'id* are the biggest tribe and the other tribes are our neighbors, the two other tribes ... All these tribes live next to each other here in the region, they are neighbors. In the past, people used to raid each other, before the region was under control [...] I think in the days the Ottomans controlled this region, people used to protect themselves in a warlike manner. I mean they used to raid each other and these ... It depends on the strength of the tribe which is facing them [...] I don't remember well, I am maybe older than seventy, it comes from what I have heard before from the elders. They said that the *Misā'id* never took, those who emerge during raids, they get out of the raid they emerge, always victorious God bless.

Speaker 2: Umm Ṣāliḥ:

ʾaxabbr-o bass ʾana mā daggēt la walla ʾi šufət walla šuft hān daggēt ʾa-l-ūd-i wĠiʾat-an u-daggēt ʾalē-(h) [. . .] bass yaḥḥa ad-damm xalaṣ yarbuṭan ʾaly-aḥ ʾādi yidall yōmēn mā tġīm-o winn-a waḥra . . . bass ḏn-nās mā ʾārif inno ḥarām gabḏl . . . ʾa-l-basāṭa ʾi walla ʾa-l-basāṭa, zīmā w-a-lāĠ ʾi lāĠ ḍarba ʾi ḍarba ḍarba gabḏl l-wāḥad lama yūĠa-o katf-o katf-o kʾyiduggoḵ ʾalē-wo ybaṭṭal yūĠa-o ʾlāĠ yaʾni [. . .] walla madri šift walla nās wāĠid ṭalāt ḏngāt ʾi billa la walla mā marrat ʾalay-yā mā ḏikart-ā . . . šuft niswān bī-hin ṭalāt nigāt ḏkbār ʾaĠāyiz ʾi . . .

I will tell him but I didn't get tattooed, by God I saw, here, I tattooed my hand [and] it hurt, I tattooed it [. . .] when the blood comes out, it's finished, they (f.) tied it normally for two days until it turns into a bruise . . . But people before didn't know it was *ḥarām* . . . Because of simpleness, by God, because of simpleness, beauty, and remedy, yes, remedy, a blow, a blow, before, when someone had a sore shoulder, they would tattoo it and the pain would stop, I mean [it's a] remedy [. . .] by God I don't know, I saw a lot of people with three dots [tattooed], yes, by God, this did not happen to me, I can't remember it . . . I saw women with three dots [tattooed], old women yes . . .

8. Discussion and Conclusions

Below (Table 1) is an overview of all the features discussed above and their distribution in the relevant dialectal groups. As mentioned earlier, Cantineau attributed a letter-code to the different groups he investigated. The two-way division is between the camel-rearer type which sub-divides into A (*Niza*) and B (*Šammar*) and sheep-rearer C group (*šāwi*). In accordance with this classification, we decided to allocate the letter D to the North-West-Arabian type. From the Table 1 below, it quickly appears that the dialect of the *Misāʾid* patterns with camel-rearer type. More precisely, it also appears to be closely connected to Cantineau's Bc type. In addition to this, our sample also reveals *šāwi*-like features such as the realisation of etymological /ġ/ as [q] (Younes and Herin 2016) and the treatment of diphthongs. Moreover, and quite surprisingly, some features that are attested in the North-West Arabian sub-group were found in the data. These are for example the resyllabification of *C₁aC₂aC₃v in derived forms such as *inC₁aC₂aC₃a, the second syllable stress in plurals of the *C₁vC₂vC₃ pattern (which may also lead to first vowel elision), and also *saʿal* for *saʿal*. In conclusion, the dialect of the *Misāʾid* matches for the most part the Bc sub-group but with *šāwi*-like features and also characteristics that are reminiscent of the North-West Arabian type. The question is how to account for such a pattern. There are at least two possibilities. The first one is that more complex configurations may have been unnoticed by Cantineau who indeed was not in a position to get large samples of data from all the tribes in the area. The second possibility is that recent dialect contact between speakers of all these sub-groups may have occurred, leading to dialect mixtures, as instantiated in our sample.

In terms of data collection and methodology, fieldwork in contexts that involve a fair amount of dialect contact can yield puzzling and conflicting linguistic output. This can also be exacerbated by short-term accommodation in the direction of the speech variety of the researcher(s). It is therefore paramount to secure the presence of an insider participant who can take the lead in carrying out data collection.

As far as the general classification of the dialects of Jordan and beyond is concerned, combining Herin and Younes' amendments to Cleveland, Palva, and Cantineau's classifications, it seems reasonable to posit the following taxonomy. We suggest that subsequent research should be framed within this canvas.

- (I) Sedentary *biġūlu*
 - a. Muʾābi (southern, Karak, Ṭafile, etc . . .)
 - b. Balgāwi-Ḥōrāni (central-north, Salt, ʿAĠlūn, etc . . .)
- (II) Southern Bedouin *yġūlu* (Ḥwētāt, Bdūl, Zawāyda, etc . . .)
- (III) Southern Bedouin *biġūlu* (mostly Nagab and Sinai)
- (IV) Central Bedouin *yġūlu* (ʿAĠarma, ʿAdwān, ʿAbābīd, etc . . .)

- (V) Northern Bedouin *ygūlūn*
 - a. ʿNizi
 - b. Šammari
 - i. Bc (Misāʿīd)
 - c. Šāwi
 - i. Ca (Bū ʿĪd et ʿĪdīn in Lebanon, so far unattested in Jordan)

Table 1. Features of the Misāʿīd and the Bedouin sub-groupings.

	A (ʿNiza)	B (Šammari)	C (šāwi)	Bc (Sattelite Šammari)	D (North-West Arabian)
<i>imāla</i> treatment	X	X		X	
Aspiration of <i>-a</i> in pause	X	X		X	
Diphthongs			X		X
Affrication			X	X	
Etymological /Ġ/	X	X		X	
<i>qalqala</i>			X		
Elision of /t/ in <i>-āt#</i>		X		X	
Resyllabification of *C ₁ aC ₂ aC ₃ v	X	X	X	X	
<i>saʿal</i> for <i>saʿal</i>					X
Resyllabification in derived verbs					X
<i>gahawa</i> syndrome	X	X	X	X	X
Stress on plural C ₁ vC ₂ ŷC ₃					X
Stressed definite article <i>al-</i>	X	X	X	X	X
Trochaism	X	X		X	
3.m.pl. perfective <i>-am/-aw</i>			X		
Vowel /a/ in the imperfective	X	X		X	
<i>kala-xaḡa</i>		X	X	X	
Form IV	X	X	X	X	X
Form V et VI ytiC ₁ aC ₂ C ₂ aC ₃ /ytiC ₁ āC ₂ aC ₃	X	X		X	
Apophonic passive	X	X		X	
Free pronouns <i>ana-iḡna</i>			X	X	
Bound 1.sg.- <i>an</i>		X		X	
Bound 3.m.sg.- <i>wo</i>		X		X	
Bound 3.f.sg.- <i>ah</i>	X	X		X	

Author Contributions: Conceptualisation, B.H. and E.A.-W.; Data collection, B.H., E.A.-W., Y.A.-S.; Transcription, I.Y.; Analysis, I.Y. and B.H.; Writing, B.H., I.Y. and E.A.-W.; Writing review and editing, E.A.-W. and B.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of the University of Essex on the 30th of November 2017.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Restrictions apply to the availability of these data. Data was obtained from the speakers and are available from the authors upon request and permission of the participants.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

- ¹ The research in Azraq is led by Bruno Herin and Enam Al-Wer and involves several local field researchers. The Druze of Jordan originally migrated from Swēda and the villages surrounding it in Syria.
- ² The interview sessions were primarily led by Youssef Al-Sirour who is a native speaker of the dialect under investigation. Also present were Enam Al-Wer, Bruno Herin, and Dina Oweidat, all of whom are speakers of urban central Jordanian dialects.
- ³ Incidentally, this sentence also features the deitic adverb *hānā*, which as far we know is typical of the Bc group (*šāwi* influenced *šammari* dialects).

References

- Blanc, Haim. 1964. *Communal dialects in Baghdad*. Cambridge: Distributed for the Center for Middle Eastern Studies of Harvard University by Harvard University Press.
- Cantineau, Jean. 1936. *Etudes sur Quelques Parlers de Nomades Arabes d'Orient: Premier Article*. Paris: Librairie Larose.
- Cantineau, Jean. 1937. *Etudes sur Quelques Parlers de Nomades Arabes d'Orient: Deuxième Article*. Paris: Librairie Larose.
- Cleveland, Ray L. 1963. A Classification for the Arabic Dialects of Jordan. *Bulletin of the American Schools of Oriental Research* 171: 56–63. [CrossRef]
- Herin, Bruno. 2019. Traditional dialects. In *The Routledge Handbook of Arabic Sociolinguistics*, 1st ed. Routledge Language Handbooks Series; Edited by Enam Al-Wer and Uri Horesh. New York: Routledge, pp. 93–105. Available online: <https://www.taylorfrancis.com/books/9781317525011/chapters/10.4324/9781317522450-7> (accessed on 20 December 2019). [CrossRef]
- Herin, Bruno. 2020. Northern Domari. In *Arabic and Contact-Induced Change*. Edited by Christopher Lucas and Stefano Manfredi. Berlin: Language Science Press, pp. 489–509. Available online: <https://zenodo.org/record/3744543> (accessed on 21 December 2020). [CrossRef]
- Holes, Clive. 2004. Tanwin in the arabic dialects of eastern and south-eastern Arabia. *Estudios de Dialectología Norteafricana y Andalusí, EDNA* 8: 89–97.
- Palva, Heikki. 1984. A General Classification for the Arabic Dialects Spoken in Palestine and Transjordan. *Studia Orientalia Electronica* 55: 357–376.
- Palva, Heikki. 2011. Northwest Arabian Arabic. In *Encyclopedia of Arabic Language and Linguistics*. Leiden: Brill, Available online: https://referenceworks.brillonline.com/entries/encyclopedia-of-arabic-language-and-linguistics/northwest-arabian-arabic-EALL_COM_vol3_0233 (accessed on 20 November 2020).
- Younes, Igor, and Bruno Herin. 2016. Šawi Arabic. In *Encyclopedia of Arabic Language and Linguistics*. Leiden: Brill, Available online: https://referenceworks.brillonline.com/entries/encyclopedia-of-arabic-language-and-linguistics/sawi-arabic-SIM_001011?s.num=8 (accessed on 20 November 2020).

Article

New Perspectives on the Urban–Rural Dichotomy and Dialect Contact in the Arabic *gəlat* Dialects in Iraq and South-West Iran

Bettina Leitner

Department of Near Eastern Studies, University of Vienna, 1090 Vienna, Austria; bettina.leitner@univie.ac.at

Abstract: This paper reevaluates the ground on which the division into urban and rural *gəlat* dialects, as spoken in Iraq and Khuzestan (south-western Iran), is built on. Its primary aim is to describe which features found in this dialect group can be described as rural and which features tend to be modified or to emerge in urban contexts, and which tend to be retained. The author uses various methodical approaches to describe these phenomena: (i) a comparative analysis of potentially rural features; (ii) a case study of Ahvazi Arabic, a *gəlat* dialect in an emerging urban space; and (iii) a small-scale sociolinguistic survey on overt rural features in Iraqi Arabic as perceived by native speakers themselves. In addition, previously used descriptions of urban *gəlat* features as described for Muslim Baghdad Arabic are reevaluated and a new approach and an alternative analysis based on comparison with new data from other *gəlat* dialects are proposed. The comparative analysis yields an overview of what has been previously defined as rural features and additionally discusses further features and their association with rural dialects. This contributes to our general understanding of the linguistic profile of the rural dialects in this geographic context.

Keywords: dialect classification; dialect contact; urban; rural; *gəlat*; *qəltu*; spoken Arabic

Citation: Leitner, Bettina. 2021. New Perspectives on the Urban–Rural Dichotomy and Dialect Contact in the Arabic *gəlat* Dialects in Iraq and South-West Iran. *Languages* 6: 198. <https://doi.org/10.3390/languages6040198>

Academic Editors: Simone Bettega and Roberta Morano

Received: 19 October 2021
Accepted: 18 November 2021
Published: 30 November 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

This study aims at a critical reevaluation of the urban–rural division in the *gəlat* dialects and the description of linguistic dynamics correlating with urbanization tendencies.¹ The urban–rural dichotomy is used in the descriptions of Arabic dialects from different regions (cf., for example, Abd-el-Jawad 1986; Abu-Haidar 1988; Ech-charfi 2020; Holes 1995; Ingham 1973; Miller 2007; Sharkawi 2014). However, until today there is only a small amount of evidence for common linguistic tendencies found among Arabic dialects in urban contexts (cf. Miller 2007, p. 2). Similarly, the clear-cut distinction into urban vs. rural regarding the *gəlat* dialects² of Iraq and Iran still seems to be built on weak ground. This study tries to sum up what we do and what we do *not* know about the division of the *gəlat* dialects into rural and urban ones. By including an areal perspective and new data from the *gəlat* dialects of Khuzestan, we hope to arrive at a more detailed description of the characterizing factors of rural dialects in general and the linguistic consequences of urbanization for dialects of the *gəlat* group more specifically.

The present work brings together hitherto used linguistic criteria for the distinction into rural vs. urban *gəlat* and other features determined by the author as possibly rural. The existence of these features is compared in the *gəlat* dialects described so far, including new data from the *gəlat* dialects in Khuzestan and two cognate dialects (Šāwi and Khorasan Arabic).

The study also tries to retrace what processes are at work when different *gəlat* dialects are in contact in urban contexts and questions the often not well defined and synonymous use of the terms ‘Bedouin’ and ‘rural’, as well as ‘sedentary’ and ‘urban’.

The study includes a small-scale sociolinguistic survey revealing what native speakers of (urban) Iraqi Arabic subjectively tend to perceive as typically rural or urban.

In general, the study and classification of the *gəlat* dialects has received only marginal attention, especially when compared with the much better studied *qəltu* dialects (cf.

Hassan 2021, p. 51). Even though a number of studies on *gələt* dialects has been published since Blanc's classification of the Iraqi dialects into *qəltu* and *gələt* in his seminal work on Baghdadi Arabic (Blanc 1964), they are still few. Among those who rely on freshly gathered data are: Hassan (2016a, 2016b, 2017, 2020, 2021), Mahdi (1985), Denz and Edzard (1966) and Abu-Haidar (2002) on South Iraqi Arabic; Ingham (1973, 1976), Leitner (2019, 2020) and Bettega and Leitner (2019) on Khuzestani Arabic; and Salonen (1980) on al-Shirqat (Širqāt)/Assur Arabic.

On lacking definitions: What do 'Bedouin', 'sedentary', 'urban', and 'rural' mean in the context of the gələt and qəltu dialects?

In his important paper on the linguistic character and development of Muslim Baghdad Arabic, Palva appears to use the labels 'urban' and 'sedentary', and 'rural' and 'Bedouin' interchangeably.³ By using the terms 'urban' and 'sedentary' as well as 'rural' and 'Bedouin' as quasi-synonyms (cf. Ech-charfi 2020, p. 67), we ignore the different nature of these terms and the different implications they have or had as socio-economic criteria at different times in history. It also ignores the fact that, for example, sedentary dialects can be rural as well. In Iraq, *urban* and *sedentary* as well as *rural* and *Bedouin* are indeed often closely linked, but a synonymous use of these concepts, especially for descriptions of the present-day linguistic classification, would be misleading. While—in our geographic context—*urban* and *sedentary* are concepts historically related to the *qəltu* dialects of Iraq, the concepts *rural* and *Bedouin* are historically associated with the *gələt* dialects.⁴

However, all four terms might denote very different things when looking at the present-day *gələt* dialects of Iraq and south-west Iran. Even though originally rural in character, many *gələt* speakers (or their ancestors) have moved to urban contexts and gradually replaced their rural identity with an urban one. Similarly, even though ultimately the *gələt* dialects are Bedouin, at present, the vast majority of their speakers lead a sedentary lifestyle. The usefulness of the latter distinction (sedentary vs. Bedouin type) has been recurrently criticized in the past years by scholars such as Janet Watson (2011, p. 859), who describes the Bedouin/sedentary split as “an oversimplification and of diminishing sociological appropriacy”.

For synchronic descriptions of present-day Iraqi Arabic, it is therefore mainly the terms 'rural' vs. 'urban' that remain useful to describe the different socio-economic circumstances people live under, whereas the importance of the question of sedentary (*ḥadār*) vs. Bedouin (*badu*) appears to be generally decreasing. This is also reflected in the sociolinguistic interviews I conducted with native speakers of Iraqi Arabic (see Section 3.3), who more often described certain features as typical of the countryside (*rīf*, *aryāf*) or the city (*madīna*) than as typical of the Bedouin (*badu*; *ʿašāyir* lit. 'tribes'). In these interviews, the participants never use the term sedentary (*ḥadār*) to characterize or specify the use of a feature. Still, it is important that these terms are apparently meaningful to native speakers in the present day.

In Khuzestan the urban–rural distinction has not played a role for a long time, as most inhabitants are of rural origins, and distinctions were made based on other socio-economic factors closer to the sedentary–nomad split (cf. fn. 13). However, modern-day Khuzestan has witnessed a rapid growth of urban centers, especially in the city of Ahvaz (cf. 3.4), for which reason the term 'urban' and its socio-linguistic and socio-economic implications (e.g., increase of contact and leveling tendencies) must at least be considered as an arising category.

Regarding the historical linguistic situation in Baghdad, the predominant Muslim dialect used to be (according to Blanc 1964, p. 170, at least until the fourteenth century) of the *qəltu* type and thus was characterized, as stated above, by the features [+urban] and [+sedentary]. Starting in the fourteenth century, the city of Baghdad was populated by incoming *gələt*-speakers (especially in the seventeenth and eighteenth centuries, cf. Palva 2009, p. 32), initially carrying the features [+rural] and [+Bedouin] and remodeling the former linguistic character of Baghdad, so that over the time its *gələt* character has become predominant. Nowadays, MBA (Muslim Baghdad Arabic) is a *gələt* dialect associated by

Arabic dialectologists with the features [+urban] and [+Bedouin], since the incoming rural Bedouin dialects in Baghdad have been urbanized (cf. Palva 2009, p. 38).

This process, the urbanization of rural dialects, contains the loss of highly marked rural features (e.g., the *gahawa*-syndrome), motivated by the speakers' wish to adapt to the urban linguistic profile. This is, of course, also observed for other urbanized *gələt*-speaking contexts, which do not have a *qəltu* substrate as we find it in MBA. The difference between MBA and other urbanized *gələt* dialects, which do not have a *qəltu* substrate, is that the *gələt* dialects in Baghdad have adopted some *qəltu* features (e.g., the marking of definite objects with a proclitic *l-*, cf. Palva 2009, p. 22). In this light, it appears useful to distinguish MBA from other *gələt* dialects that are nowadays spoken in (arising) urban contexts and which lack a *qəltu* substratum (e.g., Basra Arabic and Ahvazi Arabic).

This paper follows the assumption that eventually all *gələt* dialects are originally rural and Bedouin in character but focusses on the present-day definition of rural *gələt* features and their (lack of) prestige analyzing which features tend to be modified most readily in urban contexts.

The abovesaid shows the multifaceted nature of the terms 'Bedouin', 'sedentary', 'urban', and 'rural' and their historical and modern-day application for the regions of Iraq and south-west Iran.

Aims of This Paper

The purpose of this paper is twofold and can roughly be divided into one part focusing on synchronic aspects and the other dealing primarily with diachronic aspects. The former (Sections 3.1, 3.3 and 3.4) is dedicated to the following overarching questions:

(i) What unites rural *gələt* dialects? Which features are marked rural features, i.e., strongly associated with rural speech by *gələt* speakers themselves?

(ii) What happens to rural dialects when their speakers move to urban contexts? Which features tend to emerge (innovations) or be dropped as a consequence of the adoption of an urban lifestyle by *gələt* speakers?

The diachronic part of this study (Section 3.2) is a critical evaluation of Palva's derivations of certain MBA features (or lack of certain features in MBA) via the *qəltu* substrate, offering alternative explanations for the development of these features.

Against this background, this paper aims at reevaluating the hitherto applied linguistic criteria for the subclassification of the *gələt* group into urban and rural dialects and sheds new light on the question of the linguistic dynamics found in urban *gələt*-speaking contexts.

Section 2 presents the methods applied to answer the questions outlined above and the linguistic features focused on. Section 3 of this paper discusses the results of this study: Section 3.1 presents the distribution of the rural features analyzed, and is followed by a reevaluation of those MBA features which Palva explained as consequences of the *qəltu* substrate (Section 3.2). Section 3.3 presents the results of a small-scale sociolinguistic survey conducted among five urban Iraqis who fled to Vienna during the past five years on subjective perceptions of rural forms in Iraqi Arabic. Section 3 closes with a case study of the city of Ahvaz, pointing out linguistic tendencies found in *gələt* dialects spoken in arising urban contexts (Section 3.4).

Section 4 discusses the results of this analysis in the light of the questions proposed above. Section 5 concludes the study and provides an outlook on possible future studies on the urban–rural distinction in the *gələt* dialects.

2. Materials and Methods

In order to answer the above-outlined research questions, this paper starts with a comparative overview of seven phonological and morphological features and their existence in the *gələt* dialects of Khuzestan, Kwayriş/Babylon, al-Shirqat/Assur, Basra, and Muslim Baghdad. While the first three are usually associated with rural speech, the latter two are usually taken to represent urban-type *gələt*. The analysis further considers the existence of these linguistic variables in the Šāwi dialects of Syria and south-eastern Anatolia⁵ and

the Arabic dialects of Khorasan. Including the Šāwi dialects and Khorasan Arabic hopefully contributes to a better understanding of their obvious typological proximity to the *gələt* dialects.

The following features investigated for the purpose of this paper have either been listed by Blanc (1964, p. 166)⁶ and/or Palva (2009, pp. 21–29) as typically rural (i), or are suggested by the author of this paper as possible further rural features (ii):

- (i) Rural *gələt* features as listed by Blanc and Palva:
 - Affrication of *q > g > ġ and *k > č in the vicinity of front vowels (Section 3.1.1);
 - Use of the *gəhəwə*-syndrome (Section 3.1.3);
 - Resyllabification of CaCaC-v(C) > CCvC-a(C) (Section 3.1.4);
 - Retention of gender distinction in the plural of pronouns and verbs (Section 3.1.5).
- (ii) Further rural features suggested by the author of this paper:
 - Raising (and elision) of *a in pre-tonic open syllables (Section 3.1.2);
 - Prefix *tv-* (vs. urban *t-*) for Form V and VI verbs (Section 3.1.6);
 - Imperative M.SG of final weak roots of the form ?vCvC (Section 3.1.7).

3. Results

This section discusses possible rural *gələt* features and their distribution based on the available sources on *gələt* dialects (illustrated in Table 1; features I-III are phonological, while IV-VII are morphological features).

Table 1. Distribution of (possible) rural *gələt* features⁷.

Rural Features	Khuzestan	Kwayriš	al-Shirqat	Khorasan	Šāwi	Basra	Muslim Baghdad
I. Affrication of *q in front vowel environment	Yes	Yes	Yes	Yes	Yes	Yes	No
II. Raising of *a in pre-tonic open syllables	Yes	Yes	Yes	Yes	Yes	Yes	Yes
III. <i>gəhəwə</i> -syndrome	Only a few remnants	Yes	Only a few remnants	Yes	Yes	Only a few remnants	No
IV. CaCaC-a(C) > CCvC-a(C)	Partly	Yes	Yes	No	Yes	No	No
V. Gender distinction in the plural	Yes	Yes	Yes	Yes	Yes	Yes? ⁸	Partly
VI. Form V and VI prefix <i>tv-</i>	Partly	Partly	Yes	Arabkhane: <i>ti-</i> Khalaf: <i>it-</i>	Yes	No	No
VII. Imperative SG.M of weak verbs: ?vCvC	Partly	Partly	No evidence found	Yes	Partly	No	No

3.1. Rural *gələt* Features

3.1.1. Affrication of OA (Old Arabic) *q > g > ġ and *k > č in the Vicinity of Front Vowels

Generally, the phenomenon of a phonetically conditioned affrication of OA *k and *g is considered typical of Eastern Bedouin-type dialects of the Syro–Mesopotamian area (Palva 2006, p. 606). The phonetically conditioned affrication of *k is basically a feature shared by all *gələt* dialects but is somewhat more limited in urban varieties (Blanc 1964, p. 166): compare, e.g., MBA and Basra Arabic *akəl* (Blanc 1964, p. 166; Mahdi 1985, p. 64) and Khuzestani Arabic *ačəl*.

Similarly, the affrication of *q is traditionally more strongly associated with the rural type (Fischer and Jastrow 1980, pp. 142–43; cf. Blanc 1964, pp. 25–28, who calls the affrication of *q “a hallmark of the countryside”; Palva 2009, p. 37, fn. 19). According to

Blanc's informants (Blanc 1964, pp. 27–28), speakers perceive forms with \check{g} (< * q) instead of g or q as rural or 'provincial', or the use of g or retention of * q as urban.

According to the descriptions of Blanc (1964, pp. 26–27), in Muslim Baghdad Arabic the general reflexes of * q are g and q , thus without an affricated realization in front vowel environments. More recently, Palva notes for Muslim Baghdad Arabic that "the contrast between urban and rural $g\text{ə}l\text{ə}t$ is diminishing", because such features as the conditioned affrication of g (as well as the use of feminine plural forms in the 3rd person, cf. 3.1.5) are gaining ground in that dialect (Palva 2009, p. 37, fn. 19 and the references mentioned there).⁹

For Basra Arabic, Mahdi (1985, pp. 86–87, fn. 102) states that there is some variation between g and \check{g} .

We also find these phenomena in the Šāwi-dialects (e.g., *čitīr* 'much' and *ğidīm* 'old', Younes and Herin n.d., EALL Online), and in Khorasan Arabic (e.g., *čitab* 'he wrote'; Seeger 2013, p. 314, and *ğirīb* 'close', Seeger 2009, p. 310).

Regarding other Iraqi Arabic dialects, affrication of * q > \check{g} is further attested in texts from al-Ḥilla (Denz and Edzard 1966, p. 68: *ğiddām* 'in front of', or 70: *rfiği* 'my friend').¹⁰

Thus, at present these phenomena also appear in urban contexts, although apparently to a lesser degree: In MBA, the 'default form' is still unaffricated, in Basra there is variation between affricated and non-affricated forms, and we find the same variation in present-day Ahvazi Arabic, e.g., *mədayyəğ* ~ *mədayyəg* 'worried', *şədəğ* ~ *şədəg* < *şidq*¹¹ 'truth', *bəčān* ~ *bəkān* 'place', and the progressive marker *gāfiid* ~ *ğāfiid* cf. 3.2.2; Leitner 2020, pp. 30, 32). This might point at a tendency in urbanizing contexts towards de-affrication or replacement of \check{g} and \check{c} with the less marked or less 'provincial' g and k . In other words, its marked rural character (cf. 3.3) makes this phonetic feature prone to be given up in contact with another dialect or other dialects. This, of course, contradicts Palva's statement (as cited above) that the phenomenon of affricating g is gaining ground in MBA. This contradiction might result from analyzing data from different speech communities (Shiite vs. Sunnite; different quarters, etc.) of a city. Based on my data, however, I cannot confirm his observation, but must rather argue the contrary.

3.1.2. Raising of OA * a in Pre-Tonic Open Syllables

Examples from the dialects analyzed, which feature the raising of * a not only in *CaCin patterns, are: Kwayriš *šibāb* 'youth' (Denz 1971, p. 66); Khuzestan *səwələf* 'stories' (Leitner 2020, p. 43); Khorasan *miğlitin* 'gathering' (Seeger 2002, p. 637); Šāwi *sičāčim* 'knives' (Younes and Herin n.d., EALL Online); al-Shirqat: *dībāyeh* 'slaughter animals' (Salonen 1980, pp. 9, 28, Text 1, sentence 17); MBA: *sičāčim*¹¹ ~ *sičāčim* 'knives'; and Basra: *diğāğa* 'chicken' (Denz and Edzard 1966, p. 80, Text VII) ~ *dyāy* (Mahdi 1985, p. 247). The overall picture we get from the distribution of this feature is that synchronically this phonological change is found in both urban and rural dialects. Examples such as *maṛākub* ~ *mṛākub* 'ships' and *manāqil* ~ *mṇāqil* 'barbecues' from Basra Arabic (Mahdi 1985, pp. 141–42), which appear both with and without the raising and subsequent elision of * a in the first (pre-tonic and open) syllable, might point to a slight tendency among urban varieties to preserve * a . However, this tendency was not really confirmed by the results of the sociolinguistic survey conducted for this study (cf. 3.3), in which most speakers produced forms with a raised * a .

Even though this phonological process is often inhibited by consonants of the guttural group (cf. Younes 2018, p. 5), we find various counterexamples, such as *mṣābed* 'temples' (Salonen 1980, pp. 10, 29, Text 1, sentence 31) < *maṣābid* probably via raising and subsequent elision of * a in the first syllable. Younes argues that in the Middle East this phenomenon probably predates the appearance of the *gahawa*-syndrome (Younes 2018, pp. 7–8) and treats it as a pan-Eastern Bedouin dialect phenomenon.

3.1.3. *gahawa*-Syndrome

The so-called *g(a)hawa*-syndrome is another feature described by Blanc (1964, p. 166) as typical of rural *gəlat* dialects and not present in MBA. This morphological phenomenon denotes the reshuffling of non-final syllables closed by a guttural consonant (i.e., /x, ǧ, ħ, ʕ, ʔ, h/): CvC_G > CvC_Gv or CC_Gv, e.g., *gahwa* ‘coffee’ > *gahawa* or *ghawa*.

As for MBA, the incoming Bedouin tribes and the rural population that has settled in Baghdad apparently have given up this feature in the urban context.

Although there are many traces of this resyllabification rule still found in contemporary Khuzestani Arabic—e.g. *ʔahali* ‘my family’ and *xadār* ~ *ʔaxadār* ‘green’—it has ceased to be an active phonological process (cf. Section 3.4 and Leitner 2020, p. 50).

Also for Basra Arabic, there are only very few examples of this phonological rule to be found in Mahdi’s Ph.D. thesis, but many, which do not show the *gahawa*-syndrome, e.g., *naʕya* ‘ewe’ (not *nʕaya*) and *yiʕruf* ‘he knows’ (not *yʕaruf*) (Mahdi 1985, pp. 51, 99, respectively). The only examples found are some originally ʔ-initial words like *xadār* < *ʔaxdār* ‘green’ or *ħawal* < *ʔahwal* ‘cross-eyed’, in which the first syllable was dropped after a vowel *a* was inserted after the guttural consonant (Mahdi 1985, p. 62).

The data on Kwayriš Arabic also shows mixed results: while we find, e.g., *naxla* ‘palm tree’ (cf. Meißner 1903, p. XVIII), we also get *lighawa* ‘the coffee’ (Denz 1971, p. 55), (*a)heli* ‘my family’ (Meißner 1903, p. 26), and *yǧalub* ‘he wins’ (Denz 1971, p. 68).

As for al-Shirqat, I found one word that is subject to this phonological rule in Salonen’s texts: *ʔäheli* ‘my family’ (Salonen 1980, pp. 21 and 42, Text 8, sentence 1) and *ʔahalu* ‘his family’ (Salonen 1980, pp. 22 and 44, Text 9, sentence 4)—but later, in another text, we find the same word without insertion of a vowel after the guttural: *ʔahlu* ‘his family’ (Salonen 1980, pp. 24 and 46, Text 13, sentence 1).

This phenomenon is attested for both the Šāwi dialects (e.g., *ǧħaša* ‘female ass’, instead of *ǧahša*; Procházka 2003, p. 78), and for Khorasan Arabic (e.g., *yogodi* ‘he goes’, instead of *yogdi*).

Today, the productive use of the *gahawa*-syndrome appears to be very limited in most dialects and has thus ceased to be a good criterion for distinguishing rural from urban dialects as well as Bedouin-type from sedentary-type dialects. The reason for the loss of this feature is most likely related to its markedness (cf. Section 3.3).

3.1.4. Resyllabification of OA CaCaC-v(C) > CCvC-a(C)

Blanc described reflexes of the OA PFV verbal forms CaCaC-v(C), e.g., *katabat*, with initial CC- as typical rural *gəlat* forms (Blanc 1964, p. 166). Ingham (1982, pp. 48–49, 52) describes such forms as characteristic of the Mesopotamian *bādiya* dialects, in contrast to the Mesopotamian *ħadār* dialects that have an initial syllable structure CiC-.

Basra Arabic and MBA both have forms of the structure CvCC-v(C), e.g., **katabat* > *kətbət/kitbat* ‘she wrote’ (Blanc 1964, p. 98; Leitner et al. 2021, p. 69; Mahdi 1985, p. 93). Also in present-day Khuzestani Arabic, the most common reflex is CvCC-v(C) and not CCvC-v(C)—e.g. *kətbət* ‘she wrote’. In Khorasan, we find the structure CiCiC-v(C), e.g. *čitibat* ‘she wrote’ (Seeger 2013, p. 314).

Forms of the type CCvC-v(C), e.g., *ktibət* ‘she wrote’, are found, e.g., in Kwayriš (Meißner 1903, pp. XLI, LII)¹², al-Shirqat (Salonen 1980, p. 80), in all Šāwi dialects (Younes and Herin n.d., *EALL Online*), and in certain (ʕarab-type¹³) dialects of Khuzestan (Leitner 2020, pp. 14–15). Among the urban speakers of Ahvaz, such forms are not used and are rather perceived as clearly rural.

3.1.5. Retention of Gender Distinction in the Plural

Palva states that gender distinction in the plural is a feature retained only in the rural dialects of southern Mesopotamia (Palva 2009, p. 23; cf. also Blanc 1964, p. 166, who states that this phenomenon is “only marginal in M[BA]”). He explains the lack of gender distinction in the plural of MBA as an inherited *qəltu* trait, albeit admitting that sedentarization or urbanization processes would probably have led to the same

development (*ibid.*). Ingham states that, based on his material from Iraq, Khuzistan, Kuwait, and Northern Najd, gender “distinction was maintained almost everywhere, except in the urban centres of Zubair, Kuwait, Basra and Baghdad” (Ingham 1982, p. 38).

My data shows that gender distinction in the plural is still maintained in all Khuzestani Arabic dialects, even in (modern) urban contexts (cf. the case study of Ahvaz below, Section 3.4), and it appears to be used in modern MBA and Basra Arabic as well (see Table 1 and fn. 8). Gender distinction in the plural was also maintained by all urban Iraqi speakers interviewed for the sociolinguistic study described in Section 3.3 and none of the interviewees associated the use of feminine plural forms with rural speech. Therefore, we might need to rethink the strict association of this feature with rural contexts.

3.1.6. Prefix *tv-* in Form V and Form VI Verbs

The retention of the vowel after the prefix *t-* in Pattern V and VI verbs, e.g., al-Shirqat *taḥawwalaw* ‘they moved’ (Salonen 1980, pp. 11, 31, Text 2, sentence 12) and Kwayriš *yatalagga* ‘he meets’ (besides *ḥlagga*, Meißner 1903, p. XLIV), might be another rural feature of the broader Mesopotamian area.

At least for Khuzestani Arabic, my data suggests that forms with a vowel are typical of rural areas found, e.g., in the dialect of Ḥamidiyya, e.g., *tačabbašət* ‘I have learnt’, *nətaḥašša* ‘we have dinner’.¹⁴

Apart from rural Khuzestani Arabic, Kwayriš and al-Shirqat Arabic, we also find this trait in Khorasan Arabic (Volkan Bozkurt, pers. comm.), and in the Šāwi dialects (Behnstedt 1997, pp. 328–29, map 164; Behnstedt 2000, p. 444; Bettini 2006, p. 33). In contrast, it appears to be completely absent from the dialects of Basra and Baghdad, as well as urban Khuzestani Arabic (at least for the city of Ahvaz).

We might tentatively propose that the retention of the vowel in the prefix of Form V and Form VI verbs is a rural feature. However, this hypothesis definitely needs further elaboration and more data from other rural areas, especially in the form of sociolinguistic surveys.

3.1.7. Imperative SG.M of Final Weak Roots: ?vCvC

In MBA and Basra Arabic, the SG.M imperative of final weak roots ends on a vowel, e.g., *imši* ‘go (IMP SG.M)’ (Blanc 1964, p. 103; Mahdi 1985, p. 125). In Kwayriš, as well as in present-day Ahvazi Arabic, in addition to these forms, we also find forms lacking the final vowel, e.g., *ʔəməš* ‘go (IMP SG.M)’ (Leitner 2020, p. 19; Meißner 1903, p. XLVIII). The latter forms (lacking the final vowel) are also found in the Arabic varieties of Khorasan (?vCvC Volkan Bozkurt, pers. comm.) and Kuwait City (?vCC, Yousuf B. AlBader, pers. comm.), as well as in some Šāwi dialects (e.g., Urfa Arabic, Stephan Procházka, pers. comm.; cf. Behnstedt 1997, pp. 404–5, map 202; and Bettini 2006, p. 35).

For Khuzestani Arabic, Ingham describes the imperative form lacking the final vowel as *ʔarab*-type and, conversely, the form with the final vowel as *ḥaḍar*-type (Ingham 2007, p. 577; 1973, p. 544; cf. fn. 13 at the end of this paper on these terms). He further describes the introduction of a prothetic vowel before an imperative of the structure C_əCC-v (i.e., IMP SG.F, PL.F, and PL.M; or IMP SG.M with a vowel initial object suffix)—e.g., *əkətbi* ‘write (IMP SG.F)’—as a rural feature (Ingham 1973, p. 542). In my corpus, this feature is also attested for speakers from Ahvaz.

The small-scale sociolinguistic study conducted for this paper confirms—at least partly, as some speakers stated that such forms did not exist—that urban speakers associate imperative forms of final weak roots lacking the final vowel with rural speech (cf. Section 3.3).

3.2. Reevaluating the Urban Character of MBA: A Question of Urban Features or Inherited *qəltu* Features?

3.2.1. Indefinite Article *fadd* ~ *fard*

The use of the indefinite article *fadd* or *fard* has been described as a shibboleth of Iraqi Arabic. Its use is, however, not limited to the nation of Iraq and we also find it in Arabic

dialects in Iran (in the provinces of Khuzestan, Bushehr, and Hormozgan), and in Central Asian Arabic (cf. Leitner and Procházka 2021).

Palva (2009, p. 23) describes the indefinite article as a “sedentary feature found in the Mesopotamian dialect area” that is probably quite old. We can assume that in principle this is an old sedentary feature that has most likely developed in urban contexts that allow for contact with other languages, which make use of an indefinite article, e.g., Persian. This supports the theory that new linguistic categories are more likely to arise in urban contexts and contact situations.

Other than that, looking at the emergence of this feature does not tell us much about the synchronic urban–rural distinction in the *gələt* dialects.

3.2.2. Progressive Markers *da-* and *gāʕid*

Palva argues that the clitic progressive marker *da-* < *qāʕid* is a sedentary feature and writes that “the use of verb modifiers to mark different tense and aspect categories is a prominent sedentary feature very well developed in all *qəltu*-dialects [. . .], whereas in rural *gələt* dialects these categories as a rule are unmarked” (Palva 2009, p. 20). Palva’s view on the distribution of this feature is contradicted by data from several rural *gələt* dialects. In fact, Palva himself states several pages later that we do find progressive markers in many rural *gələt* dialects as well (Palva 2009, p. 28: “In addition to the *qəltu*-type verb modifier *da-*, MB[A] also makes use of the unshortened active participle *gāʕid* in the same function [. . .]. This is an obvious imported rural *gələt*-type form . . . ”; cf. also Denz 1971, pp. 82–82, 110, 116 on *gāʕid* in Kwayriš).

This fact is supported by Hanitsch (2019, pp. 266–71), who even states that: “Die vollen Formen [of the verbal modifiers deriving from OA *qaʕad* ‘to sit’] sind praktisch über das gesamte arabische Sprachgebiet hinweg anzutreffen. Besonders typisch sind sie für Dialekte nomadischen Typs oder mit nomadischem Adstrat” (Hanitsch 2019, p. 267). She adds that even though this verbal modifier was especially typical of nomadic dialects, it was also found in *qəltu*-type dialects spoken in the Syro–Mesopotamian area, as well as in rural sedentary dialects in Morocco, Tunisia, and Palestine (Hanitsch 2019, p. 267).

As Palva (2009, p. 28) states, the use of this progressive marker is also well documented from the Syrian Desert (*gāʕid*) and Ḥōrān (*gāʕid*). It is also used in many Šāwi dialects (*gāʕid* ~ *gāʕd* ~ *gāʕd*; cf. Bettini 2006, p. 44; Procházka 2018b, p. 281; Younes and Herin n.d., EALL Online) and in Khorasan Arabic (Volkan Bozkurt; pers. comm).

Some southern Iraqi dialects, e.g., Basra Arabic (Mahdi 1985, p. 212) as well as Najaf and Amarah Arabic (information provided by native speakers) use *gāy* (active participle of the verb ‘to come’) to mark progressive aspect.

It thus seems that only the shortened form, *da-*, is typical of urban dialects, not the use of a progressive marker per se.

3.2.3. Future Marker *rāh*

The future marker *rāh* evolved by a grammaticalization process from *rāyəh* ‘going’, which is the AP of the verb *rāh* ‘to go’. *rāh* is used as a future marker in Khuzestani¹⁵ (Leitner 2020, p. 157), Basra (Mahdi 1985, pp. 210–11) and al-Shirqat Arabic¹⁶, and in all dialects of Baghdad (cf. Blanc 1964, pp. 117–18). In addition to the dialects primarily analyzed in this study, it is found, e.g., though not very frequently, in Bahraini Arabic (Holes 2001, p. 216; Holes 2016, p. 304; Johnstone 1967, p. 152; cf. Taine-Cheikh 2004, pp. 219–220; 231 for a good overview regarding its distribution including examples from North Africa and the Levant). In Kuwaiti Arabic it is used to express proximal intent ‘to be about to’ (Holes 2016, p. 304).

According to an informant from Kerbala (participant D in Section 3.3), some dialects in central Iraq, e.g., Kerbala Arabic, also use *ha* to mark future tense.

Kwayriš (cf. Denz 1971, p. 109, fn.11) and Šāwi Arabic have no future marker. Khorasan Arabic uses the particle *ʕūd* to mark future tense (Volkan Bozkurt, pers. comm.).

Based on the fact that cognate forms of this future marker occur in Baghdadi Arabic (Muslim, Jewish, and Christian), as well as in Egypt, Damascus, and Beirut, Palva describes this future marker as an old urban and sedentary feature (Palva 2006, p. 612; Palva 2009, p. 21).

If the future marker *rāḥ* is indeed an old sedentary feature, it must have at one point been adopted by Bedouin Arab tribes (e.g., those that then settled in Khuzestan). To determine at what point this adoption happened is made difficult by the lack of historical data of, for example, Khuzestani Arabic. Similarly, the question of *why* this feature was adopted—that is, whether it was motivated by, e.g., intensive contact with Iraqi speakers or via the spread of urban features (often connected with a certain prestige)—remains unanswered.

Regarding its current distribution, the use of this future marker is definitely a feature typical of both the sedentary-type *qəltu* and the Bedouin-type *gəlat* dialects (like the indefinite article, cf. 3.2.1). It would be interesting to find out, for example, whether this future marker is still not used in Kwayriš or in other Iraqi rural areas nowadays. In case it is used in present-day Kwayriš Arabic, this would support the theory of features found in urban centers spreading to rural areas. From a synchronic point of view, and based on our limited available data, we cannot, however, solidly claim that the use of the future marker *rāḥ* is an exclusively urban feature of present-day *gəlat* dialects.

3.2.4. Emphatic Imperative Prefix *d-*

The prefix *də-* ~ *d-* is used to express an emphatic, more energetic form of imperative, as in the following example from Khuzestani Arabic (cf. Leitner 2020, p. 166 for more examples):

(1) Ahvazi Arabic (Leitner 2020, p. 166)

də-xáll *asələf* *xayya!*
EMP-HORT tell\IPFV.1SG sister.DIM

‘Let me tell (my story), sister (and don’t interrupt)!’

In addition to Khuzestan, this prefix is also used in the described function in the Arabic varieties of Baghdad (Blanc 1964, p. 117), Basra (Mahdi 1985, p. 107), Kwayriš (Meißner 1903, p. XXXIV), Mardin (*qəltu*), and Harran-Urfa (Şāwi) in eastern Anatolia (Procházka 2018a, p. 169), in Christian-Maslawi Arabic (Hanitsch 2019, p. 61), and in some sedentary-type Bahraini Arabic village dialects (Holes 2016, p. 202). I was unable to find evidence for the use of this feature in Salonen’s work on al-Shirqat (Salonen 1980), nor in Denz and Edzard’s text recorded from a speaker from al-Shirqat (Denz and Edzard 1966).¹⁷

Colloquial Persian (especially the northern varieties) also uses a prefix *d-* for strong or emphasized imperatives, e.g., *de-boro* ‘Go (now)!’ (pers. comm., Nawal Bahrani and Babak Nikzat, May 2021) and the Mandaic imperative prefix *d-* (see Häberl 2019, pp. 694–95) may also be related.

Due to its co-occurrence in all Arabic varieties of Baghdad, Palva describes it as an old *qəltu* and with that as a sedentary feature (Palva 2006, p. 612; Palva 2009, pp. 21–22).

As for the origin of this prefix, Grigore (2019, p. 114) derives it from Ottoman Turkish (as an abbreviated form of *haydi/hayed/hadi* ‘Come on!’) and states that *de-* is also found in this function in contemporary varieties of Anatolian Turkish. Procházka questions this derivation arguing that “Turkish possesses a distinct suffix to intensify imperatives (*-sana/-sene*) and the use of *haydi* together with such forms is only optional” and instead points out that the particle might as well be of Arabic origin and a reflex of the OA demonstrative *dā/dī* (Procházka 2018a, pp. 183–84).

Whatever its ultimate source (language), its present-day distribution allows us to consider it an old areal feature of the broader Mesopotamian linguistic area, presently found in both rural and urban *gəlat* dialects. One possible scenario for its distribution is that the Bedouin dialects that have arrived in the Mesopotamian area between the fourteenth and eighteenth centuries have adopted this feature from the (rural and urban) sedentary dialects.

In case its origin is actually Persian, the question of an ultimate sedentary or Bedouin character of this feature is in principle redundant, even though, of course, it might again be

the sedentary dialects that first adopted this feature from Persian and then passed it on to the Bedouin dialects.

3.2.5. Lack of Features: Feminine Plural Forms, Resyllabification Rules, and Form IV Verbs

To explain the absence of feminine plural forms as “an inherited *qəltu* trait” (Palva 2009, p. 23) seems a bit counterintuitive. Instead, we suggest to explain this phenomenon as a modification (reduction/loss) of dialectal features of the incoming *gələt* speakers rather than an adoption of the absence of a category. Even though the difference in this explanation might appear minor, we deem it important to acknowledge the directions of language change. This alternative interpretation is actually touched upon by Palva himself (ibid.), stating that “the natural drift combined with dialect contact would probably have led to the same development, as it has actually done as part of sedentarization process, e.g., in urban centers such as Basra, Zubair and Kuwait.”

Although the loss of feminine plural forms is indeed often connected with urbanizing processes, this is not necessarily the case (cf. Ahvazi Arabic 3.4), nor is the lack of feminine plural forms perceived as an urban feature by urban speakers of Iraqi and Khuzestani Arabic (cf. 3.3).

Similarly, Palva lists the absence of the Bedouin-type resyllabification rules, such as the *gahawa*-syndrome or the rendering of OA CaCaC-v(C) > CCvC-a(C), among the inherited *qəltu*-type features, but explains this as a “phonetic adaptation by immigrant Bedouin speakers” (Palva 2009, p. 24). As was shown in Sections 3.1.3 and 3.1.4, the active use of these phonological rules has reduced greatly, and its absence is not limited to *gələt* dialects that have been in contact with a *qəltu* dialect or been influenced by a *qəltu* substratum. It thus seems that this process (loss of phonological or morphological features) is triggered by the markedness of these features (cf. Section 3.3).

Finally, Palva also lists the absence of Form IV as a productive morphological category in MBA among the inherited *qəltu* traits, even though he himself admits that this feature was absent in Jewish and Christian Baghdadi as well (Palva 2009, p. 24). This verbal pattern has ceased to be productive in most dialects of this area (and beyond), also in rural dialects, and is therefore much rather a general tendency of spoken Arabic than a specific *qəltu* feature.

3.3. Markedness of Rural Features—A Small-Scale Sociolinguistic Survey among Urban Iraqis

Inspired by the question proposed in the introduction and the features previously described as rural (cf. Table 1), the author has undertaken a small-scale sociolinguistic survey among five urban Iraqis. Audio-recordings of all the interviews were made—of course, with the participants’ consent—and later partly transcribed. Some examples taken from these recordings are cited below with English translations.

The five participants, aged between 26 and 41 (three male, two female), are from different urban backgrounds (Falluja, Baghdad, Kerbala, Amarah), and all fled to Vienna within the past few years.

Participant A is a 33-year-old male graphic designer from Falluja, who fled to Vienna in 2015.

Participant B is a 41-year-old professional painter and was born and lived in Al-Adhamiyah in Eastern Baghdad until he fled to Austria in 2015.

Participant C is 26 years old and is also from Al-Adhamiyah, Eastern Baghdad. He went to study for some time near Tikrit. He then too fled to Vienna in 2015.

Participant D is a 28-year-old (female) doctor, who was born and studied in Kerbala, where she has also worked in a hospital after finishing her studies. She has moved to Vienna in September 2021.

Participant E is a 39-year-old mother of four children, who was born in the city of Amarah (Maysan), but then lived for more than a decade in Kirkuk before she and her family fled to Vienna in 2015.

A short interview was conducted with each interviewee individually (mostly in Iraqi Arabic) while focusing on three aspects:

- (i) Do the interviewees themselves reproduce rural features as described in Section 3.1?
- (ii) What do the participants think about the features described in Section 3.1 regarding their distribution, status, and use? (Which speakers, or in which regions, would we find them? Are they rural, *rīfi*, features?)
- (iii) What other features do the participants associate with rural speech?

To answer question (i), the participants were asked to translate certain words and phrases from English, German, or Standard Arabic into their native dialect to see whether they themselves reproduced rural features.

The other two questions were asked within an ‘open questions’ part of the interview, in which the participants could tell me what came to their mind when hearing specific words or features—e.g., the use of *ʔimiš* for ‘go’ (IMP SG.M) instead of *ʔimši*—or when thinking about the urban–rural distinction in the Iraqi dialects in general. Whereas for denoting ‘rural’ the Arabic adjective *rīfi* was used by all speakers and some of the younger speakers even used the English term ‘rural’ additionally (cf. the quote below), for the concept of *urban* no direct equivalent was used. Instead, the interviewees, for example, explained that *ahl il-madīna* ‘the city dwellers’ would use a certain form or not.

This survey yielded the following results represented in what we shall call a ‘Rurality Scale’. The stronger a feature is associated by speakers with rural contexts, the farther the bar related to these features goes to the right¹⁸ and, in consequence, the more likely it is that such a feature is modified or given up in urban contexts.

Of all (possibly) ‘rural’ features discussed in Section 3.1, only the raising of **a* and the use of gender distinction in the plural were found in the speech of my interviewees, who all lived in cities when living in Iraq. For the translation of ‘knives’, for example, all participants used the form *sičāčim* (< *sakākin* via raising of **a*), and the sentence ‘The mothers are baking bread’ was translated by all as *əl-ʔummaḥāt da-yixubzan*, i.e., using the third person feminine plural form of the verb.

As illustrated in Figure 1, among the ‘overtly rural features’ (cf. Abu-Haidar 1988, p. 75) in Iraqi Arabic are the *gahawa*-syndrome, the resyllabification of CaCaC-v(C) > CCvC-a(C), the affrication of **q*, and, albeit to a lesser degree, SG.M imperatives of IIIy/w verbs of the structure ʔvCvC, i.e. lacking the final vowel, and the pronunciation of **ğ* as [q].

The pronunciation of **q* as *ğ* in forms like **qalīl* > *ğīlīl* and **qarīb* > *ğirīb* was considered a rural feature by four participants (B, C, D, and E). Speaker D even emphasized that this was a *very* rural feature as such forms were used mainly in *əl-mukānāt əl kullīš rural* ‘areas that are totally rural’ and therefore in general not that commonly heard. Speaker B associated this with the speech of northern tribes, where he has relatives, and said it was possibly also found among rural speakers from the south, but that he was not sure about that because he did not know any southerners. Speaker A considered it a northern feature that was, however, not limited to rural contexts but also found in urban contexts in the north.

As for the resyllabification of CaCaC-v(C) > CCvC-a(C), only the youngest participant (C) explained that forms such as *ktəbat* ‘she wrote’ or *ğlubat* ‘she turned’ did not exist. The other four speakers (A, B, D, and E) considered both the resyllabification of CaCaC-v(C) and the *gahawa*-syndrome as typically rural (e.g., A *bi-l-aryāf* ‘in rural areas’) or Bedouin (e.g., B *šašayirna yğulūha* ‘our tribes say it’) features. Speaker A and B again associate these features particularly with the speech of northern tribes.

As for the SG.M imperative of final weak roots, all of the interviewees used forms ending in a vowel and two of them (B and C) considered forms lacking the final vowel—e.g. *ʔəməš* ‘go!’—as simply wrong or non-existent. The other three participants (A, D, and E) associated said forms with rural speech (e.g., A: *ʔimiš yistaxdimūnha b-ir-rīf ʔaktar* ‘*ʔimiš* is mainly used in rural areas’).

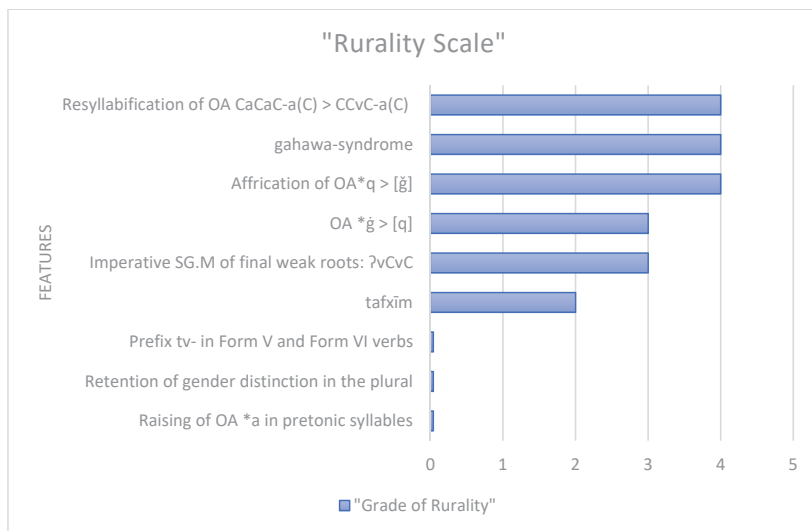


Figure 1. Scale of markedness of rural speech elements.

Interestingly, none of my interviewees associated the use of feminine plural forms with rural speech. In fact, most of them used verbs in the feminine plural form when translating sentences like ‘the girls washed . . .’ or ‘the mothers are baking bread’, but only the speaker from Amarah (E) explicitly stated that the sentence would be wrong if I used a masculine plural verb instead (which would have been acceptable for the others).

Similarly, the retention of the vowel after the prefix *t-* in Pattern V and VI verbs or the raising of **a* were not mentioned as rural features by any of the interviewees. All of the participants produced forms of Pattern V and VI verbs without vowel retention only, e.g., *nət ʔašša* ‘we eat dinner’, but did produce forms in which **a* was raised to *i*, e.g., *diḡāḡa* ‘a chicken’. The retention of the vowel in the prefix of Pattern V and Pattern VI verbs thus might be limited to rural areas but does not seem to be marked as a feature of rural speech. The raising of **a* is neither limited to rural areas nor a marked rural feature.

In addition to these features, three participants (A, C, E) mentioned the realization of **ḡ* as [q] and two (A and C) the use of *tafxīm* ‘emphasis’ (also described as ‘heavy speech’) as rural features. About the realization of **ḡ* as [q], participant A stated that *b-ir-rif mā nistaxdim il-ḡayn b-il-ʔumūm, nistaxdim il-qāf, maṭalan id-Deckel* [German for ‘lid (of a pot)’] *iḡna b-il-ʔāmmiyya inḡūl ʔalē qabaḡ humma iḡūlūn ʔalē qabaq ē kulla qāf* ‘on the countryside, they [lit. ‘we’] don’t use the [letter] *ḡayn* generally, they use the [letter] *qāf* [instead]; for example, for lid [of a pot] we say *qabaḡ* in our dialect, (while) they call it *qabaq*, yes, it’s all *qāf*s’. About the stronger use of emphasis in the dialects spoken in the north-western Iraqi province of Anbar, the same speaker says: *lahḡathum kulliḡ iḡīla* ‘their dialect is very heavy [i.e., characterized by emphasis]’.

Furthermore, participant A and C also mentioned the use of several lexical features as typically rural, e.g., the use of *xasš* ‘enter’ (instead of urban *ṭabb*, participant A¹⁹; cf. Abu-Haidar 1988), *dahḡiḡ~dahḡiḡ* ‘to see’ (vs. urban *bāwa ʔ*, participant C), and *ḡādi* ‘there’ (participant C).

As for the city of Baghdad, participant A (from Falluja) described its people and dialect as the most educated: *fa-tkūn lahḡathum hiyya l-lahḡa l-Karxiyya l-bēda lli hiyya qarība ʔa-l-fuḡḡā ya ʔni bī-ha kalimāt ʔāmmiyya u bī-ha fuḡḡā* ‘and their dialect is the “white” dialect of Karkh [Western part of Baghdad], which is close to the literary language, that is, it has dialectal words as well as words derived from the literary language’.

Participant E described the dialect of Baghdad as effeminate and unmanly, while she described the rural dialects as masculine.

In general, it appears that the purely geographical division into *ḡanūbi* ‘southern’, *b-il-waṣaṭ* ‘in the center, central’, and *ḡarbi* ‘western’ often plays a bigger role in the participants’ descriptions of the dialectal landscape that we find in Iraq.

Finally, it has to be mentioned that two pejorative terms used by city-dwellers to describe people from rural areas were mentioned in the interviews. These terms are *šrūgi* (PL *šrūḡ*)²⁰ and *mḡēdi* (PL *mḡāḡdān*). While the former mainly denotes people from the south, the latter essentially refers to the marsh-dwellers, many of whom live in the Eastern province of Maysan, but is now often used to derogatorily describe an uneducated, uncultivated person. People of both groups have moved to (the suburbs of) Baghdad during the past decades (cf. Miller 2007, p. 14) and thus more contact situations with the city dwellers have arisen. Most of my participants mentioned that they only used those terms for people who lacked education, good taste in clothing, and had a more conservative lifestyle, but not to people who came from rural areas but were educated and have adapted to the city lifestyle. Only participant E, who was born in Maysan herself, said she was proud of the *mḡāḡdān* heritage of her people and considered it an important part of Iraqi culture.

3.4. Case Study Ahvaz, Khuzestan: Urbanization of a Rural *ḡolot* Dialect

The city of Ahvaz, capital of the south-western Iranian province of Khuzestan, witnessed rapid urbanization and population growth in the twentieth century. In the nineteenth century, Ahvaz was no more than a village and in the early twentieth century it still had less than 50,000 inhabitants (see Oppenheim 1967, p. 22, fn. 1). During the Iran–Iraq War (1980–1988) numerous houses were destroyed (especially in the southern Khuzestani cities of Muḡammara/Khorramshahr and Abadan) and many families were forced to flee their hometowns. Many of the Khuzestani Arab war refugees left the province of Khuzestan altogether or went to comparably safer cities, such as Ahvaz. For this reason, during that time, the city of Ahvaz witnessed an immense population growth. According to Nejatian (2015), the number of inhabitants in Ahvaz grew from 334,399 in 1976 to 724,653 in 1991, and to 1,112,021 in 2011.

To look at the loss of rural features in Ahvaz might give us a hint as to what *ḡolot* features are highly marked and first to be given up in arising urban contexts. Arising urban contexts are here defined as contexts which permit contact with other *ḡolot* dialects (urban and rural) but not necessarily have an old sedentary or *ḡoltu* substratum.

Of the (possibly) rural features discussed in Section 3.1, Feature 1 (affrication of **q*), Feature 2 (raising of **a* in pre-tonic open syllables), and Feature 5 (gender distinction in the plural), are commonly found among all speakers of Ahvazi Arabic.

The remaining features discussed in Section 3.1 are not found in Ahvazi Arabic or are in the process of being dropped:

Feature 3: In Ahvazi Arabic as well as most present-day Khuzestani Arabic dialects, the use of forms that show the typical *ḡahawa*-type resyllabification is limited to certain frozen examples—e.g., *ḡahali* ‘my family’ and *xadār* ~ *ḡaxadār* ‘green’—and not productive, e.g., Ahvazi Arabic *ḡahwa* (not *ḡhawwa*) ‘coffee’, and *naḡya* (not *nḡaya*) ‘ewe’ (cf. Leitner 2020, p. 50 for more examples).

Feature 4: Ahvazi Arabic does not have forms that show the Bedouin-type resyllabification rule of CaCaC-v(C)-structures, e.g., Ahvazi Arabic *kəḡbat* ‘she wrote’, and *šəḡgat* ‘she hugged’. The Bedouin-type form is still typical of north-western Khuzestani towns and villages, such as Xafaḡīya (Pers. Susangerd) and Ḥuwayza.

Feature 6: Most speakers of Ahvazi Arabic do not show retention of the prefix vowel in Form V and Form VI verbs. As stated above (3.1.6), this feature is more typical of rural areas and smaller Khuzestani towns and villages such as Ḥamīdiyya.

Feature 7: In present-day Ahvazi, but not among all speakers, the rural form of the SG.M imperative of IIIw/y verbs is still found, i.e., the one lacking the final vowel—e.g., *ḡaməṣ* ‘go! (IMP SG.M)’, *ḡahəṣ* ‘speak! (IMP SG.M)’. However, the form ending on a vowel that is associated with urban contexts (cf. Ingham 2007, p. 577; 1973, p. 544)—e.g.,

ʔəmsi ‘go (IMP MSG)’—is also used in present-day Ahvazi Arabic. Thus, this rural feature is apparently still in the process of being dropped and substituted by the less marked urban forms.

Finally, present-day Ahvazi Arabic appears to use both typically *ʕarab*- and *ḥadar*-type words (cf. fn. 13 on these terms; Ingham 1973, p. 538). *ʕarab* words (associated with a rural lifestyle) include, e.g., (*le-*)*ḡād* ‘there’ (however, its *ḥadar* equivalent *hnāk* ‘there’ is equally attested in Ahvaz). *ḥadar*-type lexical items in Ahvazi Arabic are, for instance, *taʕadda* ‘to pass (i.e., go past something)’. Some lexemes that Ingham mentioned appear to be given up completely, e.g., the word for ‘meal’ today is neither *marag* (which Ingham lists as *ḥadar*) nor *ydām* (which Ingham lists as *ʕarab*), but *ʔakəl*; and the most commonly heard word for ‘mirror’ in present-day Ahvazi Arabic is neither *mnədra* (*ḥadar*) nor *mrāya* (*ʕarab*), but *məšifa* (PL *məšāwəf*) and *məšaffa* (PL *məšāffāt*). In turn, sometimes items from both types are used, for example, ‘to look (at)’ may be expressed by *əšfəba* (*ḥadar*), *bāwa* *ʕ* (*ʕarab*), or *ʕāyan* (*ʕarab*) in present-day Ahvazi Arabic. Even though this distinction cannot be equated with the urban–rural distinction but is rather connected to occupational differences (cf. fn. 13), these processes found in the lexical domain support the assumption that Ahvazi Arabic has been subject to linguistic levelling tendencies since the times of Ingham’s fieldwork in 1969 and 1971. The result of this development is a dialect which does not clearly correspond to one of these sociolinguistic categories anymore and may rather be considered as of mixed typology. The reasons for the linguistic leveling and mixture of dialectal features observed for Ahvazi Arabic lie mainly in the rapid demographic changes that this city has witnessed during recent years. Its fast growth during and after the Iran–Iraq War, especially, has allowed for much (linguistic) contact among people of different geographic origins within Khuzestan and southern Iraq, calling for linguistic accommodation and triggering leveling processes (cf. Ech-charfi 2020, pp. 70–71, 75 on leveling tendencies in other new cities, such as Amman and Casablanca).

The fact that of all possible rural features discussed in Section 3.1, Feature 2 (raising of **a* in pre-tonic open syllables) and Feature 5 (gender distinction in the plural) are not modified or dropped in Ahvazi Arabic is partly paralleled by the results of the sociolinguistic interview (cf. Section 3.3), according to which these features are not marked as rural features. The third feature that is not dropped in Ahvazi Arabic, Feature 1 (affrication of **q*), shows that affrication of **q* is apparently less marked in Khuzestan. This might be related to the fact that the urban category is newer in the Khuzestani society and the dichotomy of urban vs. rural features not as strong or long established as in Iraq, where urban centers have already existed for hundreds of years.

All features that are not found (anymore) in Ahvazi Arabic as described above—except for Feature 6—are the same features that were perceived as highly marked by the participants of the sociolinguistic interview.

4. Discussion

In the light of the scarcity of linguistic descriptions of *gələt* dialects in general (of both urban and rural contexts) and the fact that some of the descriptions date back more than 100 years (e.g., Meißner 1903 on Kwayriš), we must be careful when drawing general conclusions about the present-day classification of this dialect group. The following interpretation of our results is therefore to be understood as tentative and as re-opening the floor for debating this issue.

4.1. Historical and Modern MBA and the Quest for Urban *gələt*

The analysis in Section 3.2 lets us safely conclude that many features of MBA must not necessarily be explained as a consequence of the *qəltu* substrate but can also be interpreted as consequences of the urbanization of *gələt* dialects.

The fact that some features which have been explained as old urban and sedentary features are nowadays also found in rural Bedouin dialects that have not (or not likely)

been directly in contact with *qəltu* supports the theory that features often spread from prestigious urban centers to rural areas.

For example, the emphatic imperative prefix *d-* (cf. Palva 2009, pp. 22, 35), the use of the future marker *rāh* and the indefinite article *fard* are nowadays attested for several urban and rural *gələt* dialects alike and (partly) also for the Šāwi dialects and dialects of Khorasan. In these dialects, the existence of these features cannot be explained via a *qəltu* substrate, which they do not have. More likely, they have spread—probably at a very early stage—from the longer established sedentary *qəltu* to the later incoming Bedouin *gələt* dialects. Nowadays, we may consider them areal features general to the southern Mesopotamian area and beyond.

Instead of explaining the absence of certain features of MBA—the lack of the resyllabication rules and the feminine plural forms—via the *qəltu* substrate, we suggest, rather, that these developments be seen as consequences of the urbanization of *gələt* speakers and the subsequent loss of marked features (cf. subsequent subsections). Some but not all of these marked features are mostly absent in modern urban contexts, e.g., Baghdad and Ahvaz, and are still strongly associated by speakers themselves with Bedouin and rural-type dialects (at least for Iraq). Notably, this is not the case with the gender distinction in the plural, which is gaining ground in modern MBA and is retained in the modern city of Ahvaz.

As for the question why—at least at the time of Haim Blanc’s descriptions of Baghdadi Arabic—the only urban or urbanized *gələt* dialects are found in Lower (and none in Upper) Iraq, we must keep in mind that most towns of Lower Iraq were built (or re-populated) not before the eighteenth and nineteenth centuries after the massive depopulation of this area following the Mongol invasions (cf. Blanc 1964, p. 170, fn. 189 and the references mentioned there). This implies that we are mostly not dealing with longstanding sedentary populations (that were later Bedouinized or marginalized by the Bedouin immigrants, as in the case of MBA²¹) but much later (nineteenth, early twentieth century) sedentarized and urbanized Bedouin populations. In contrast, towns in Upper Iraq have had a more continuous (*qəltu*-speaking) population. In addition to that, the topography of Upper Iraq has been described as “more conducive to polarization between sedentary and non-sedentary life”, as the steppes allowed for grazing only, while the fixed banks of the river courses are well suited for permanent sedentarism (cf. Blanc 1964, pp. 170–71). This situation has definitely changed as more and more Bedouins have given up the nomadic lifestyle. The question of whether or not and to what degree we nowadays find urbanizing tendencies in the *gələt* dialects of Upper Iraq remains to be answered in future studies as we still lack the data needed for such an analysis.

4.2. Who Speaks Urban *gələt*?

Traditionally, Muslim Baghdad Arabic (MBA) has been considered the main representative of an urban *gələt* type besides the Arabic dialect of Basra (Blanc 1964, p. 165).²² However, considering the unique linguistic history of MBA, a *gələt* dialect that has a *qəltu* substratum (Palva 2009 argues for a mixed *qəltu/gələt* character of MBA; cf. Section 3.2 for a discussion of this description), it is questionable whether it is a good reference point for a general description of an urban *gələt* type. The specific character of MBA has indeed arisen in an urban context (though the sense of ‘urbanization’ here is not a socio-economic one), however, the contact between (Bedouin) *gələt* and (Sedentary) *qəltu* speakers has also shaped the linguistic profile of this dialect.

On the urban character of MBA and the urban–rural split among the *gələt* dialects, Blanc tentatively stated that “... [MBA] is closest to the urban dialects on which some data are available (Basra, Qal’at Şāleh) so that one dimly foresees a possible classification of urban vs. rural *gələt* dialects, as yet not solidly established” (Blanc 1964, p. 165).

Mahdi writes in the introduction to his thesis on Basra Arabic that “in studying BA, I found no justification for dividing BA [Basra Arabic] into two groups, i.e., urban and rural, [...] The linguistic boundaries between the town and the surrounding countryside are

simply not apparent [...] mainly because the town society is rural in origin and the towns depend basically on the surrounding villages and countryside for filling the needs and manpower. Those who live in the town are most of them originally villagers or cultivators who moved to the town for various reasons" (Mahdi 1985, p. XV). Basra has been almost completely destroyed in the fourteenth century and was subsequently moved to its modern location at al-ʿUbulla (Pellat and Longrigg 2012; Oppenheim 1952, p. 178) and was subject to massive immigration by speakers from rural areas of Lower Iraq. A major difference between Basra Arabic and MBA is that the former has no *qəltu* substrate.

In a similar vein, Bruce Ingham states in the introduction to his book *Arabian Diversions* (Ingham 1997, pp. ix–x) that geographically and demographically all Khuzestani Arabic dialects are really rural in character, for which reason he prefers to use the terms *ḥadār* vs. *ʿarab* instead of urban vs. rural for the subclassification of these dialects (cf. fn. 13 below).

Of course, since Ingham's descriptions (based on his fieldwork carried out in the 1960s and 1970s), new urban centers have developed (e.g., Ahvaz, cf. Section 3.4 below) and others have considerably grown. In both Iraq and Khuzestan, the past decades have witnessed massive population movements (to a considerable part caused the Iran–Iraq war in the 1980s), from rural to—already existing as well as newly arising—urban areas. The question of possible linguistic effects of these urbanizing tendencies among the *gəlat* dialects will be discussed in Section 4.4.

Thus, we can see that even though the term 'urban' is clearly associated with cities, this does not necessarily mean that the speakers of a city speak an urban variety and even less does it mean that their families are of an old, established urban background.

4.3. About Rural *gəlat* and the Markedness of Rural Features

The analysis of the possible (phonological and morphological) rural features in Section 3.1 yields no unanimously clear picture regarding their distribution. As we can see in Table 1, only Features 6 and 7 are clearly absent in the dialects of the urban centers Basra and Baghdad.²³ However, even regarding these features the picture is not clear when it comes to the supposedly rural dialects of Khuzestan and Kwayriš. The distribution of all other features (1–5), doesn't show a clear-cut distinction between rural and urban contexts either.

While Feature 1 (affrication of **q*) is found in all dialects analyzed but MBA, Feature 2 is found in all dialects analyzed. Even though Features 3 and 4 are virtually absent from the dialects of Basra and Baghdad, they have also been dropped or are in the process of being dropped from what have usually been described as rural *gəlat* dialects. Feature 5 stands out by being found in all varieties analyzed (urban and rural), albeit in MBA it can often be substituted by masculine plural forms.

This picture is in many points corroborated by the results of the sociolinguistic survey presented in Section 3.3 and the theoretical assumption that "overt stigmatization attached to certain rural features seems to be the main reason why these features are reduced during the accommodation process to urban speech" (Abu-Haidar 1988, p. 76).

How markedness may lead to the loss of features, how non-markedness can foster retention of features, and which features are relevant in our context, will be discussed in the following Section 4.4.

4.4. Linguistic Consequences of Urbanization

In the following, some general tendencies found in urban(izing) contexts will be discussed. 'General' is to be understood in the geographical context of the Mesopotamian area only, as supra-regional tendencies in urban and rural dialects of Arabic are difficult to find, and the analysis of such supra-regional tendencies was not within the scope of this paper. One possible supra-regional tendency might be the perception of urban dialects or urban features as effeminate vs. rural dialects as masculine (cf. Section 3.3; Miller 2007, p. 13; Ech-charfi 2020, p. 75). In general, urbanizing contexts are characterized by an

increase of contact (be it urban–rural, or rural–rural of two different rural backgrounds) and leveling tendencies; they also often go hand in hand with complexity reduction.

4.4.1. Loss of Features

It is often the case that a language becomes structurally simpler in contexts permitting a high degree of language (or variety) contact leading to linguistic leveling (cf. Kerswill 2013, p. 521 and the references mentioned there). Based on this rule, we would expect categories such as gender distinction in the plural for verbs and pronouns to be given up in urban contexts.

Like most other *gələt* dialects, including those spoken in urban contexts such as Basra, all Khuzestani Arabic dialects have retained gender distinction in the plural. This is a feature that in general is often given up in urban-type dialects, even when of Bedouin origin (cf. Procházka 2014, p. 129). The use of feminine plural forms has even seemed to gain ground in the city of Baghdad. This fact was confirmed in the sociolinguistic survey conducted for this paper, which also showed that it is not indexed with rural speech or overtly marked as rural. This explains why it can be easily reintroduced in urban contexts, such as Baghdad, with the immigration of rural speakers, and why it is not rapidly given up in newer urban contexts, such as Ahvaz. Finally, the fact that feminine plural forms exist in the literary language might raise its prestige or at least increases the exposure of speakers to this feature, which in turn lends to the readiness to adopt a feature (Baghdad), as well as its resistance to be dropped (Ahvaz).

Due to increased literacy in cities, features that deviate from the literary language in a way that is clearly perceivable for speakers are often highly marked. Examples of such marked features (be it for their deviation from the literary language or for other reasons) are, e.g., the *gahawa*-syndrome, the Bedouin-type resyllabification rule, and SG.M imperative forms of final weak roots lacking a final vowel. At least for our geographic context, our small-scale sociolinguistic survey shows that these are among the features most readily given up in urban contexts and most strongly marked as rural among urban speakers. Inversely, the sociolinguistic survey also shows that those features that are apparently not indexed with rurality—especially Features 2 and 5 as described in Section 3.1—tend to be retained in both urban and rural contexts.

The current developments in Ahvazi Arabic (cf. Section 3.4) underline the proposed tendencies found in urban contexts regarding the loss of highly marked rural features.

4.4.2. Innovations

Reevaluating Heikki Palva’s diachronic discussion of urban *qəltu* features as found in MBA has shown that urban contexts, especially those where contact with other languages plays an important role, may encourage the development of new linguistic categories. Examples of this are the emergence of an indefinite article (which in general is absent in Arabic), as described in Section 3.2.1, and of the emphatic imperative particle *d-* (provided we consider it a Persian loan; cf. Section 3.2.4).

The use of time (future), aspect (progressive) and indefiniteness markers may have arisen historically among sedentary communities but were at some point later adopted by Bedouin speakers. Most likely, this happened in urban contexts, which strongly facilitate contact situations.

5. Conclusions and Outlook

The distinction between urban and rural must not be given up completely for the *gələt* dialect group. Trying to write about this distinction with respect to present-day dialects, we should, however, shift the focus onto synchronic sociolinguistic differences that we can observe among speakers who live in cities vs. speakers who live on the countryside. Following this, we should focus on the study of current linguistic trends as observed in arising urban contexts, such as Ahvaz.

Approaching the question of the urban–rural split among the *gələt* dialects, it also appears highly necessary to ask what is subjectively perceived as rural or urban speech by native speakers, rather than imposing what we think to be rural or urban based on the few dialectal descriptions we have from over a hundred years ago. By this critical reevaluation, the author by no means wants to lower the value of these seminal contributions made by Arabic dialectologists, but merely proposes a new way of approaching the classification of the *gələt* dialects.

One major factor that should be considered in any new attempt at classifying the *gələt* dialects by using the urban–rural dichotomy is the different nature of older or longer established urban communities, as found in Baghdad, and the communities of new cities, such as Ahvaz. While in the former, longer established urban communities witnessed rural immigration, in the new urban contexts most inhabitants—or at least their (grand)parents—still are of rural origin themselves. This means that there is no established urban community which would define the linguistically urban character of this city in the first place. Rather, it is the “cohabitation of different ethnic groups” (Ech-charfi 2020, p. 72) and groups of different geographical origin that is shaping the new urban profiles. This should not mean, however, that the linguistic profile of such urban spaces that *do* have established urban communities and that are facing rural immigration is only defined by the linguistic traits of the old urban community. In such scenarios we often observe that new urban sociolinguistic identities are coined by combining both old urban features and part of the rural linguistic heritage (cf. Ech-charfi 2020, pp. 72, 75–76 on a similar observation in Rabat, Amman and Casablanca Arabic). At least for modern MBA, the following citation from Ech-charfi applies “New urban identities are constructed linguistically by combining traditional urban and rural variants while rural stereotypes serve as the background against which urban identities are defined” (Ech-charfi 2020, p. 76).

Thus, in no scenario can we speak of a homogenous urban group that is clearly distinct from the rural population, as such groups are partly (Baghdad) or completely (Ahvaz) descended from rural populations themselves (cf. 4.1. and Mahdi 1985, p. XV, who states that all inhabitants of Basra are rural in origin). Rather, we can only try and capture linguistic trends found in urban contexts by observing which features are most readily modified or dropped and which adopted, and by asking speakers what features are perceived as rural. By this we can get to define the (socio-)linguistic profiles of the (modern-day!) urban and rural societies in Iraq and Khuzestan.

Importantly, the definition of the ‘default’ urban *gələt* type should not be limited to the scenario of MBA, a Bedouin type *gələt* dialect with a sedentary type *qəltu* substratum, but should be extended to include *gələt* dialects spoken in the context of newly arising urban spaces. Of course, the different nature of these two urban scenarios must be borne in mind.

In addition to the urban–rural distinction as treated in this paper, the *gələt* dialects can be divided by geographic aspects, e.g., into a southern and a northern group. For this question, the reader is referred to Hassan (2020, 2021), who discerns two geographic subgroups of Iraqi Arabic: *šrūgi* (south of Baghdad) and *non-šrūgi* (north of Baghdad) dialects, corresponding roughly to the Shiite and the Sunni groups of *gələt* speakers in Iraq, respectively (cf. fn. 20 on the derogatory nature of these terms).

The classification of the *gələt* dialects is still far from being solidly established, a situation which primarily results from the scarcity of data available on this dialect group.

We do hope, however, that this modest contribution has brought forth aspects of this classification hitherto not considered and presented some already considered aspects in a new light. Hopefully, this study will motivate other researchers to continue research on the classification, the historical development, and the modern urbanizing tendencies of this still under-researched dialect group.

One major desideratum in the investigation of the urban–rural split in the *gələt* dialects is a large-scale sociolinguistic survey showing how linguistic variables are perceived in terms of prestige, markedness, and other sociological factors, such as masculinity

vs. femininity, in various varieties of the *gələt*-speaking area. Ideally, this large-scale sociolinguistic survey would include additional variables: on the one hand, new variables from the domains of syntax and lexicon (e.g., *lē-gād* ‘there’ described as a typical rural *gələt* feature by Fischer and Jastrow 1980, p. 151), which could not be treated within the scope of the present study; and on the other hand, phonological features like the realization of *ǧ as [q] and the use of *tafxīm* ‘emphasis’ that were mentioned as rural features by the participants of the small-scale sociolinguistic survey conducted for this study.

Funding: Open Access Funding by the University of Vienna.

Institutional Review Board Statement: The only part of this study that included ethically relevant human interaction were the sociolinguistic interviews. Since it was not possible due to the research agenda to anonymize gender, age, and locations, this was only but thoroughly done for names. All subjects were informed that their anonymity is assured, why the research was being conducted, and how their data was going to be used. All subjects gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki. Approval by an Ethics Committee was not deemed necessary as this is a non-interventional study, all conventions in terms of data usage were upheld, and there are no perceivable risks for the participants due to the conduction of the study or the publication of the material.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Both publicly available and private datasets were analyzed in this study. The publicly available datasets can be found in the references cited in this article. The private dataset stems from the author’s own fieldwork conducted in Khuzestan in 2016 and is available on request from the corresponding author.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

DIM	Diminutive
EMP	Emphatic marker
F	Feminine
HORT	Hortative
IMP	Imperative
IPFV	Imperfective
M	Masculine
MBA	Muslim Baghdad Arabic
OA	Old Arabic
PL	Plural
SG	Singular

Notes

- ¹ I would like to express my gratefulness to my dear friends and colleagues Stephan Procházka and Ana Iriarte Díez as well as the reviewers for their valuable thoughts and critical remarks on draft versions of this article.
- ² Following Haim Blanc’s classification of Mesopotamian Arabic dialects into *gələt*- and *qəltu*-type dialects. These terms are the 1SG PFV verb forms for ‘to say’ (cf. Blanc 1964, pp. 7–8) which indicate certain phonological and morphological characteristics of these dialect groups.
- ³ Compare, e.g., “rural *g*” (Palva 2009, p. 35) and “the voiced reflex of OA *q* is the most exclusive Bedouin feature” (*op.cit.*: 24).
- ⁴ Historically, probably all *gələt* dialects outside Arabia were Bedouin or Bedouinized dialects (cf. Blanc 1964, pp. 167–68).
- ⁵ The term ‘Šāwi-type Bedouin Arabic’ is used here to refer to a bundle of closely related dialects spoken by semi-nomads in various regions of the Fertile Crescent. Typologically similar dialects are found in many rural parts of Iraq, which is why the Šāwi and the rural Iraqi *gələt*-type dialects are often grouped together as ‘Syro-Mesopotamian (fringe) dialects’ or pre-ʿAnazī dialects (Palva 2006, p. 606).
- ⁶ At the time Haim Blanc wrote his book on the Arabic dialects of Baghdad, the main scientific data available on *gələt* dialects were limited to (Meißner 1903; Weissbach 1968) on Kwayriš, his own data from one informant from the Musayyib district, and (Van Wagoner 1944) as well as his own data on the Arabic spoken in the Amarah district.

- 7 The sources used for Table 1 are: (Leitner 2020) for Khuzestani Arabic; (Meißner 1903) and (Denz 1971, which is based on Meißner 1903 and Weissbach 1908) for Kwayriš, (Salonen 1980) for al-Shirqat, (Seeger 2013, 2002; Volkan Bozkurt, pers. comm.) for Khorasan, (Behnstedt 1997; Bettini 2006; Jastrow 1996; Fischer and Jastrow 1980; Procházka 2003, 2018a; 2018b; Younes and Herin n.d., EALL online) for Šawi Arabic, and (Mahdi 1985) for Basra, (Leitner et al. 2021; Blanc 1964) for Muslim Baghdad Arabic (MBA); as well as (Palva 2009) and (Hanitsch 2019, pp. 266–71).
- 8 The sources are in fact contradictory on this question: while Mahdi (1985, pp. 94–106, 152–55) provides feminine plural forms for verbs and pronouns, Ingham (1982, p. 38) states that gender distinction in the plural has been given up in urban centers, such as Basra.
- 9 It is important to note that MBA shows interdialectal variation due to the subsequent immigration of rural people to the city; cf. Abu-Haidar (1988) on a number of phonological differences.
- 10 Blanc (1964, p. 204) reports that, even though the town was historically rather stable in comparison to most other towns in Lower Iraq, the population of Hilla has been Bedouin since its foundation; cf. Oppenheim 1952, pp. 185, 189, who writes that Hilla was deserted after the second Mongol invasion in the fourteenth century.
- 11 In several dialects, the raised vowel was elided subsequently, e.g., Khuzestani Arabic (ə)mrākəb ‘boats’ < *mərākəb < OA *marākib¹⁴.
- 12 Meißner (1903, p. XLI) also adds that the urban structure would be CvCCvC (*kitbet*).
- 13 Ingham (1976, p. 64; 1997, p. x) discerns three main socio-economical groups in Khuzestan: the *šarab*, the *ħaḍār*, and the marsh Arabs. While the term *šarab* denotes a group of larger territorially organized tribes, who live—sometimes as semi-nomads—away from the river in the plain (*bādiya*) and are involved in occupations such as cereal, rice, and date cultivation, sheep herding, and water buffalo breeding, the *ħaḍār* group are riverine palm-cultivating Arabs of mixed tribal descent, who live along the banks of the Shatt al-Arab and the lower parts of the river Karun. Ingham notes that while the *šarab* dialect shows “considerably more resemblance to the dialects of Arabia”, the *ħaḍār* was “more strictly Mesopotamian” (Ingham 1997, p. ix). Cf. Ingham (2009) on some ‘fringe’ Bedouin dialects in Kuwait and north west of Nasiriyah, which share features with both the southern Mesopotamian *gəlot* group, e.g., affrication of OA *k > č and *q > ġ, and the northern Najdi dialects.
- 14 Cf. Ingham (1976, p. 74), who contrasts ‘nomadic’ forms like *tahāča* and ‘sedentary’ forms like *thāča*; elsewhere, Ingham describes the prefix sequence CCv- in such structures, e.g., *nta š ašša*, as rural (Ingham 1973, p. 541).
- 15 In Khuzestani Arabic, the active participle form *nāyəl* *nāyħa* is also used to express future intent (*āna nāyħa (a) sawwō-lak* ‘I will make (for) you...’). We also find Khuzestani Arabic sentences with future reference that do not feature any future particle. In such cases, future reference is usually indicated by an IPFV verb or an active participle together with a temporal adverb like ‘tomorrow’ or ‘soon’.
- 16 There is one instance of *rāħ* and two instances of *rāyehl* used as future markers in Salonen’s texts: *rāħ teħergu* ‘it will burn him’ (Salonen 1980, pp. 13 and 33, Text 3, sentence 8), *rāyehl yiṭla š* ‘he will come up’, *rāyehl yiğrag* ‘he will drown’ (Salonen 1980, pp. 14 and 34, Text 4, sentence 4).
- 17 Of course, this particle is less common in narratives than in conversations and the fact it does not appear in these texts from al-Shirqat might be attributed to the nature of the text genre and must not necessarily mean that it does not exist in this Arabic variety.
- 18 The grade of rurality of a certain feature was measured quantitatively by the number of its mentions by the interviewees as a *rifi* ‘rural’ feature. This means, if a feature reaches, for example, the number 3 on the ‘Rurality Scale’, three of the five interviewees have described this feature as typical of rural dialects.
- 19 Participant C, in contrast, stated that the form *xašš* was used in Baghdad as well, but only by elderly women.
- 20 Cf. Hassan (2020, 2021), who uses this term for dividing the Iraqi Arabic dialects into a northern and a southern group (the latter being associated with the term *šrūgi*), even though he acknowledges its pejorative and derogatory use (Hassan 2021, p. 52).
- 21 Cf. Blanc (1964, p. 170): “In the fourteenth century, the Baghdad Muslims were still speaking a *qəltu* type dialect . . .”.
- 22 Blanc also refers to the dialect of the town of Qal’at Saleh as a representative of the urban group. The only description of this dialect, however, is an unpublished dissertation (Van Wagoner 1944) not available to the author of this paper.
- 23 Although we did not find evidence for the use of Feature 6 in the one source that exists for al-Shirqat, we cannot rule out that it does not exist in this variety. It is, however, also absent in the modern dialect of the city of Ahvaz.

References

- Abd-el-Jawad, Hassan R. 1986. The Emergence of an Urban Dialect in the Jordanian Urban Centers. *International Journal of the Sociology of Language* 61: 53–63. [CrossRef]
- Abu-Haidar, Farida. 1988. Speech Variation in the Muslim Dialect of Baghdad: Urban vs. Rural. *Zeitschrift Für Arabische Linguistik* 19: 74–80.
- Abu-Haidar, Farida. 2002. Negation in Iraqi Arabic. In “Sprich Doch Mit Deinen Knechten Aramäisch, Wir Verstehen Es!”: 60 Beiträge Zur Semistik. *Festschrift Für Otto Jastrow Zum 60. Geburtstag*. Edited by Werner Arnold and Hartmut Bobzin. Wiesbaden: Harrassowitz, pp. 1–13.
- Behnstedt, Peter. 1997. *Sprachatlas von Syrien. I: Kartenband*. Semitica Viva 17. Wiesbaden: Harrassowitz.

- Behnstedt, Peter. 2000. *Sprachatlas von Syrien. II: Volkskundliche Texte*. Wiesbaden: Harrassowitz.
- Bettega, Simone, and Bettina Leitner. 2019. Agreement Patterns in Khuzestani Arabic. *Wiener Zeitschrift Für Die Kunde Des Morgenlandes* 109: 9–37.
- Bettini, Lidia. 2006. *Contes Féminins de La Haute Jézireh Syrienne. Matériaux Ethno-Linguistiques d'un Parler Nomade Oriental*. Florence: Dipartimento di Linguistica.
- Blanc, Haim. 1964. *Communal Dialects in Baghdad*. Cambridge: Harvard Univ. Press.
- Denz, Adolf, and Otto Edzard. 1966. Iraq-arabische Texte nach Tonbandaufnahmen aus al-Hilla, al-'Afač und al-Bašra. *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 116: 60–96.
- Denz, Adolf. 1971. *Die Verbalsyntax des neuarabischen Dialektes von Kwayriš (Irak): Mit einer einleitenden allgemeinen Tempus- und Aspektlehre*. Wiesbaden: Steiner.
- Ech-charfi, Ahmed. 2020. The Expression of Rural and Urban Identities in Arabic. In *The Routledge Handbook of Arabic and Identity*. Edited by Reem Bassiouney and Keith Walters. London: Routledge, pp. 65–78.
- Fischer, Wolfdietrich, and Otto Jastrow. 1980. *Handbuch Der Arabischen Dialekte*. Wiesbaden: Harrassowitz.
- Grigore, George. 2019. Deontic Modality in Baghdadi Arabic. In *Studies on Arabic Dialectology and Sociolinguistics: Proceedings of the 12th International Conference of AIDA Held in Marseille from May 30th to June 2nd 2017*. Edited by Catherine Miller, Alexandrine Barontini, Marie-Aimée Germanos, Jairo Guerrero and Christophe Pereira. Livres de l'IREMAM. Aix-en-Provence: IREMAM, Available online: <https://books.openedition.org/iremam/3917> (accessed on 24 January 2021).
- Häberl, Charles. 2019. Mandaic. In *The Semitic Languages*. Edited by John Huehnergard and Na'ama Pat-El. Milton: Routledge, pp. 679–710.
- Hanitsch, Melanie. 2019. *Verbalmodifikatoren in den arabischen Dialekten: Untersuchungen zur Evolution von Aspektsystemen*. Porta Linguarum Orientalium 27. Wiesbaden: Harrassowitz.
- Hassan, Qasim. 2016a. Concerning Some Negative Markers in South Iraqi Arabic. In *Arabic Varieties: Far and Wide. Proceedings of the 11th International Conference of AIDA—Bucharest, 2015*. Edited by George Grigore and Gabriel Bițună. Bucharest: Editura Universității din București, pp. 301–6.
- Hassan, Qasim. 2016b. The Grammaticalization of the Modal Particles in South Iraqi Arabic. *Romano-Arabica* 16: 45–55.
- Hassan, Qasim. 2017. On the Synonymous Repetition of Greeting Questions in South Iraqi Arabic: A Sociopragmatic Study. In *The International Scientific Conference: Oriental Studies—Yesterday, Today, Tomorrow/8th December 2016*. University of Belgrade, Department of Oriental Studies. Belgrad: Faculty of Philology, University of Belgrad, pp. 429–36. [CrossRef]
- Hassan, Qasim. 2020. Reconsidering the Lexical Features of the South-Mesopotamian Dialects. *Folia Orientalia* LVII: 183–93. [CrossRef]
- Hassan, Qasim. 2021. Phonological Evidence for the Division of the Gələt Dialects of Iraq into Šrūgi and Non-Šrūgi. *Kervan—International Journal of Afro-Asiatic Studies* 25: 51–61.
- Holes, Clive. 1995. Community, Dialect and Urbanization in the Arabic-Speaking Middle East. *Bulletin of the School of Oriental and African Studies* 58: 270–87. [CrossRef]
- Holes, Clive. 2001. *Dialect, Culture, and Society in Eastern Arabia. Vol. 1: Glossary*. Leiden: Brill.
- Holes, Clive. 2016. *Dialect, Culture, and Society in Eastern Arabia. Vol. 3: Phonology, Morphology, Syntax, Style*. Leiden and Boston: Brill.
- Ingham, Bruce. 1973. Urban and Rural Arabic in Khūzistān. *Bulletin of the School of Oriental and African Studies* 36: 523–53. [CrossRef]
- Ingham, Bruce. 1976. Regional and Social Factors in the Dialect Geography of Southern Iraq and Khuzistan. *Bulletin of the School of Oriental and African Studies* 39: 62–82. [CrossRef]
- Ingham, Bruce. 1982. *North East Arabian Dialects*. London: Kegan Paul.
- Ingham, Bruce. 1997. *Arabian Diversions: Studies on the Dialects of Arabia*. Reading: Ithaca Press.
- Ingham, Bruce. 2007. Khuzestan Arabic. In *Encyclopedia of Arabic Language and Linguistics*. Leiden: Brill, vol. 2, pp. 571–78.
- Ingham, Bruce. 2009. The Dialect of the Euphrates Bedouin, a Fringe Mesopotamian Dialect. In *Arabic Dialectology. In Honour of Clive Holes on the Occasion of His Sixtieth Birthday*. Edited by Enam Al-Wer and Rudolf de Jong. Leiden and Boston: Brill, pp. 99–108.
- Jastrow, Otto. 1996. The Shawi Dialects of the Middle Euphrates Valley. In *Encyclopaedia of Islam*. 376a-b. Leiden: Brill.
- Johnstone, Thomas M. 1967. *Eastern Arabian Dialect Studies*. London: Oxford University Press.
- Kerswill, Paul. 2013. Koineization. In *The Handbook of Language Variation and Change*. Edited by J. K. Chambers and Natalie Schilling. Oxford: John Wiley & Sons, Inc., pp. 519–36. [CrossRef]
- Leitner, Bettina, and Stephan Procházka. 2021. The Polyfunctional Lexeme /Fard/ in the Arabic Dialects of Iraq and Khuzestan: More than an Indefinite Article. *Brill's Journal of Afroasiatic Languages and Linguistics* 13: 1–44. [CrossRef]
- Leitner, Bettina, Fady German, and Stephan Procházka. 2021. *Lehrbuch des Irakisch-Arabischen: Praxisnaher Einstieg in den Dialekt von Bagdad*. Wiesbaden: Harrassowitz.
- Leitner, Bettina. 2019. Khuzestan Arabic and the Discourse Particle ča. In *Livres de l'IREMAM*. Edited by Catherine Miller, Alexandrine Barontini, Marie-Aimée Germanos, Jairo Guerrero and Christophe Pereira. Aix-en-Provence: IREMAM, Available online: <https://books.openedition.org/iremam/3939> (accessed on 24 January 2021).
- Leitner, Bettina. 2020. *The Arabic Dialect of Khuzestan (Southwest Iran): Phonology, Morphology and Texts*. Vienna: University of Vienna.
- Mahdi, Qasim R. 1985. *The Spoken Arabic of Bašra, Iraq: A Descriptive Study of Phonology, Morphology and Syntax*. Ph.D. dissertation, Exeter University, Exeter, UK.
- Meißner, Bruno. 1903. Neuarabische Geschichten Aus Dem Iraq. *Beiträge Zur Assyrologie Und Semitschen Sprachwissenschaft* 5: 203.

- Miller, Catherine. 2007. Arabic Urban Vernaculars: Development and Change. In *Arabic in the City: Issues in Dialect Contact and Language Variation*. Edited by Catherine Miller, Enam Al-Wer, Dominique Caubet and Janet C. E. Watson. London: Routledge, pp. 1–31.
- Nejatian, Mohammad Hossein. 2015. Ahvaz Iv. Population, 1956–2011. In *Encyclopaedia Iranica Online Edition*. Available online: <http://www.iranicaonline.org/articles/ahvaz-04-population> (accessed on 24 January 2021).
- Oppenheim, Max Freiherr von. 1952. *Die Beduinen. Band III: Die Beduinenstämme in Nord- Und Mittelarabien Und Im Irāk*. Wiesbaden: Harrassowitz.
- Oppenheim, Max Freiherr von. 1967. *Die Beduinen: 4,1: Die Arabischen Stämme in Chūzistān (Iran). Pariastämme in Arabien*. Leipzig: Harrassowitz.
- Palva, Heikki. 2006. Dialects: Classification. In *Encyclopedia of Arabic Language and Linguistics*. Edited by Kees Versteegh. Leiden: Brill, vol. 1, pp. 604–13.
- Palva, Heikki. 2009. From qəltu to gələt: Diachronic Notes on Linguistic Adaptation in Muslim Baghdad Arabic. In *Arabic Dialectology. In Honour of Clive Holes on the Occasion of His Sixtieth Birthday*. Edited by Rudolf de Jong and Enam Al-Wer. Leiden & Boston: Brill, pp. 17–40.
- Pellat, Charles, and Steven Helmsley Longrigg. 2012. Al-Bašra. In *Encyclopaedia of Islam*, 2nd ed. Edited by Peri Bearman, Thierry Bianquis, Clifford Edmund Bosworth, Emeri Johannes van Donzel and Wolfhart Heinrichs. Online. Leiden: Brill.
- Procházka, Stephan. 2003. The Bedouin Arabic Dialects of Urfa. In *AIDA: 5th Conference Proceedings, Association Internationale de Dialectologie Arabe (AIDA): Cádiz, September 2002*. Edited by Ignacio Ferrando and Juan José Sánchez Sandoval. Cadiz: Servicio de Publicaciones de la Universidad de Cádiz, pp. 75–88.
- Procházka, Stephan. 2014. Feminine and Masculine Plural Pronouns in Modern Arabic Dialects. In *From Tur Abdin to Hadramawt: Semitic Studies. Festschrift in Honour of Bo Isaksson on the Occasion of His Retirement*. Edited by Tel Davidovich, Ablahad Lahdo and Torkel Lindquist. Wiesbaden: Harrassowitz, pp. 129–48.
- Procházka, Stephan. 2018a. The Arabic Dialects of Eastern Anatolia. In *The Languages and Linguistics of Western Asia*. Edited by Geoffrey Haig and Geoffrey Khan. Berlin and Boston: De Gruyter Mouton, pp. 159–89.
- Procházka, Stephan. 2018b. The Northern Fertile Crescent. In *Arabic Historical Dialectology: Linguistic and Sociolinguistic Approaches*. Edited by Clive Holes. Oxford: Oxford University Press, pp. 257–92.
- Salonen, Erkki. 1980. *On the Arabic Dialect Spoken in Širqāt (Assur)*. Helsinki: Suomalainen Tiedeakatemia.
- Seeger, Ulrich. 2002. Zwei Texte Im Dialekt Der Araber von Chorasān. In *“Sprich Doch Mit Deinen Knechten Aramäisch, Wir Verstehen Es!” 60 Beiträge Zur Semitistik. Festschrift Für Otto Jastrow Zum 60. Geburtstag*. Edited by Werner Arnold and Hartmut Bobzin. Wiesbaden: Harrassowitz, pp. 629–46.
- Seeger, Ulrich. 2009. ‘Khalaf – Ein Arabisches Dorf in Khorasan’. In *Philologisches Und Historisches Zwischen Anatolien Und Sokotra: Analecta Semitica in Memoriam Alexander Sima*. Edited by Werner Arnold, Michael Jursa, Walter W. Müller and Stephan Procházka. Wiesbaden: Harrassowitz, pp. 307–17.
- Seeger, Ulrich. 2013. Zum Verhältnis Der Zentralasiatischen Arabischen Dialekte. In *Nicht Nur Mit Engelszungen. Beiträge Zur Semitischen Dialektologie: Festschrift Für Werner Arnold Zum 60. Geburtstag*. Edited by Renaud Kutu, Ulrich Seeger and Shabo Talay. Wiesbaden: Harrassowitz, pp. 313–22.
- Sharkawi, Muhammad al-. 2014. Urbanization and the Development of Gender in the Arabic Dialects. *Journal of Arabic and Islamic Studies* 14: 87–120. [CrossRef]
- Taine-Cheikh, Catherine. 2004. Le(s) futur(s) en arabe. Réflexions pour une typologie. *Estudios de Dialectología Norteafricana y Andalusi* 8: 215–38.
- Van Wagoner, Merrill Y. 1944. *A Grammar of Iraqi Arabic. Unpublished dissertation*. New Haven: Yale University.
- Watson, Janet C. E. 2011. Arabic Dialects (General Article). In *The Semitic Languages: An International Handbook*. Edited by Stefan Weninger. Berlin & Boston: De Gruyter Mouton, pp. 851–95.
- Weissbach, Franz Heinrich. 1968. *Beiträge Zur Kunde Des Irak-Arabischen*. Leipzig: J.C. Heinrichs'sche Buchhandlung.
- Younes, Igor, and Bruno Herin. n.d. ‘Šāwī Arabic’. In *Encyclopedia of Arabic Language and Linguistics*. Online Edition. Leiden: Brill.
- Younes, Igor. 2018. Raising and the Gahawa-Syndrome, between Inheritance and Innovation. *Zeitschrift Für Arabische Linguistik* 67: 5–11. [CrossRef]

Article

The Southern Moroccan Dialects and the Hilālī Category

Felipe Benjamin Francisco

Seminar für Semitistik und Arabistik, Freie Universität Berlin, 14195 Berlin, Germany;
fbenjamin@zedat.fu-berlin.de

Abstract: The aim of this paper is to review the classification of the southern Moroccan dialects, advancing on the general description of these varieties. Recent descriptive studies provided us with new sources on the linguistic reality of southern Morocco, shedding light on the status of dialects commonly classified as Bedouin or ‘Hilālī’ within the Maghrebi context. To do so, the paper highlights conservative and innovative features which characterize the dialects of the area, focusing mainly—but not exclusively—on the updated data for two distant localities in southern Morocco: Essaouira and its rural outskirts—the Chiadma territory (Aquerroud and Sidi Iṣḥāq)—and Tafilalt, in south-eastern Morocco. The southern dialects have been situated in an intermediary zone between pre-Hilālī and Hilālī categories for a long time. Discussing their situation may contribute to understanding what distinguishes them as a dialectal group and also the validity of the ‘Hilālī’ category in the Moroccan context.

Keywords: Arabic dialectology; Moroccan Arabic; Essaouira; Tafilalt; southern Morocco; Bedouin dialects

Citation: Francisco, Felipe Benjamin. 2021. The Southern Moroccan Dialects and the Hilālī Category. *Languages* 6: 192. <https://doi.org/10.3390/languages6040192>

Academic Editors: Simone Bettega and Roberta Morano

Received: 3 August 2021
Accepted: 10 November 2021
Published: 23 November 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Traditionally, dialectologists have divided the Maghrebi dialects into two categories—pre-Hilālī and Hilālī—within a diachronic perspective which associates linguistic features to the waves of Arabization in North Africa, from the works of W. Marçais, such as the seminal text *Comment l’Afrique du Nord a été arabisée* (Marçais [1938] 1961), to more recent scholarship (cf. Aguadé 2018). Based on these two types, Colin ([1937] 1945, 1986) proposed a sub-classification to Western Maghrebi dialects, or Moroccan dialects precisely, grouping them into: *parlers citadins*, *parlers montagnards*, *parlers bédouins* and *parlers juifs*.

Regarding the Hilālī-Bedouin type in Morocco, authors have attempted to tackle the problem of grouping different linguistic varieties under this category. Colin (1986, p. 1196) proposed that the Moroccan Bedouin dialects could be divided according to their levels of conservatism. That is the case of some dialects of the Sahara area—but not exclusively (e.g., Casablanca, Kampffmeyer 1912)—which retain features such as the realization [g] of *qāf and the maintenance of interdental (e.g., /d/ and /t/). The same aspect was observed by Lévy (1998, p. 19) who points out that Hilālī and Ma^ṣqili dialects found in the Atlantic plains are quite different from the Ma^ṣqili type in the Sahara (e.g., Ḥassāniyya). In agreement with this view, Heath (2002, p. 8) drew a distinction between Hilālī central type dialects and the Saharan ones, which—according to him—are restricted to southern oases and parts of the Atlantic plains in Morocco.

Regarding the Bedouin category in Morocco, Taine-Cheikh (2017) points out: “*la situation reste complexe à décrire pour les parlers qui ne sont ni pré-hilaliens ni du type ‘saharien’*” (p. 25). That is the case of the southern Moroccan dialects, for which the application of the Hilālī category remains doubtful, despite of the confirmation of the [g] realization and the loss of interdental, both commonly associated to it. In this manner, we pose the question of whether the findings on the southern dialects, and the revision of their classification, might contribute to shedding the light on the Hilālī category within the global linguistic reality of Morocco?

More recently, the endeavor of dialectologists for classifying the so-called Hilāli-Bedouin dialects has come out again after ‘Hilāli’ and ‘Ma^ṣqili’ terms were called into question. Benkato (2019) criticized that dialectologists erroneously linked Medieval historical facts—originally incorporated from Ibn Khaldun by French orientalist—with the modern linguistic reality of the Maghreb. He argued that there is a lack of evidence on the direct connection between Medieval tribes, taken as a “reliable unit of sociolinguistic analysis” (*sic*)—such as the Ma^ṣqil—and the Arabic dialects spoken nowadays in the region (p. 21). In this way, understanding the condition of southern Moroccan dialects might contribute to understand the validity of categories, such as ‘Hilāli’. Nevertheless, it is important to say that the link between historical factors and the current linguistic reality should never be totally discarded.

To explain the distribution of the “southern” linguistic features over this part of Morocco, I argue that they are associated not only to the process of Arabization of this area, but also to modern historical factors, including the trans-Saharan trade route—connecting the Sahel to the Atlantic—and its effects on the populational movements on this area until the nineteenth century. In this way, I also try to explain the reason why distant localities in the south share common linguistic features and how their nature impacts on the validity of Hilāli-Bedouin category for classifying southern Moroccan dialects. This may be a complementary explanation to defining the origin of common features found in southern dialects.

In this context, the aim of this paper is to review the classification of the southern Moroccan Arabic dialects, highlighting some of the features which might single out these varieties. To do so, the study relied on the recent collection of dialectological data for the Atlantic strip, in Essaouira region (Francisco 2019a, 2019b, 2022), and south-eastern Morocco, represented by Tafilalt (Heath 2002; Behnstedt 2004)—without ignoring previous descriptive studies on other varieties of the region.

2. Materials and Methods

2.1. Descriptive Studies on the Southern Varieties

According to the traditional classification of Moroccan dialects, the southern dialects belong to the Hilāli or Bedouin type, given that they spread from the Atlantic plains—in the area of Mogador—to the eastern part of the country, the Moulouya basin and the Moroccan Sahara (Colin [1937] 1945, p. 230). Part of these dialects is commonly distinguished from the “truly” Bedouin dialects—Saharan or Ma^ṣqili type—, given that their classification was thought based on maintenance and loss of conservative features (Colin 1986, pp. 1195–96). Most of the dialects from the Atlantic strip to Tafilalt lost traditional conservative features (e.g., the interdental), due to different degrees of Arabization of the Berber tribes (Heath 2002, pp. 8–9).

Taine-Cheikh (2017, pp. 25–26) proposes the category ‘*parlers “hilaliens” du Sud marocain*’ to set together the dialects spoken from the Atlantic coast as far as the Algerian border. Her description relied on studies for the dialects of Skoura, Sous and Essaouira, which exhibit the realization of *qāf as [g] and the loss of interdental. In fact, the realization of *qāf in southern Moroccan dialects remains a complex issue to the classification of these varieties under the label ‘Hilāli’, given that both *g* (Bedouin) and *q* (sedentary) alternate—as phonemes and allophones—, varying lexically. Moreover, the voiced *g* continues to be very usual in these varieties (Heath 2002, p. 9).

Taine-Cheikh (2017, p. 26) considers as general common features for the southern varieties the following:

- *ũ* as variant for /ə/ in velar and uvular consonantal contexts;
- labializations or the occurrence of ultra-short *ũ*;
- reduction of diphthongs into *ī* and *ū*;
- no gemination in syllabic structures (*‘ressautées’*);
- *-ti* (2s. perf.) for both masculine and feminine, except for Skoura;
- gender distinction in the 2s. imperf. and imperative;

- no distinction of gender for the 2s. clitic pronoun;
- 3s.m. suffix *-u* or *-h* and *-ah*;
- the suffixes *-āw* or *-īw* (pl. imperf.) for defective verbs;
- passive prefix: *t-*, *tt-* and *nt-*;
- preverbs: *ta-* and/or *ka-*;
- genitive particles: *ntāʿ* and/or *dyāl* (~ *d-*);
- future particle: *ḡādi*.

Given the new data for the dialects of the region, this paper gives special attention to the dialects encountered in two geographical extremities of the south not considered in the study mentioned above. For Tafilalt¹: Rissani and Erfoud (Heath 2002); Īgli, Erfoud, Ma^ʿdid, z-Zrīgāt and z-Zāwya ž-Zdīda (Behnstedt 2004) and ʿṛab Ṣabbāḥ (Behnstedt n.d.²); and Judeo-Arabic of Ksar es-Souk and Rich (Heath and Bar-Asher 1982). For Essaouira (Mogador), I considered the Muslim and Jewish³ dialects of the city, and also the variety of Chiadma territory (Aquermoud and Sīdi Īshāq)⁴ in the rural surroundings of the city (Francisco 2019a, 2019b, 2022). Essaouira data, specifically, may prove to be a valuable source to understanding the linguistic reality of southern Morocco, due to the nature of the settlement in the city, which attracted speakers from different parts of the south since its foundation in the second half of the eighteenth century, as we may see in the next session.

Apparently, the dialects of both southern regions share most of the common features indicated by Taine-Cheikh previously, with a few exceptions, as we may see in Section 3, which consequently bring implications to the classification of these varieties. That may be explained due to historical facts related to the Arab settlement in these localities and also the lasting linguistic contact between the southern dialects.

2.2. Historical and Linguistic Connections in Southern Morocco

The History of population settling in southern Morocco may explain the linguistic proximity between south-eastern Morocco—the Tafilalt area—and the Essaouira region on the Atlantic strip, which includes the Chiadma territory. Concerning the Arabization process of both areas, well known historical sources indicate that these territories were occupied by Ma^ʿqil tribe members at some point, after the beginning of the second wave of Arabization in the Maghreb with the Banū Hilāl invasions in the XI century. Modern sources continued to narrate the movements of these groups in southern Morocco, region which became gradually more connected by centuries-long trade routes.

It is well known that the Ma^ʿqil tribes entered the Maghreb accompanying the Banū Hilāl (XI–XIII) and settled mainly on the outskirts of the latter’s territory, specially the Sous and the region corresponding to current Mauritania. In eastern Morocco, the D̄wī Maṣūr⁵ settled along the Moulouya River and the deserts of Tafilalt, from Taourirt—in northern Morocco—to the Draa Valley, as far as Sijilmassa (Ibn Khaldun 2011, p. 2361). La Chapelle (1930, p. 89) claims that they remained in Tafilalt until the nineteenth century living among other tribes under Berber rule.

In southwestern Morocco, the settlement of groups of Ma^ʿqil origin happened more lately, during the Saadian rule (XVI–XVII), groups such ṣ-Ṣabbānāt⁶ and lə-Mnābha emigrated from the Sous and established themselves on the territories of ʿAbda and Z^ʿir, on the Atlantic plains in central Morocco, but also at the surroundings of Marrakesh⁷ (Colin [1937] 1945, p. 224). Moreover, in 1765, the foundation of the port town of Essaouira (Mogador), on the limits between Chiadma (Arabic speaking) and Haha (Tachelhit speaking) territories, attracted peoples, not only from these two neighboring territories, but also from distinct parts of the Sous, and among them ṣ-Ṣabbānāt and lə-Mnābha once again were attracted to the Atlantic plains taking part in the formation of Mogador’s population (al-Kānūnī 1932; ar-Ragrāgī 1935; as-Sūsī [1966] 2005; aṣ-Ṣiddīqī 1969).

Later in the nineteenth century, the flow of the trade of the Trans-Saharan route shifted westward to the Atlantic coast, due to the important role of the port of Essaouira⁸ for the international trade. The city became connected with southern Moroccan cities by routes

with Akka and Guelmim⁹. In this way, Essaouira was connected indirectly to Tafilalt and West African regions¹⁰. In the second half of the nineteenth century (1860–70s), the greater portion of the West Africa trade comes into Morocco via Tindouf and Sous to Mogador (Dunn 1971, pp. 278–80). Caravans were moving between Essaouira, Tafilalt and sub-Saharan regions, connecting their populations who probably used Arabic as a lingua franca in commercial relations¹¹. In the nineteenth, Essaouira used to receive annually one or two caravans composed of thousand camels and smaller caravans as well, trading export commodities—such as gum and ostrich feathers—but also gold and slaves for the local market (Dunn 1971, p. 271). As Lévy (1998, p. 13) points out, certainly linguistic exchanges took place due to contact of the caravans with local populations while passing by rural markets on their routes across the south.

The trans-Saharan slave trade was also very active by that time. El-Hamel (2013) shows that there was a continuous import of thousands of slaves into Morocco by well-established trade routes (Tindouf, Ijil and Twat). He estimates that, by the end of the nineteenth century, the total black population was of half a million people (pp. 245–46). According to him, many were sold in the markets of Fez, Mogador and Marrakesh (p. 251), and besides that, a part of the enslaved people from sub-Saharan Africa could be found in the sugar refineries near Essaouira, in Haha and Shishawa territories (p. 152).

It cannot be ignored that Essaouira and Tafilalt were connected, despite of the distance, by the caravans moving between the two regions due to the trans-Saharan trade. And this fact may be important to explain certain singular linguistic features in both localities.

3. Southern Moroccan Features (Results)

The following selected features may help understanding more deeply what set the southern dialects together or apart, according to the innovative or conservative nature of these traits.

3.1. Retention of Diphthongs: /āw/ and /āy/

In general, the southern varieties present the contraction of diphthongs /āw/ > /ū/ and /āy/ > /ī/, as expected for Hilali or central-type varieties, described by Heath (2002, p. 9), such as for: **Essaouira** *līl* “night”, *sūq* “market”, *ṣūf* “summer” (Francisco 2019b, p. 143), **Chiadma** *l-yūma* “today”, *zīt* “oil” (Francisco 2019a, p. 5); **Skoura** *bīd* “eggs”, *lūz* “almonds”, *ṣīb* “pocket” (Aguadé and Elyacoubi 1995, p. 25); **Sous** *īm* “water spring”, *fūq* “over, on” (Destaing 1937, p. 27). Sometimes diphthongs are accepted as variants in pharyngeal contexts, e.g., **Essaouira** *īṭb* ~ *īṭyb* “shame”, *ṣūf* ~ *ṣūwf* “wool” (Francisco 2019b, p. 77).

In other southern varieties, the predominant feature above occurs along with the retention of diphthongs which can be realized phonetically as the vowels [o:] and [e:]—also represented by *ō* and *ē*—as found in: **Tafilalt** *lōn* “color”, *lēl* ~ *līl* “night”, *rmāyṭu* “I threw it” (Behnstedt n.d., Notes sur le parler “bedouin” des ṣṣab Sobbāḥ, p. 3), *fōk* “above, on” (Heath and Bar-Asher 1982, p. 46). For the rural area of Essaouira: **Chiadma** *zōz*¹² “plough drawn by oxen” (Francisco 2019b, p. 79), *ḥāndu kōma* “he has got a stack (of money)” (Francisco 2019a, p. 6), *nsāyṭ* “I forgot”, *bḡāyṭ* “I want”, *bnāyṭna* “we built”, *dzādāyṭna* “we were born”, *ḡāyṭna* “we came” (Francisco 2019b, p. 108). The diphthong in defective verbs is also preserved in Ḥassānīya: *ḡāyṭna* “we came” (Cohen 1963, p. 110), *ṣrāyṭna* “we bought” (ibid, p. 102). The same feature is found in Saharan type dialects in neighbouring areas, as in southwest Algeria: **Saoura** *wēn* ~ *weyn* “where”, *ṣōt* ~ *ṣowt* “voice” (Grand’Henry 1979, p. 215); **Mzāb** *nsēit* “I forgot” (Grand’Henry 1976, p. 24), *ṣrīna*~*ṣrēina* “we bought” (ibid, p. 26).

Nevertheless, the retention is not attested in most of the southern dialects analyzed here. Even in the few dialects that it is attested, Chiadma and Tafilalt, the feature still occurs along with the reduction of diphthongs in /ī/ and /ū/. The maintenance of diphthongs and allophones *ō* and *ē* in both localities does not appear to be a result of Trans-Saharan trade connections, given that the feature is absent from the urban Essaouira dialect. Therefore,

the feature seems to be of Saharan origin, as attested by its occurrence in Saharan dialects, and might be an evidence of the nature of the Arab settlement in both localities.

3.2. The Verbal Suffix *-āt* (3f. perf.)

In Morocco, the suffix *-āt* is a variant of *-ət* in the 3f. perf. conjugation of trilateral strong verbs. The suffix appears to be predominant in the majority of southern dialects: **Essaouira** *šərbāt* “she drank” (Francisco 2019b, p. 96), **Essaouira (J)** *okʕt* “happened” (3.f.) (Lévy 2009, p. 367), *sərbūt* “she drank” (Francisco 2022, in press)¹³; **Tafilalt** *šərbāt* “she drank” (Behnstedt 2004, p. 55); **Sous** *xərzāt* “she went out”, *hərbāt* “she ran away” (Destaing 1937, p. 7).

The absence of *-ət* in strong trilateral verbs seem to be characteristic of the southern dialects. Exceptions are found in **Skoura** *ktəbt* “she wrote” (Aguadé and Elyaccoubi 1995, p. 151) and **Sous (Houwara)**: *šərbət* “she drank” (Socin and Stumme 1894, p. 22). The same is found in the old data for Essaouira: *žəbrət* “she found” (Socin 1893, p. 164), *dərbət* “she hit” (p. 180)¹⁴. More recently, the suffix is seldomly attested, except for a unique occurrence in the Jewish dialect of **Essaouira (J)** *xərzət* “she went out” (Lévy 2009, p. 368). The suffix is also found in **Tafilalt (J)** (Heath and Bar-Asher 1982, p. 64).

The prevailing opinion is that the occurrence of *-āt* (3f. perf.), in Moroccan dialects, is due to the analogy with weak verbs (e.g., *mšāt* “she went”) (Heath 2002, p. 223; Aguadé 2008, p. 291). Regarding the diffusion of the feature, urban centers—such as Casablanca, Meknes and Marrakesh—may play an important role in it. For instance, Aguadé interprets the occurrence of *-āt* in **Settat** *xādmāt* “she worked” (Aguadé 2013, p. 4) as a convergence towards the Casablanca variety. In my opinion, regarding southern Morocco, Marrakesh may also have diffused the suffix in the region, given that it is well attested in the city, e.g., **Marrakesh** *səmʕāt* “she listened” (Sánchez 2014, p. 121).

It is not clear whether *-ət* spread earlier than *-āt*. On one hand, the neighboring Saharan type varieties do not exhibit the ending *-āt*, as one can attest in Ḥassāniya: *ktəbt* “she wrote” (Cohen 1963, p. 91); or in Algerian Sahara in the Mzāb region: *kətbət* “she wrote” (Grand’Henry 1976, p. 43). On the other hand, the ending *-āt* is found in other parts of the Maghreb: **Eastern Libya** *ik’tib-at* “she wrote” (Owens 1984, p. 105). It is also found in the Maghreb neighbouring areas, like in West Sudanic¹⁵ *katabat* “she wrote” (Owens and Hassan 2009, p. 713).

The fact is that *-āt*—in the Moroccan case—must consist of a conservative feature just like in other parts of the Arabic speaking world. The ending *-at*, with a short vowel, in strong verbs, is found in many eastern dialects, not only inside the Arabian Peninsula, but also outside of it in Bedouin-type dialects (Gaash 2013, p. 49).

3.3. The Clitic *-ki* (2s.f.)

The occurrence of the clitic *-ki* (2s.f.) is very common in semiverbs all over Morocco. Heath (2002, p. 242) confirms it, but he did not analyze the use of *-ki* in the possessive function.

In southern Morocco, apparently, we find it with possessive and object functions in two regions exclusively: **Tafilalt**: *šəftki* “I saw you”, *gānki* “he told you”, *rəzniki* “your feet” (Behnstedt 2004, p. 56), *ḥḥāki* “your father”, *mšāki* “with you”, *wuldkī* “your son”, *dārki* “your house”, *šandki* “you have”, *šuftki* “I saw you”, *gallki* “he told you” (Behnstedt n.d., Notes sur le parler “bedouin” des Ḥrab Səbbāh, p. 6); and **Essaouira** *xūki* “your brother”, *ḥḥāki* “your father”, *šāndki* *ūliyydāt?* “do you have children?”, *āna kənt hna šāndki* “I was here at your place”, *dyāki* “yours”, *nsūwəwəlki* “I will ask you”, *yəqdəḥ i šāwūnki* “he will be able to help you”, *hāki ktābki* “here is your book” (Francisco 2019b, p. 164); and **Chiadma** *iḥḥāki* “your father”, *xūki* “your brother”, *ḥḥāki* “your mother”. The suffix *-ki* with possessive and object functions seems not to be attested in other southern localities though, like in the vernaculars of Marrakesh and Sous, for instance.

Regarding the origin of this feature, I claimed previously (cf. Francisco 2019b, p. 164) that the occurrence of *-ki* in the possessive function—in Essaouira and Tafilalt—resulted

probably from a morphological analogy with semiverbs, given that it is absent in Saharan type varieties such as Ḥassāniya *nsāk anti* “he forgot you” (2f.) (Cohen 1963, p. 151) or the Mzāb region variety in Algeria, which do not distinguish the gender, using only *-ak/-k* (Grand’Henry 1976, p. 67). However, the historical links between southern Morocco and the West African region could provide us with a new hypothesis. It is not unrealistic to think that *-ki* entered Morocco, and remained restricted to the south, due to the slave trade connecting Tafilalt to the sub-Saharan region, given that enslaved boys and girls were brought to Morocco from parts of Western Africa, such as Nigeria and Chad (El-Hamel 2013, pp. 130–31). Moreover, the clitic is found in West Sudanic *buyūt-ki* “your houses (f.)” (Owens and Hassan 2009, p. 712), being clearly a retention. Such hypothesis would deserve a more in-depth discussion though.

Regarding the exclusive occurrence of *-ki* in Essaouira and Tafilalt, it could be explained by the linguistic link resulted from the caravans of Arabic speakers which connected both localities.

3.4. The Suffix *-u* (pl. imperf.) for Defective Verbs in *-i*

This is a retention (Cl. Ar. **yamšū-na > yəmš-u*) in defective verbs ending in *-i* well attested in Saharan type dialects, e.g., in Ḥassāniya *nəšru* “we buy” (Cohen 1963, p. 103), Saoura *imšu* “they go” (Grand’Henry 1979, p. 220), Mzāb *yəmšu* “they go” (Grand’Henry 1976, p. 49). In southern Morocco, we find it in: **Sous** (Houwara) *ka-yībku* “they cry” (Socin and Stumme 1894, p. 16), along with **Sous** *ibnīw* “they build” (Destaing 1937, p. 39); **Essaouira** *ta-yīšru* “they buy”, **Chiadma**: *ka-yīžru* “they run” (Francisco 2019b, p. 103) along with the variant *-īw* as well. Apparently, the suffix *-u* is not attested in other southern localities, such as for **Skoura** *təmšiw* “you go” (Aguadé and Elyacoubi 1995, p. 48) and **Tafilalt** *ḡādyīn nəmšīw* “we will go” (Behnstedt 2004, p. 56), *tānu izīw* “they were coming” (Heath and Bar-Asher 1982, p. 74).

In the south, this feature appears to be evidently of Saharan origin, but restricted to Sous and Essaouira, probably due to the settlement of Saharan dialects speaking tribes as mentioned in the Section 2.

3.5. Future Preverb *ba-bḡa*

The use of the perf. verb *ba-bḡa* “to want” with imperf. verbs to express the future consists of a structure predominant all over the south of Morocco, from the Atlantic strip to Tafilalt (Heath 2002, p. 217). It is found in **Marrakesh** *ba-yīžri* “he will run” (Sánchez 2014, p. 182) and **Skoura** *bīt nšūfhuīm* “I will see them” (Aguadé and Elyacoubi 1995, p. 86). In Essaouira, the verb *ba-bḡa* developed into an invariant particle *b(ə)-* to express future: *fīn bə-ṭkūn ḡōdda?* “where will you be tomorrow?”, *b-nāhḍru dāba ṡāl-lə-bḡūṛ* “now we are going to talk about the sea”, *ḡōdda b-yīšru l-ḡwāyōž* “tomorrow they will buy clothes” (Francisco 2019b, p. 140). A similar particle occurs in **Sous** (Houwara) *‘bunnīmšī* “I will go” (Socin and Stumme 1894, p. 54).

The particle is a Hilāli feature attested in other parts of the Maghreb as well. The preverb *ba-* is found in the Sahara, in Algerian southwest: **Saoura** *ba-iṣarrəs* “he will get married” (Grand’Henry 1979, p. 224); and the particle *b-* is also used to express the future in Bedouin Libyan dialects, e.g., **Al-Khums** *b-yṡāwəd* “he will repeat” (Benmoftah and Pereira 2017, p. 317).

Sánchez (2014, p. 183) mentions the occurrence of the future particle *bā-* in some dialects of Yemen, for which he suggests a common etymology with the verb *ba* “to want” in Marrakesh. Despite the fact that Ma^tqil tribes are assumed to have come from Yemen, the verb *bḡa* is employed in Ḥassāniya to express intention only, not expressing the future (Taine-Cheikh 2004, p. 225). However, Ḥassāniya does apply the same structure above with the verb *idōr* “to want” as an auxiliary to express the future: *idōr iṡḡ* “he is going to fall” (*ibid*, p. 224). This structure to express future is an innovation common to Hilāli-Bedouin dialects in southern Morocco, but also in the Sahara and other parts of the Maghreb.

4. Discussion

Taine-Cheikh (2017, p. 26) proposed that the southern Moroccan dialects were similar to the Casablanca dialect according to a list of common features presented above. On one hand, the southern dialects exhibit traditional Hilālī features, such as the realization [g] for *q, but, on the other hand, they also exhibit the loss of interdental, a distinctive trait of Bedouin dialects. Considering the difficulty to classify these dialects as Hilālī, the selection of features by Taine-Cheikh (cf. Section 2.1) attempted to draw a group of southern dialects, but it was not able to single these varieties out, distinguishing them from other varieties in Morocco. To give a few examples, the following features cited above (cf. Section 2.1) are quite spread all over the country: the future particle *ḡādi*; the no gender differentiation for the 2s. clitic *-k*; and the ending *-ṭw* for defective verbs (3pl.imperf).

More recent data, especially from Essaouira and Tafilalt, shed a new light on the reality of southern dialects. Comparing them with the well-known dialects of the region—Sous and Skoura—and also with the dialects of the Saharan neighbouring areas—Ḥassāniya and the dialects of Algerian regions of Saoura and Mzāb—demonstrated that the varieties of the southern region are not so homogenous as we thought previously. The findings revealed a Bedouin color for the southern area due to the occurrence of retentions and innovations, some of them comprehending Saharan variants.

The Saharan traits attested in the southern varieties co-occur with the variants presented by Taine-Cheikh (cf. Section 2.1). They are the following retentions: maintenance of diphthongs /*āw*/, /*āy*/ in pharyngealized and plain contexts, sometimes realized as [ō] and [ē]; and the suffix *-u* (pl. imperf.) for defective verbs. Both features are spread in Bedouin dialects beyond the Moroccan borders. Nevertheless, in the south, the dialect of Skoura is an exception, not exhibiting these features. Curiously, regarding the suffix *-u*, there is no register of it for Tafilalt.

Concerning the conservative trait *-āt* (3.f. perf.), the variant is well spread all over Morocco along with *-at*, however, it proved to be dominant in the south, where the latter is seldomly registered nowadays, except for Skoura. Despite of the absence of *-āt*, in Ḥassāniya and other Saharan varieties of the region, it does seem to have a Bedouin origin, as the suffix is attested in other parts of the Maghreb (e.g., Eastern Libya) and also in neighbouring varieties, like West Sudanic. And even though the feature occurs in other parts of Morocco, it can be considered characteristic of southern dialects.

Another representative case is the verb *ba-bḡa* “to want” and the particles derived from it (*b-*, *ba-*) attached to imperfective verbs to express future. This feature consists of the single variant connecting all the southern dialects apparently, distinguishing them from the dialects of northern Morocco. Since the feature has reflexes on Ḥassāniya and is attested in other varieties across the Maghreb, being registered even in Yemen, this innovation may be an evidence of a common Bedouin or Hilālī origin for the southern dialects, along with traditional traits such as [g] for *qāf and the occurrence of interdental in a previous stage.

Within southern dialects, the clitic *-ki* (2f.), in the possessive and object functions, builds a bridge between Essaouira and Tafilalt, not being attested in any other dialects across the whole country until now. The retention of the clitic restricted to these southern extremities, West and East, could be explained by the contact performed by caravans in the Trans-Saharan trade routes linking distant parts of the south with the Sahara and the Sub-Saharan Africa. This shared past in the south must have played an important role to the diffusion of the other features mentioned here as well.

5. Conclusions

Southern Moroccan varieties proved to exhibit more Bedouin features than previously thought. Colin (1986) tackled the Bedouin-Hilālī issue, pointing out that some dialects could not be classified fully as Bedouin-type for not maintaining conservative features. That was the case of the southern Moroccan dialects, which were thought not to exhibit Saharan-type features. Here, linguistic findings revealed the opposite though.

In agreement with the traditional dialectology scholarship, the occurrence of Saharan-type retentions in southern sites could be explained by the settlement of the Ma^ṣqil in the area as confirmed in historical sources. Nevertheless, it is important to note that the Ma^ṣqil tribe settled in Tafilalt and in the Atlantic strip with a gap of centuries. Furthermore, there is no linguistic evidence that the Ma^ṣqil tribe presented a dialectal unity—as the tribe was subdivided into distinct groups (*buṭūn*)—and, if so, that it preserved this unity through centuries until modern times. Therefore, populational movements caused by multiple factors, such as the Trans Saharan trade, famine crisis and epidemics which swept southern Morocco, should not be ignored when trying to explain the spread of certain variants.

The only innovation feature which seems to gather the bunch of southern dialects, linking them with other Bedouin varieties in Algerian Sahara and Libya, is the future construction *ba~bġa* “to want” > *b-/ba-* + imperf. Apparently, the origin of this feature goes back to the Yemeni particle *bā-*, nevertheless, it is not attested in Ḥassānīya, though the dialect is associated to the Ma^ṣqil tribe, supposedly from Yemen. Despite of that, Ḥassānīya attests a similar structure with the same function (e.g., *idōr itih*, cf. Section 3.5), which may have developed after the former with *ba~bġa* > *b-/ba-*. Moreover, the occurrence of the structure (*b-/ba-* + imperf.) in other parts of the Maghreb may warn us that not all the features in current Saharan varieties, especially Ḥassānīya, may be representative of the “purest” Bedouin-type, or Hilālī, neither in Morocco nor in the Maghreb.

The occurrence of certain conservative features in certain parts of Morocco and the Maghreb also corroborates the previous argument. That is the case of the retention of the ending *-āt* (3f. perf.) and the conservative clitic *-ki* (2f.) both absent from Saharan-type varieties.

The current linguistic situation of southern Moroccan dialects highlights the limitation of the ‘Hilālī’ category, including the ‘Ma^ṣqilī’ label, to deal with dialectal layers within Moroccan varieties. The difficulty in applying this category is due to the co-occurrence of features, or variants, of distinct origins in the current dialects. Trying to determine dialectal groups based on the Arabization waves does not prove to be sufficient anymore, given that speakers of distinct Arabic dialects have been in movement and contact for centuries in the area.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

Notes

¹ In this paper, ‘Tafilalt’ refers to south-eastern Morocco in general. The reader can access the specific locality of the linguistic findings according to the cited author.

² Behnstedt (n.d.) Notes sur le parler “bédouin” des Ṣāb Səbbāḥ (Tafilalt/Maroc).

³ The Jewish dialects of Essaouira and Tafilalt are indicated by: (J).

⁴ I collected the data for the rural area of Essaouira in Aquermoud and Sīdi Iṣḥāq during linguistic fieldwork in the city of Essaouira in the years of 2016, 2017 and 2018. Some inedit data for Essaouira and the Chiadma territory are published in this paper.

⁵ Ibn Khaldun (2011) claims that the Ma^ṣqil tribe were Bedouin from Yemen, being divided into three groups (*buṭūn*): Ḍwi Ṣūbidullāh, at-Ta^ṣāliba, Ḍwi Maṣūr e Ḍwi Ḥassān (pp. 2363–70).

⁶ Dialectal pronunciation of the name. In *Kitāb al-Ṣībar* (Ibn Khaldun 2011) the name is registered *aṣ-Ṣūbbānāt*.

⁷ On the reasons these tribes moved to this area, see: az-Zayyānī (1886); ad-Ḍu^ṣayyif (1986) and al-^ṢUfrānī (1998).

⁸ See Schroeter (1988) on the role of Essaouira for international trade in the nineteenth century Morocco.

⁹ Guelmim was one of the redistribution and places of concentration of caravans such as Aboudam, Ghadames and Assiout (Miège 1981, p. 96).

- 10 In pre-modern times, the Sahel, or the sub-Saharan West Africa was designated in Arabic under the term *bilād as-Sūdān*. The recent Historical publications on the Trans-Saharan trade, cited in this paper, maintained the use of the term “Sudan”.
- 11 For the possible use of Arabic as a lingua franca in the trade with West Africa, see [Levtzion \(2000, p. 64\)](#) and [Bouwman \(2008, p. 135\)](#).
- 12 The Chiadma realization of this vowel is very similar to Ḥassānīya as described by [Cohen \(1963, p. 53\)](#) who sees it as a long vowel followed by an “*appendice labial ou palatal*”, which he represents by $\text{ɒ}^h, \text{ɔ}^w$.
- 13 *-īl* is the Jewish variant for *-āt* of strong regular verbs, in Essaouira ([Francisco 2022, in press](#)). [Heath \(2002, p. 224\)](#) posed two hypothesis on the origin of the suffix *-īl*: (i) a mutation of **-at* or (ii) a lengthening of **-at* to keep the 3. perf. f. suffix distinct from the first and second persons suffixes.
- 14 We should be cautious about Socin’s data, given that he seems to have transcribed the oral texts based on a manuscript ([Socin 1893, p. 157](#)) read out loud by a single informant. He explains that the speech of the informant presented both citadin and Bedouin features (p. 155). Despite of that, we do find many common features between his data and recent collected ones (cf. [Francisco 2019b](#)).
- 15 By ‘West Sudanic’, [Owens and Hassan \(2009\)](#) refer to a dilect region which comprehends the Arabic of Nigeria, Chad and Cameroon.

References

- ad-Ḍu^ʿayyif, Muḥammad. 1986. *Tārīx ad-Ḍu^ʿayyif: Tārīx ad-dawla as-sa^ʿīda li-Muḥammad ad-Ḍu^ʿayyif ar-ribāfī (1165-1233 H.)*. Rabat: Dār Al-Ma^ʿtūrāt.
- Aguadé, Jordi. 2008. *Morocco. EALL*. Leiden and Boston: Brill, vol. 3, pp. 287–97.
- Aguadé, Jordi. 2013. Zum arabischen Dialekt von Settat (Marokko). In *Nicht nur mit Engelszungen: Beiträge zur semitischen Dialektologie: Festschrift für Werner Arnold zum 60. Geburtstag*. Edited by Renaud Kutry, Ulrich Seeger and Shabo Talay. Wiesbaden: Harrassowitz, pp. 1–6.
- Aguadé, Jordi. 2018. The Maghrebi dialects of Arabic. In *Arabic Historical Dialectology: Linguistic and Sociolinguistic Approaches*. Edited by Clive Holes. New York: Oxford University Press, pp. 29–63.
- Aguadé, Jordi, and Mohammad Elyaacoubi. 1995. *El Dialecto Árabe de Skūra (Marruecos)*. Madrid: C.S.I.C.
- al-Kānūnī, Muḥammad bin ^ʿAḥmad al-^ʿAbdī. 1932. *ʿĀsft wa-mā ʿīlayhi qadīman wa-ḥadītan*. Cairo: Maṭba^ʿat Muṣṭafā Muḥammad.
- al-^ʿUfrānī, Muḥammad aṣ-Ṣaḡīr. 1998. *Nuzhat al-ḥādī bi-ʿaxbār mulūk al-qarn al-ḥādī*. Casablanca: Maṭba^ʿat an-naḡāḥ al-ḡadīda.
- ar-Raḡrāḡī, ^ʿAḥmad bin al-Ḥāḡḡ. 1935. *Aṣ-Ṣumūs al-munīra fī ʿaxbār madīnat aṣ-Ṣuwayra*. Rabat: Al-Maṭba^ʿa al-Waṭāniyya.
- aṣ-Ṣiddīqī, Muḥammad bin Sa^ʿīd. 1969. *ʿIqāḍ as-sarīra li-tārīx aṣ-Ṣuwayra*. Casablanca: Dār al-Kitāb.
- as-Sūsī, Muḥammad al-Muxtār. 2005. *ʿĪlīḡ, qadīman wa-ḥadītan*. Ar-Ribāt: Al-Maṭba^ʿa al-Malikīya. First published 1966.
- az-Zayyānī, ^ʿAbū l-Qāsim bin ^ʿAḥmad. 1886. *at-Turḡumān al-mu^ʿarib ^ʿan duwal al-Maṣriq wa-l-Maḡrib*. Paris: Imprimerie Nationale.
- Behnstedt, Peter. 2004. Von an-ʿĀṣor (Al-Qaṣr) nach Īgni (Īgli): Ein Vorbericht zu einigen arabischen Dialekten der Provinz ar-Raṣīdiyya (Marokko). In *Approaches to Arabic Dialects: A Collection of Articles Presented to Manfred Woidich on the Occasion of His Sixtieth Birthday*. Edited by Martine Haak, Rudolf De Jong and Kees Versteegh. Leiden and Boston: Brill, pp. 47–65.
- Behnstedt, Peter. n.d. Notes sur le parler “bedouin” des Ṣrab Səbbāḥ (Tafilalt/Maroc).
- Benkato, Adam. 2019. From Medieval Tribes to Modern Dialects: On the Afterlives of Colonial Knowledge in Arabic Dialectology. *Philological Encounters* 4: 2–25. [[CrossRef](#)]
- Benmoftah, Najah, and Christophe Pereira. 2017. Preliminary Remarks on the Arabic spoken in Al-Khums (Libya). In *Tunisian and Libyan Arabic Dialects: Common Trends, Recent Developments, Diachronic Aspects*. Edited by Veronika Ritt-Benmimou. Zaragoza: Instituto de Estudios Islámicos y del Oriente Próximo, pp. 301–25.
- Bouwman, Dinie. 2008. *Mali. EALL*. Leiden and Boston: Brill, vol. 3, pp. 135–41.
- Cohen, David. 1963. *Le Dialecte Arabe Ḥassānīya De Mauritanie (parler de la Gobla)*. Paris: Librairie C. Klincksieck.
- Colin, Georges. 1945. L’arabe. In *Initiation au Maroc*, 2nd ed. Paris: Vanoest (Les Éditions d’Art et d’Histoire), pp. 219–47. First published 1937.
- Colin, Georges. 1986. Al-Maḡhrib (VII. Aperçu Linguistique). *EI* 5: 1193–98.
- Destaing, Edmond. 1937. *Textes Arabes en parler des Chleuhs du Sous (Maroc)*. Paris: Imprimerie Nationale.
- Dunn, Ross R. 1971. The trade of Tafilalt: Commercial Change in Southeast Morocco on the Eve of the Protectorate. *African Historical Studies* 4: 271–304. [[CrossRef](#)]
- El-Hamel, Chouki. 2013. *Black Morocco: A History of Slavery, Race and Islam*. New York: Cambridge University Press.
- Francisco, Felipe Benjamin. 2019a. Preliminary Notes on the Arabic Dialect of the Chiadma (North of Essaouira). In *Studies on Arabic Dialectology and Sociolinguistics, Proceedings of the 12th International Conference of AIDA, Marseille, France, 30 May–2 June 2017*. Edited by Catherine Miller, Alexandrine Barontini, Marie-Aimée Germanos, Jairo Guerrero and Christophe Pereira. Aix-en-Provence: IREMAM. [[CrossRef](#)]
- Francisco, Felipe Benjamin. 2019b. O dialeto árabe de Essaouira: Documentação e Descrição de uma Variedade do sul do Marrocos. Ph.D. dissertation, Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, Brazil. Available online: <https://teses.usp.br/teses/disponiveis/8/8159/tde-29102019-180034/pt-br.php> (accessed on 5 June 2021).

- Francisco, Felipe Benjamin. 2022. The Judeo-Arabic of Essaouira revisited. In *Semitic Dialects and Dialectology: Fieldwork—Community—Change*. Edited by Maciej Klimiuk. Heidelberg: Heidelberg University Publishing, in press.
- Gaash, Amir. 2013. The Verbal and Nominal Feminine Endings -at and -it in Neo-Arabic. *Zeitschrift für arabische Linguistik* 57: 48–69.
- Grand'Henry, Jacques. 1976. *Les parlers arabes de la région du Mzâb (Sahara algérien)*. Leiden: Brill.
- Grand'Henry, Jacques. 1979. Le parler arabe de la Saoura (Sud-ouest algérien). *Arabica* 26: 213–28. [CrossRef]
- Heath, Jeffrey. 2002. *Jewish and Muslim Dialects of Moroccan Arabic*. London: RoutledgeCurzon.
- Heath, Jeffrey, and Moshe Bar-Asher. 1982. A Judeo-Arabic Dialect of Tafilalt (Southeastern Morocco). *ZAL* 9: 32–78.
- Ibn Khaldun, ⁵Abd ar-Rahmān bin Muḥammad. 2011. *Tārīḫ Ibn Khaldūn al-musammā bi-kitāb al-⁵ībar wa-dīwān al-mubtada⁷ wa-l-xabar ft ⁷ayyām al-⁵arab wa-l-⁵aḡam wa-l-barbar wa man ⁵aṣarahum min dawī aṣ-ṣulṭān al-²akbar*. Beirut: Dar Ibn Hazim, vol. III.
- Kampffmeyer, Georg. 1912. *Marokkanisch-Arabische Gespräche, im Dialekt von Casablanca*. Berlin: Druck und Verlag von Georg Reimer.
- La Chapelle, Frederic (de). 1930. Esquisse d'une histoire du Sahara occidental. *Hespéris* XI: 35–95.
- Levtzion, Nehemia. 2000. Islam in the Bilad al-Sudan to 1800. In *The History of Islam in Africa*. Edited by Nehemia Levtzion and Randall L. Pouwels. Oxford: Ohio University Press, pp. 63–91.
- Lévy, Simon. 1998. Problématique Historique du Processus d'arabisation au Maroc: Pour une histoire linguistique du Maroc. In *Peuplement et arabisation au Maghreb Occidental: Dialectologie et Histoire*. Edited by Jordi Agudé, Patrice Cressier and Ángeles Vicente. Madrid and Zaragoza: Casa de Velázquez, Universidad de Zaragoza, pp. 11–26.
- Lévy, Simon. 2009. *Parlers Arabes des Juifs du Maroc: Histoire, Sociolinguistique et Géographie Dialectale*. 'Estudios de Dialectología Árabe' 3. Zaragoza: Instituto de Estudios Islámicos y del Oriente Próximo.
- Marçais, William. 1961. 'Comment l'Afrique du Nord a été arabisée'. In *Articles et Conférences*. Paris: Adrien-Maisonneuve, pp. 171–92. First published 1938.
- Miège, Jean-Louis. 1981. Le commerce transsaharien au XIX^e siècle. *Revue de l'Occident musulman et de la Méditerranée* 32: 93–119. [CrossRef]
- Owens, Jonathan. 1984. *A Short Reference Grammar of Eastern Libyan Arabic*. Wiesbaden: Harrassowitz.
- Owens, Jonathan, and Jidda Hassan. 2009. West Sudanic. *EALL* 4: 708–18.
- Sánchez, Pablo. 2014. *El árabe Vernáculo de Marrakech. Análisis Lingüístico de un Corpus Representativo*. 'Estudios de Dialectología Árabe' 8. Zaragoza: Prensas de la Universidad de Zaragoza.
- Schroeter, Daniel. 1988. *Merchants of Essaouira: Urban Society and Imperialism in Southwestern Morocco, 1844–1886*. Cambridge: Cambridge University Press.
- Socin, Albert. 1893. 'Zum arabischen Dialekt von Marokko'. *Abhandlungen der philologisch-historischen Classe der Königlich-Sächsischen Gesellschaft der Wissenschaften* 14: 151–203.
- Socin, Albert, and Hans Stumme. 1894. Der arabische Dialekt der Houwāra des Wād Sūs in Marokko. In *Abhandlungen der philologisch-historischen Classe der Königlich-Sächsischen Gesellschaft der Wissenschaften* 15. Leipzig: Bei S. Hirzel.
- Taine-Cheikh, Catherine. 2004. Le(s) futur(s) en arabe: Reflexions pour une typologie. *EDNA* 8: 215–38.
- Taine-Cheikh, Catherine. 2017. La classification des parlers bédouins du Maghreb: Revisiter le classement traditionnel. In *Tunisian and Libyan Arabic Dialects: Common Trends, Recent Developments, Diachronic Aspects*. Edited by Veronika Ritt-Benmimoun. Zaragoza: Instituto de Estudios Islámicos y del Oriente Próximo, pp. 15–42.

Article

Auditory and Acoustic Evidence for Palatalization of the Nasal Consonant in Cairene Arabic

Navdeep Sokhey

Department of Modern and Classical Languages and Literatures, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA; navsokhey@vt.edu

Abstract: This paper introduces the palatalized nasal [nʲ] as an allophonic realization of coronal /n/ in Cairene Arabic. The palatalized variants of the phonemes previously described in acoustic and sociolinguistic terms include the alveolar stops [t, d] and their pharyngealized counterparts [tʰ, dʰ], which can be palatalized preceding the high, front vowel [i:]. While previous studies have anecdotally noted that the coronal nasal /n/ can undergo palatalization in the same environment, this variant has not been systematically investigated. Focusing on syllable-final /-ni:/ segments, I first use auditory measures to show that the palatalized variant occurs with some regularity (~50%) in the read speech of seven speakers of Cairene Arabic. Then, I provide acoustic evidence that this perceived difference significantly correlates with the difference in F2 values taken from the onset and midpoint of the vowel following the nasal consonant. There is also evidence of a lexical effect, such that borrowings exhibit less palatalization than non-borrowings. This study contributes data for the unexamined Cairene nasal and supports the likelihood of palatalization of coronals at the typological level.

Keywords: palatalization; nasal; Cairene Arabic; sociophonetics; acoustic phonetics

Citation: Sokhey, Navdeep. 2021. Auditory and Acoustic Evidence for Palatalization of the Nasal Consonant in Cairene Arabic. *Languages* 6: 190. <https://doi.org/10.3390/languages6040190>

Academic Editors: Simone Bettega and Roberta Morano

Received: 9 July 2021
Accepted: 5 November 2021
Published: 18 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

This paper presents an acoustic and auditory study of palatalization in the nasal consonant /n/ in Cairene Arabic (CA). While the mechanics of palatalization have not been widely studied across Arabic dialects, the few studies that have explored this phenomenon have centered on the sociolinguistics, phonology and phonetics of the palatalized stops, /t, d/ and the palatalization of their pharyngealized counterparts (Haeri 1996a; Youssef 2013). There exist some informal observations that the Cairene coronal nasal can undergo palatalization, but no work has systematically examined this sound.

Palatalization of coronals generally involves fronting and heightening articulatory gestures. Thus, a raised F2 and lowered F1 is expected due to the fronting and raising gestures associated with a shortened front cavity. Previous acoustic literature on palatalization have used F2-F1 as a possible cue in distinguishing between palatalized and non-palatalized consonants (Kochetov 2017; Iskarous and Kavitskaya 2010; Purcell 1979). However, this measure can be problematic when studying nasals, as antiformants are known to have an obscuring effect, rendering F1 measures unusable. Studies concerned with nasals have alternatively examined duration, intensity, or nasal murmurs, and have found F2 transition alone to be a highly effective cue for establishing palatalization (Recasens 1983; Harding and Meyer 2003; Kerdpol 2012). Upon examining palatalization in nasals, a linear relationship was established between the formant frequencies at the vowel onset and midpoint, in which F2 at vowel onset varies in relation to the coarticulatorily produced vowel (Sussman et al. 1991; following Lindblom 1963b). Due to F2's reliability in determining plain vs. palatalized nasals in particular, this paper utilizes two points along the F2 transition—one at vowel onset and the other at vowel midpoint/steady-state—and provides a method to establish palatalization in the CA nasal. Auditorily coded tokens are further analyzed in comparison to triangulate the acoustic findings, therefore providing two types of results:

one that indicates the strength/degree of palatalization measured in continuous terms, and the other in terms of frequency or proportions palatalized. This paper provides the first detailed phonetic descriptions of the Cairene [nʲ].

The participants in this study are seven native speakers from Cairo in their early 20 s, and the data include word-final /ni:/ segments elicited from word-list readings. Results show that using formant transition (F2Onset-F2Midvowel) as a measure effectively distinguishes between plain [n] and palatalized [nʲ]. By contributing new acoustic data for an unexamined variant in Arabic, my results provide a foundation for those interested in conducting further sociolinguistic and comparative work on palatalization across Arabic dialects.

1.1. Articulatory Gestures and the Phenomenon of Palatalization

The linguistic phenomenon of palatalization is not uncommon among the languages of the world, let alone within the dialects of Arabic. As a speech process involving the production of a secondary articulation, palatalization entails shifting the primary place of articulation towards the palatal region (Kochetov 2011), or the “superimposition of a raising of the front of the tongue toward a position similar to that for *i* on a primary gesture” (Ladefoged and Maddieson 1996, pp. 363–65). In coronal primary articulations, this involves a displacement of the tongue surface, which would have been realized to support movement of the tongue-tip in the non-palatalized production, to a slightly different primary constriction location (Ladefoged and Maddieson 1996, p. 365).

Two general types of palatalization are often discussed: secondary palatalization (Bateman 2007; Hall 2000; Kochetov 2011) and full palatalization (Bateman 2007). Secondary palatalization, also referred to as “tongue-raising” (Bhat 1978), refers to the addition of a secondary, palatal articulation without changing the initial place of articulation, such as [t → tʲ] (Bateman 2011, p. 589). This type is extremely common in the labial, coronal and dorsal places in many languages of the world (Kochetov 2011). Full palatalization (Bateman 2007), can include palatalization to a posterior coronal and to an anterior coronal (Kochetov 2011). A shift to the posterior coronal may result in a non-sibilant sound, e.g., [t, k → c], or a sibilant sound, e.g., [t, k → tʃ]. A shift to the anterior coronal can result in a non-sibilant sound, e.g., [p, k → t], or a sibilant, which is rare, e.g., [p, k → ts], and [t → ts], which is relatively more common (Kochetov 2011, p. 1671).

The most likely phonetic triggers of palatalization are the high front vowel /i/ and palatal glide /j/ (Chen 1973; Bhat 1978; Hall 2000; Hall and Hamann 2006; Hall et al. 2006; Bateman 2007, 2011, p. 596), followed ‘at a considerable distance by mid front vowels’ (Kochetov 2011, p. 1672). The acoustic similarities between the high front vowel and the palatal glide play a role, but it is also worth noting that different consonants in various languages have been found to have different triggers. At the typological level, however, Kochetov noted the dependencies between triggers and targets, in which coronals are commonly targeted by high vocoids and dorsals by /i/ and other front vowels (Kochetov 2011, p. 7). Bateman noted that if there were only one vowel trigger of palatalization in a language, that vowel should be /i/ due both to its high and front qualities (Kochetov 2011). These findings aid in better understanding palatalization in Cairene Arabic, as the triggers in this language variety follow the aforementioned pattern (discussed in the next section).

In describing the articulatory gesture of palatalization, it is recognized that co-articulation has to do with the occurrences of two different articulations at the same time. (Catford 1988, p. 106), and that palatalization thus occurs as a type of coarticulation. In CA, palatalization occurs as phonetic coarticulation, as opposed to a phonemic shift (Youssef 2013). We will thus henceforth treat palatalization of /n/ as a gradient, phonetic feature.

As the data analyzed in this study all contain /n/ in the syllable-final /ni/ position, the articulatory gestures involved are necessarily unpacked. First, to produce the nasal /n/, the soft palate is lowered, and there is a complete closure in the mouth: between tongue-tip and teeth or teeth-ridge for [n], so that all airflow is shunted through the nose (Catford 1988, p. 74). The realization can be apico- and lamino- articulations against the dental zone

and against the front and back of the alveolar zone (Catford 1988, p. 82). Secondly, on producing the high front [i, i:] vowel, the Jones' system of a "vowel limit" posits that since the tongue is tense and the dorsal surface is pushed close enough to the hard palate, there is a point at which the approximant [i] turns into a palatal fricative [j] (Catford 1988, p. 125). In other words, if the vowel is 'high enough,' it will ultimately result in a palatal-like production. Articulatorily speaking, [i] and [j] have an identical starting point, and "the highest point of the tongue in front vocoids lies on the front of the tongue, underneath the palatal zone" (Laver and John 1994, pp. 276–77). As the CA vowel [i:] is generally high and tense, it is feasible and likely for palatalization to occur as an assimilatory process (as documented in Youssef 2013, 2015).

During consonant production of the /ni/ segment, the tongue dorsum height and dorsopalatal contact size in the nasal change as a function of the adjacent vowel in the progression [i] > [u] > [a] (Recasens 1999, p. 89; Recasens 1984). As such, on contact with /i/, which is a vowel that requires the raising and fronting of the tongue, /n/ shifts from a coronal to a lamino-alveolar consonant (Recasens 1999, pp. 88–89). This results in a larger contact surface between the tongue and the palate.

Similarly, during palatalization, dentoalveolars such as /n/ undergo tongue dorsum raising and fronting, causing the palatalized coronal to become lamino-alveolar. Thus, there appears to be some overlap in the coarticulatory gestures for producing both the plain and palatalized /ni/ segments. Given the explanation above on Jones' vowel limit, we recognize that the transition from a 'regular' coarticulatory effect of the following vowel to one that is palatalized is gradient, and this is reflected in the auditory coding process, which recognizes that what is counted as palatalized /n/ can be wide ranging (see Section 2.2).

While the displacement of the tongue can have many acoustic and auditory consequences, the observation of the data presented in this study is one of a /j/-like quality in the release of /n/ into /i/, suggesting palatalization, which is further demonstrated through F2-raising in the acoustic analyses.

On coarticulation in VCV contexts, there is evidence that vocalic anticipation is blocked when [i] contributes to the raising of the tongue dorsum in [ni] (Recasens 1999, p. 99). Due to the increase in tongue-dorsum constraint when producing /n/, C-to-V carryover (left-to-right) effects are said to be more prominent for [ini] than for [ana] (Recasens 1999, p. 99). However, what is observed to be occurring in CA contradicts this, as it indicates anticipatory (right-to-left) coarticulation, with the influence of /i/ on the preceding consonants. Bladon and Al-Bamerni (1976, p. 148) note that anticipatory coarticulation occurs whenever an articulator is free to anticipate later segments (following Daniloff and Hammarberg 1973), implying a high-level encoding process of scanning ahead, or due to postulating the unit of speech encoding to be an articulatory syllable (consisting of a CV sequence). The triggering effect of /i/ in CA can thus be viewed as anticipatory coarticulation.

Additionally, relevant to our discussion of nasals is their observation by early Arab and Muslim phoneticians. According to Al-Khalil (d.175/791 in Darwish 1967)¹ and Sibawayhi (d.177/793)², the nasals /n, m/ were described as containing nasality (Bakalla 1981, p. 286) and as prone to assimilation, a phenomenon that is widely observed to be intrinsic to the nasal class (Bakalla 1981, p. 286). According to Sibawayhi, the Arabic nasal sounds are produced in a similar way to the modern phonetic description of the nasals, with the complete closure of the air in the oral cavity, and are further described as *munfatih*, or non-velarized, while Ibn Jinni (Ibn Jinni 1954, d.392/1002 in Al Halabi) notes they are *munkhafid*, or with lowering of the tongue body. Sibawayhi discusses '*ikhfā*' or hidden, *m* and *n*, referring to homorganic assimilation in place of articulation of *n* to the following consonant, e.g., *man jā'a* → *majjā'a* (Bakalla 1981, p. 290). This type of assimilation is also present in the observation on palatalization in cases where *n* is followed by the approximant *y* [j], observed by Al-Saqqaf (1999) and Haeri (1996a); below, see Table 1.

Table 1. Triggers of palatalization in Cairene (/t, d/ from Haeri 1996a, p. 51; Youssef 2015, pp. 25–27; /n/ from my own informal observations).

Environment	/t, d/	/n/
[j] glide	hadja ‘quiet f.sg’ → hadʃja	ha:nʲa ‘female name’ → ha:nʲja
[i] (word-final)	fa:dʲi ‘empty m.sg’ → fa:dʲi	tæ:ni ‘again’ → tæ:nʲi
[i:]	gidi:d ‘new’ → gidʲi:d	sini:n ‘years’ → si nʲi:n
[ɪ] epenthetic ^a	ruħt -i- gɪbt ‘(I) went and bought’ (lit. brought) → ruħtʲ -i- gɪbt	–
[i] (non-final) ^a	tiktib ‘you write, masc.’	–
[ee] ^a	sanateen ‘two years’	mine:n ‘from where’ → minʲe:n
low vowels ^a	uskutuu ‘be quiet, plur.’	–
at word boundaries ^a	sitt ʔawi ‘woman very’ → sittʲ ʔawi	–
near sibilants	wisti → ‘my waist’ wisʃtʲi	ħosni ‘male name’ → ħos nʲi

^a While Haeri (1996a) found palatalization in these environments, Youssef (2015) found them to block or lack WP.

While nasality is an accompanying feature that, similar to palatalization, involves lowering of the soft palate so that the air stream passes through the nasal cavity as well as through the oral cavity (Ladefoged and Maddieson 1996, p. 131), the acoustic perception in this study is of palatalization: Coders, who were trained phoneticians, searched specifically for /j/-like qualities in the syllable-final /ni/ segment. Acoustically, while an F1 bandwidth increase is indicative of nasalization, the acoustic results are difficult to interpret (Pruthi and Espy-Wilson 2007) and aerodynamic measures (not employed in this study) are better at capturing these effects. While nasalization is worthy of further examination in a more detailed study of nasals in Arabic, it is not discussed further here, as our scope is limited to the process of palatalization on one of multiple CA coronals undergoing this phenomenon. This paper establishes a premise for further study of the potential spread of palatalization onto other consonants not previously discussed.

1.2. Palatalization in Cairene Arabic

In CA, the palatalization of /t, d/ stops and their pharyngealized counterparts have been examined by Bhat (1978), Haeri (1996a) and Youssef (2013). Haeri noted two types of palatalization in Cairene, one of which is termed weak palatalization (WP), which refers to the secondary palatalization described above. The other is termed strong palatalization (SP), and refers to full palatalization. In representing the auditory effects, WP in Cairene Arabic can be represented in IPA as [t, tʰ → tʲ], and [d, dʰ → dʲ], while SP can be described as affricates: [t, tʰ → tʃ], and [d, dʰ → dʒ].

Weak and strong palatalization in CA are triggered when stops are followed by the palatal glide /j/, long /i:/, word-final /i/, as well as by the phonetically lower word-internal or epenthetic short /i/ or by the long mid vowel /ee/ (Haeri 1996a; Youssef 2013). The environments listed in Table 1 were described by Haeri and Youssef as the observed conditions for palatalization in Cairene.

In addition to the aforementioned coronal stops, Geenberg briefly observed in her study on palatalized stops that the Cairene coronal nasal may also undergo palatalization (Geenberg 2012, p. 21). Al-Saqqaf’s descriptive work on Hadramawti Arabic (Al-Saqqaf 1999) again briefly mentioned the palatalized Cairene /n/ as a comparative point to Yemeni, and noted that the palatalized Hadramawti /n/ is not limited to the high-front vowel environment, unlike in Cairene. As he noted, “n, which escaped Haeri’s attention (Haeri 1992, p. 171), is also among the consonants that become palatalized in the environment of i

or *i* in, e.g., Ar. women's speech, e.g., *inti* [ɪntʃi] 'you' f.s., *ya'ni* [jæʃni] 'it means; I mean' (Al-Saqqaf 1999, p. 95). However, neither Al-Saqqaf nor Geenberg undertook acoustic analyses of the palatalized nasal, despite its presence in multiple speech communities.

The triggers described for WP are relevant to that of the palatalized /n/, as it was observed that palatalized nasals occur in all of the environments for WP, but not necessarily for SP (Geenberg 2012). This is congruent with my own informal observations, although similar to Youssef, I did not observe palatalized [nʲ] in all environments listed by Haeri (see Table 1 for [nʲ] triggers). Grammatically, final [i] in Cairene has several roles: a noun-derived adjective (e.g., ʔamrikæ:n-i:, 'American'), and the first-person possessive or object pronoun (e.g., ʔibn-i, 'my son'; *istannu:-ni*, 'wait (2nd, pl.) for me!'). Word-final /ni:/ may also occur in names (e.g., hæ:ni, *ħosni*), in other common words such as *tæ:ni*, 'again,' and extends to borrowed English words such as 'any,' and 'funny,' which commonly occur in the speech of educated Cairenes. Though a gradient, phonetic feature, the palatalization observed in CA is assimilatory and contains an anticipatory coarticulatory effect, since the consonants become more similar in place of articulation to the following vocoids (see Kochetov and Alderete 2011).

1.3. Acoustics of Palatalization

Acoustically speaking, F2 formant transitions have been widely used as cues to determining nasal places of articulation. Palatalization in particular is often more apparent at the consonantal release than at the formation of the primary constriction, with a higher F2 value at the release (Ladefoged and Maddieson 1996, pp. 363–64). In a study comparing the interactive effects of nasal murmurs, transitions, and release as possible cues for place of articulation, Recasens (1983) noted that examining formant transitions proved useful in determining the palatal nasal's place of articulation, while the nasal murmur was not a useful cue for identifying this sound. It should be noted that the F2 transition may not provide a sufficient place cue for other types of nasal. Bilabial and velar nasals, for example, may rely on other features such as the nasal murmur or quality of the nasal release instead of, or in addition to, using F2 as a place cue (Recasens 1983). In examining formant transitions in the event of coarticulation, formant shapes will vary according to the surrounding vowels (Öhman 1966), but will generally be directed towards the same 'locus' or 'juncture' between the vowel and the sonorant, although F2 loci are not invariant in natural speech (Fant 1973; Kewley-Port 1982; Lehiste and Peterson 1961; Öhman 1966).

In describing the palatalized stops [t, d, tʃ, dʃ] in Cairene, the typical raising/fronting of F2 in these palatalized stops were seen in Youssef's acoustic data (Youssef 2015; from a 38-year-old female speaker), with F2 raising by approx. 390–460 Hz in the palatalized stops compared to their plain counterparts. Note that both the pharyngealized and plain stops can undergo palatalization in Cairene, and that the pharyngealized stops de-pharyngealize, or weaken in their pharyngealized quality, via fronting (Haeri 1996a; Youssef 2015).

Assuming that fronting is a necessary accompanying articulatory gesture for palatalization in CA, and for the above stated reasons which prevent the ability to use F1 as a measure for the nasal, this paper makes use of F2 as a cue for tracing palatalization in the presented data generated from word-list readings. The following sections define the auditory and acoustic measures used to triangulate the study of palatalized [nʲ] in CA, and proceeds with discussing the implications of the main findings and future directions on the study of this phenomenon.

2. Materials and Methods

2.1. Participants, Data Collection and Selection

Speech data gathered from 7 participants were used for the acoustic analyses in this section. They include three women and four men aged 20–27 who are native Egyptian colloquial speakers residing in the greater Cairo area at the time of data collection (month of July 2014; Due to the increasingly unstable political situation in Egypt, it became and

remains nearly impossible to do social science fieldwork in the country). All speakers were born, raised, and either attended or completed private or public university in Cairo.

For the purpose of eliciting controlled, comparable data that can be efficiently measured and analyzed, all speakers were given a wordlist written in Arabic script in dialectal spelling (e.g., *بيحبوني*, *تاني*). Of the 40 words, the 20 target words contained the coronal nasal preceding the high front vowel in the word-final /-ni:/ environment, which reflects the phonetic condition for palatalization in CA. A few tokens were noisy and produced unclear spectrograms, and were thus omitted. This yielded a total of 138 target tokens containing word-final /-ni:/ that were analyzed. The preceding vowel was not controlled for, but the words selected were items that I had informally observed to variably contain palatalization. To minimize overstressing/overemphasis of items containing the word-final /-ni:/ 20, other filler words that did not contain the /-ni:/ segment were included alongside the target items. Among the filler words, 6 tokens containing the nasal in the /-na/ word-final environment were produced, as well as 3 nasals in the word-medial /-ne-/ environment, while the rest contained no /n/ consonant at all. Although a small-scale study, this paper provides a detailed work to support the anecdotal evidence observed by the aforementioned scholars of variation in the variable (n).

Participants were recorded in spaces with as minimal noise as possible, and took place in either my home or the participant's home, or in a rented meeting room at a local cafe. However, some street noises that permeate the bustling city of Cairo could not be entirely avoided. This is a common challenge for fieldworkers in greater Cairo, where quiet, soundproof recording studios belonging to institutions, if available, have restricted access and are not always practical or possible to use. The speakers were recorded in wav format using a Zoom H1 Recorder (48 kHz) with an external cardioid lavalier microphone (SP-CMC-2), which they were instructed to hold 5–10 inches from their mouths.

2.2. Auditory Coding and Acoustic Measurements

An auditory coding method was first employed by the author, and 75% of the data was coded by two other linguists: a sociophonetician who is not a speaker of Arabic and a linguist who is a speaker of Cairene Arabic with familiarity of phonetic variation in Arabic dialects. The tokens are coded as 'n' for the plain, unpalatalized nasal, or 'nj' for tokens that contained any degree of palatalization (from lighter to stronger degrees of palatalization). Recognizing that palatalization is a gradient, phonetic feature in CA, the coding originally allowed for three categories: 'non-palatalized,' 'somewhat palatalized,' and 'strongly palatalized,' but the last two categories were collapsed into a larger 'palatalized' category for analysis. The two speakers of Arabic coded the data with 61% agreement, a point further discussed later in this paper. When these two coders did not agree on a token, the coding of the third, non-Arabic speaking phonetician was used in a tie-breaker system to code that token. These auditorily coded items are labeled Auditory Code in the statistical model.

The acoustic measurements used to determine the cue to palatalization of the nasal in this study are performed by obtaining the frequency value of F2 by hand, in Praat, measured in hertz (Hz), at two points in the word-final /-ni:/ segment: one point at the release of the nasal murmur into the following vowel (coded: F2Onset), and another at the midpoint of the same vowel (coded: F2Midpoint). For each word, the F2Midpoint was subtracted from F2Onset, and the resulting value is henceforth referred to as "F2Diff". This follows Lindblom (1963a, 1963b), Gibson and Ohde (2007) and Sussman et al. (1991), who found that comparing the formant transition between these two points served as a useful cue in distinguishing between palatal/palatalized and plain alveolars. A larger, negative F2Diff value indicates a steeper, upward CV transition, while a smaller value indicates a flatter, upward transition. To visually illustrate this, the two points measured (F2Onset and F2Midpoint) are marked in each spectrogram in Figure 1, which shows spectrograms from two speakers: one non-palatalizer (left), and one palatalizer (right), saying [hæ:ni:] 'name, Hanny'. The left, non-palatalized spectrogram shows F2 coming out of the 'n' closure at a lower point and heightening rapidly and into the steady-state/midpoint of the [i:] vowel.

The right, palatalized spectrogram, however, shows F2 starting at a higher point at [i:] onset, and shows little to no upward transition into the vowel midpoint. In the acoustic analysis, taking the difference between the two points measured in each vowel allows for alternative way of normalizing between vocal tract length, so no additional vowel normalization techniques were employed.

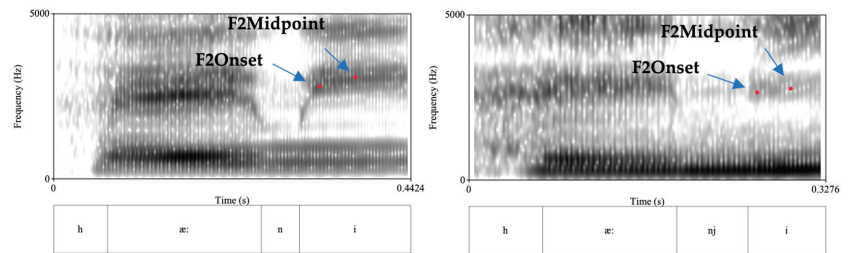


Figure 1. Spectrograms from two women: one non-palatalizer (left), showing a lower F2 at the ‘i’ vowel onset, and one palatalizer (right), showing higher F2 at onset, saying [hæ:ni] ‘male name, Hanny’.

2.3. Statistical Analyses

Using the package AFEX, a wrapper for lmer in R (Singmann et al. 2015; R Core Team 2013), the data were first fitted with the following linear mixed effects regression model: $F2Diff \sim AuditoryCode * Word + (1 | speaker)$. The dependent variable F2Diff is the subtracted value of F2Onset-F2Midpoint, and is expressed as a continuous variable (Hz). The fixed effects are the auditorily coded tokens, coded as AuditoryCode (categorically coded: n/nj) and word (categorically coded: 20 levels). The variable speaker was included as a random intercept to account for individual variation.

3. Results

3.1. Auditory Results

The total number of auditorily coded tokens include 66 tokens coded as palatalized [nʲ], and 72 tokens coded as plain [n], yielding a total of 47.8% palatalized tokens (Table 2). This is evidence in itself that the palatalized variant is a robust realization of /n/ before /i:/ in CA. Table 2 shows the breakdown of (nj) versus (n) codes by individual speaker. While two of the women show categorical realizations (interestingly, in different directions), most participants produce both variants, and averaging across participants, there is not a compelling difference based on speaker gender (41.67% for women compared to 53.63% for men) in this data pool.

Table 2. Token count and percent palatalized by speaker and gender.

Speaker	Gender	[nʲ]	[n]	Total	% Palatalized	Speaker	Gender	[nʲ]	[n]	Total	% Palatalized
1	F	19	0	19	100%	4	M	17	3	20	85%
2	F	0	19	20	0%	5	M	6	14	20	30%
3	F	5	15	20	25%	6	M	2	18	20	10%
						7	M	17	2	19	89.5%

Some of the intra-speaker variation may be driven by lexical effects. Table 3 contains the proportion of palatalization in five of the seven speakers (two speakers were omitted as they were categorically either a palatalizer or non-palatalizer, revealing no by-word variation). The table and visualized data in Figure 2 show that the words produced with the fewest palatalized tokens—i.e., only 1 instance of (nj) code in each—were *yaʒni* [jaʒni], ‘I mean/meaning,’ which contains the voiced pharyngeal fricative known to have a formant-lowering effect, and the borrowed words (from English) *funny* and *any*. The word with the

highest proportion of palatalized tokens is *kallimiini* ‘call me, 2nd, f.sg.’ (80%; Table 3). This is unusual, as other words with final /i:ni/ syllables should have had similar coarticulatory patterns, but this is not the case in the current findings. It is otherwise not immediately clear what unites the words that were or were not frequently palatalized, but English borrowings certainly seem to be among the least palatalized.

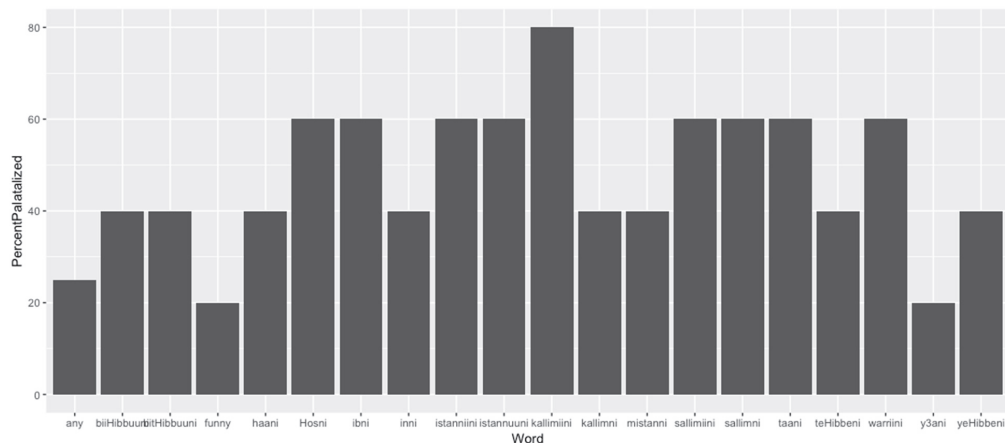


Figure 2. Visualization of proportion of palatalized tokens by word (based on Table 3).

Grouped by context, words ending in words ending in /i:ni/ (*kallimiini*, *istanniini*, *warriini*, *sallimiini*) are somewhat more likely to be palatalized (60–80%), compared to other contexts, such as the /u:ni/ (*istannuuni*, *biHibbuuni*, *biHibbuuni*), /ani/, /a:ni/ and /Cni/ environments (40–60%). The geminate /nn/ (*mistanni*, *inni*) and geminate + epenthetic /bb-1/ groups (*teHibbeni*, *yeHibbeni*) are somewhat less likely to be palatalized (40%) than all other contexts.

3.2. Acoustic Results

Based on the raw token counts above and thus following the methodology from the auditory analysis, the two speakers who categorically produced either (n) or (nj) 100% of the time were omitted in the acoustic analysis. This controls for any influential points that would have caused errors when examining lexical variation.

Upon examining F2Diff, it is apparent that these values distinguish tokens we heard as plain (n) vs. palatalized (nj). This is displayed in Figure 3, which demonstrates less of a difference between the two points measured in (nj), since the F2Onset is starting very high and barely moves to reach the F2Midpoint height of the /i:/ vowel. Contrarily, in plain (n), F2Onset starts at a lower point and moves higher into the vowel midpoint, so there is a greater height difference. Thus, a main effect of Auditory Code (n, nj) on F2Diff is observed ($F = 4.06$, $df = 1$, 112.67 , $p = 0.01$, $N = 99$). The distinguishing cue appears to be around the 100 Hz mark, indicating that if F2Diff is less than 100 Hz, the auditory quality is likely palatalized, while an F2Diff greater than 100 Hz covaries with an auditorily non-palatalized nasal. The average F2Diff is -62.4 Hz for palatalized codes, and -209.5 Hz for the non-palatalized ones.

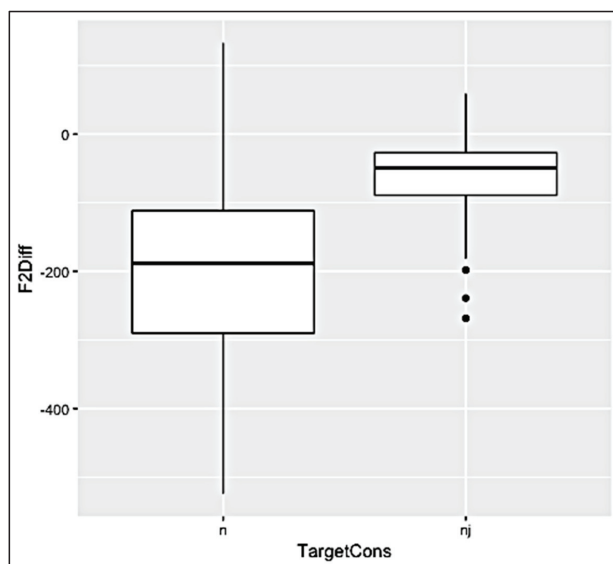


Figure 3. Plot of F2Diff (in Hz) by Auditory Code (n, nj) in all speakers. N = 138.

4. Discussion

This study was motivated by the anecdotal evidence of palatalized [nʲ] as an allophone of [n] and the absence of any acoustic or auditory study of this sound. The acoustic and auditory analyses on the speech productions of the seven speakers presented here indicate that the two types of production exist and are acoustically distinct from one another in CA.

The first main effect found for Auditory Code on the dependent variable F2Diff (the subtracted value of F2Onset-F2Midpoint) contributes to works surrounding the palatalized nasal in a few ways. First, it suggests that using auditory coding as an approach to measuring palatalized /n/ is reliable, as it correlates with the acoustic measure. This finding, along with previous informal commentaries, confirms that the list of palatalized consonants in Cairene Arabic must be expanded beyond the /t, d/ stops to include the palatalized nasal. Additionally, the linear relationship found in the points utilized in the F2Diff measure also confirm the methodology proposed by Sussman et al. (1991) and Lindblom (1963a) for analyzing nasals, and further supports the use of linear regression as a method of analysis.

The findings from the auditory coding revealed a rater agreement of 61%, which indicates that while auditory coding was effective, palatalized [nʲ] can be difficult for listeners to code. This may be an effect of perceptual compensation—a type of perceptual bias that ordinarily leads listeners to ignore, or correct for, coarticulatory effects (Garrett and Johnson 2013). It is possible that coders ignored palatalization in instances where the coarticulation was relatively milder, leading to some discrepancies in the coding process. Despite the difficulty in auditorily coding this sound, F2Diff remained effective in distinguishing (n) from (nj) codes, which suggests that the triangulated method employed can be used to measure the degree of palatalization in future studies.

The effect on borrowed words was unexpected, since palatalization had been previously observed in my informal observations on both borrowed words included in this study (particularly in *funny*) in casual speech from speakers with similar educational backgrounds to the participants. Yet, it is known that loanwords can be realized with native or non-native sounds and that topic, speaker- and word-specific sociolinguistic factors can determine the selection of one variant over another (Hashimoto 2019). Some factors that can affect this selection include level of bilingualism (Poplack and Sankoff 1984), degree of

linguistic integration (Haugen 1950), and language dominance (Aktürk-Drake 2015, 2017). More recently, Hashimoto's work on tap [ɾ]-borrowing from Māori into New Zealand English uses an exemplar-based approach (Pierrehumbert 2001; Docherty and Foulkes 2014) which posits that exemplars with native sounds and those with non-native sounds are represented in the cognitive system of a borrower and updated based on linguistic experience (Hashimoto 2019). Relevant to our discussion is the idea that exemplars with non-native sounds are stored in relation to a social category associated with the source language and its culture (Hashimoto 2019). In this study, it is possible that the context of the word-list reading task had an effect and, intersecting with the participants' mental representations of the borrowed words, in turn elicited plain [n] as the appropriate selection for the task. Future studies may be interested in further examining the effects of palatalization across speaking contexts, or specifically on loanwords, at greater length.

As for the word *kallimiini* containing the highest proportion of palatalization in the auditory coding, without similar coarticulatory effects in other words from the same environment, the motivation is unclear. There may be socially motivated reasons related to the affective quality of this particular word, potentially combined with sound symbolism. As exemplified in Japanese, a type of palatalization exists that is not phonologically conditioned, but rather contains an iconic function, and is linked with "smallness", "childishness" or "affection" (Nichols 1971; Ferguson; Ohala 1994). Such types of "expressive palatalization" occur cross-linguistically in sound symbolism, diminutive morphology, hypocoristics, and in "babytalk" (Kochetov and Alderete 2011, p. 346). A sociolinguistic perception study on palatalization of /n/ in CA would be enlightening.

Another factor is the prosodic position of the final morpheme /ni/. The list of words analyzed include words with stress systematically placed on the penultimate, which has an effect of articulatory reduction: The morpheme /ni/ is phonologically /ni:/ with a long vowel, which undergoes vowel temporal reduction, resulting in an undershoot (Lindblom 1963b, pp. 1776–79). In languages with heavy stress, vowel reduction is a characteristic feature, especially in weakly stressed syllables (Lindblom 1963b, p. 1773). In CA, word-final vowels in open syllables are never stressed, except in monosyllables such as *dī* 'this, f.' (Youssef 2013, p. 242). Youssef further recognizes the syllable-final reduction in his observation that "Short /i/ triggers palatalization only when it is word final", e.g., *h'idi* 'he became satisfied,' and *nab'a:vi* 'vegetarian' (Youssef 2013). The results observed in this paper follow this observation, and support the finding that syllable-final undershooting does not prevent the occurrence of palatalization in CA.

Additionally, as stated above, the high-front vowel /i/ is reported to be the main trigger of palatalization in CA (see Table 1). As Youssef's findings on weak palatalization of the stops /t, d/ revealed, WP is observed to occur as a phonetic co-articulatory effect of following /i/, since the articulation of the target consonant is affected by the high and front position of the tongue in the production of the following vowel (Youssef 2015). It is further reported that the vowel height of /i/ is a distinguishing articulatory feature of CA, and that there exists a vowel hierarchy in which long [i:] is higher than short [i], and word-final [i] is higher than non-final [i] (Haeri 1996a, p. 57; Youssef 2015).

As well, the auditory results show that words ending in /i:ni/ are somewhat more likely to be palatalized than words ending in /u:ni/, /ani/, /a:ni/ and /Cni/. This supports the idea of palatalization as an assimilatory process in which surrounding vocoids act as triggers (Youssef 2015; Kochetov and Alderete 2011): in this case, having an i:Ci syllable somewhat increased the likelihood of palatalization.

My data on word-final /-ni:/ show that palatalization occurs nearly 50% of the time in this environment, which is fairly similar to the probability of palatalization Haeri found in the /t, d/ stops word-finally (63%, Haeri 1996b, p. 58). This probability is second to palatalization in the *stop + j* environment (i.e., a stop followed explicitly by the [j] glide and assimilating to its feature: *nadya* [nad.ja → nad^ɹ.ja] 'name, Nadia') in Haeri's data, with a proportion of 68% (Haeri 1996b, p. 58). If we assume that palatalization of /t, d/ occurs more frequently than /n/ (Sokhey 2015), and that the two have similar phonological conditions

(see Table 1), then the fact that my data show a proportion of palatalized /n/ that is just below Haeri's data on /t, d/ support the hypothesis that the word-final environment is one of the most likely triggers of palatalization of /n/. At the typological level, my findings further support the notion that coronals are the most common sounds to undergo at least secondary palatalization in the coronal range (Bateman 2011).

The lower proportion of palatalization observed in /n/ compared to /t, d/ (based on Haeri's and my data) further begs the question of whether a change in progress is taking place, whereby palatalization is spreading from /t, d/ to /n/. Sokhey (2015) hypothesized this based on synchronic, auditorily coded /t, d/ and /n/ tokens, but a historical study tracing the emergence of the palatalized nasal, and/or an acoustic study on the relationship between the palatalized nasal and stops would be further revealing.

Sociolinguistic categories must also be taken into consideration. While the data presented here do not reveal a gendered pattern, preliminary work using free speech data reported avoidance of [nʲ] by men and more frequent use by women (Sokhey 2015). Given that there is strong evidence that palatalization of /t, d/ are features of a sociolect in CA (i.e., palatalization covaries with the larger sociolinguistic categories such as socioeconomic class and gender; Haeri 1996a; Youssef 2015), a sociolinguistic examination of the palatalized nasal is warranted. Haeri found in her work that weak palatalization of the stops /t, d/ is an innovation of upper-middle class women, and Youssef found that strong palatalization (i.e., affrication) of the same stops have been phonologized into the sociolect of a group of speakers who use it to index covert prestige in opposition to the upper classes (Youssef 2015). Decades after the first sociolinguistic work was conducted on /t, d/ in CA, it is not unlikely that weak palatalization has advanced to neighboring consonants. The Cairene nasal, as an available palatalized sound, is an ideal candidate for such ideological extensions.

5. Conclusions

This paper concludes that the palatalized Cairene [nʲ] is acoustically distinguishable from the coronal /n/, and examining the CV transition proves to be a useful distinguishing cue. Borrowed English words observed in the word-list data produced fewer palatalized nasals while intervocalic /n/ with both a preceding and following high front vowel appears to render palatalization stronger. Given that the palatalized nasal is not uncommon in Cairene speech today, future studies may examine the sociolinguistic status of palatalized /n/ in relation to the palatalized /t, d/ stops, as well as determine whether palatalization has spread to other consonants. This advancement warrants further examination of the status of the proposed sociolects involving WP and SP—i.e., whether WP continues to covary with upper-classness, and if/how this affects the social status of SP and those who use it to display opposition to upper-classness.

Furthermore, a comparative study of palatalization across other dialects and/or a historical study that traces the appearance of this phenomenon would be a worthwhile and informative study on the progression of social salience that can be used not only to study Arabic dialects, but also to study other speech communities. As a phenomenon that is linked to socioeconomic class in CA, palatalization is a sociolectal feature that requires a culturally appropriate index for measuring social class in order to be studied at a larger scale. While this has not been widely done in Arabic sociolinguistics, some scholars began work on this: Haeri created an index for her study of palatalization using a group of socioeconomic indicators with varying degrees of importance (Haeri 1996a). Rania Habib examined socioeconomic indicators in Christian rural migrant speakers in Hims, Syria and found that income, followed by residential area, were the strongest indicators of social class, which differs from the situation in the western world (Habib 2010). In the Gulf (e.g., Bahrain and the UAE), family name and communal background have weighty impacts on the internal evaluations of social status and economic opportunities, but no Arabic sociolinguistic work has considered these variables. It is apparent that socioeconomic indicators differ between Arabic speaking communities, and the aforementioned communities are

promising venues for further (and updated) investigation on the intersection of linguistic variation and social class. It is hoped that the work outlined here provides a basis for future studies involving not only palatalization, but further sociolinguistic work within other Arabic speech communities.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of The University of Texas at Austin (protocol code: 2014-04-0121; date of approval: 4 May 2014).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy restrictions.

Acknowledgments: This work was made possible by the friends and acquaintances who kindly volunteered their time to participate in this study in Cairo back in 2014. Thank you to Kristen Brustad, Abby Walker, Corinne Stokes, Aarnes Gudmestad, and Katie Carmichael for their support, encouragement and guidance throughout this project and while preparing this manuscript. Thanks to the audiences at Jil Jadid and ASAL Conferences for thoughts and reflections on the earlier iterations of this project. Any shortcomings that remain are my own.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Gloss of items in word list (transliteration note: H = ħ, 3 = ʕ)	
kallimiini	all, 2nd f.sg. me
Hosni	male name
ʔibni	my son, m.sg.
istanniini	wait 2nd f.sg. for me
istannuuni	wait 3rd, f.sg. for me
sallimni	hand over, 2nd m.sg. to me
taani	again, m.sg.
warriini	show, 2nd sg. me
sallimiini	hand over, 2nd fs. to me
biiHibbuuni	they love me
bitHibbuuni	you, 2nd pl. love me
haani	male name
kallimni	call, 2nd m.sg. me
mistanni	waiting, m.sg.
teHibbeni	you, m.sg. love me
yeHibbeni	he loves me
inni	that I
any	any
funny	funny
ya3ni	I mean/meaning

Notes

- ¹ 'Abu 'Abd al-Rahman Al-Khalil Ibn 'Ahmad. *Kitab al-ʿayn*. 1967. Ed. by 'Abdallah Darwish. Baghdad: Matba'at al-ʿAni, vol. I.
- ² Sibawayhi, 'Abu Bishr 'Amr b. d.177/793. In 'Uthman Sibawayhi, *Al-Kitab*. 1889–1900 Repr., Baghdad: Al-Muthanna, n.d. Cairo: Bulaq.

References

- Aktürk-Drake, Memet. 2015. Phonological Adoption through Bilingual Borrowing: Comparing Elite Bilinguals and Heritage Bilinguals. Ph.D. dissertation, Department of Swedish Language and Multilingualism, Stockholm University, Stockholm, Sweden.
- Aktürk-Drake, Memet. 2017. Language dominance as a factor in loanword phonology. *International Journal of Bilingualism* 21: 584–99. [CrossRef]
- Al-Saqqaf, Abu Bakr. 1999. *The Yemeni Unity: Crisis in Integration. Le Yemen Contemporain*. Paris: Karthala, pp. 141–49.
- Bakalla, Muhammad Hasan. 1981. The Treatment of Nasal Elements by Early Arab and Muslim Phoneticians. *Historiographia Linguistica* 8: 285–305.

- Bateman, Nicoleta. 2007. A Crosslinguistic Investigation of Palatalization. Ph.D. dissertation, University of California San Diego, San Diego, CA, USA.
- Bateman, Nicoleta. 2011. On the typology of palatalization. *Language and Linguistics Compass* 5: 588–602. [CrossRef]
- Bhat, Darbhe NS. 1978. A general study of palatalization. *Universals of Human Language* 2: 47–92.
- Bladon, R. Anthony W., and Ameen Al-Bamerni. 1976. Coarticulation resistance in English/1. *Journal of Phonetics* 4: 137–50. [CrossRef]
- Catford, John Cunnison. 1988. *A Practical Introduction to Phonetics*. Oxford: Clarendon Press, p. 183.
- Chen, Matthew. 1973. Predictive power in phonological description. *Lingua* 32: 173–91. [CrossRef]
- Daniloff, Raymond. G., and Robert. E. Hammarberg. 1973. On defining coarticulation. *Journal of Phonetics* 1: 239–48. [CrossRef]
- Docherty, Gerard J., and Paul Foulkes. 2014. An evaluation of usage-based approaches to the modelling of sociophonetic variability. *Lingua* 142: 42–56. [CrossRef]
- Fant, Gunnar. 1973. Stops in CV-syllables. *Speech Sounds and Features* 10: 110–39.
- Ferguson, Charles A. Baby talk as a simplified register. In *Talking to Children*. Edited by Catherine E. Snow and Charles A. Ferguson. Cambridge: Cambridge University Press, pp. 209–35.
- Garrett, Andrew, and Keith Johnson. 2013. Phonetic bias in sound change. *Origins of Sound Change: Approaches to Phonologization* 10: 51–97.
- Geenberg, Katherine Rose. 2012. “The People Who Say TSH TSH”: The Social Life of Cairene Arabic Palatalization. *University of Pennsylvania Working Papers in Linguistics* 18: 4.
- Gibson, Terrie, and Ralph N. Ohde. 2007. F2 locus equations: Phonetic descriptors of coarticulation in 17-to 22-month-old children. *Journal of Speech, Language, and Hearing Research* 50: 97–108. [CrossRef]
- Habib, Rania. 2010. Towards determining social class in Arabic-speaking communities and implications for linguistic variation. *Sociolinguistic Studies* 4: 175–200.
- Haeri, Niloofar. 1992. Synchronic variation in Cairene Arabic: The case of palatalization. In *Perspectives on Arabic Linguistics IV*. Amsterdam: John Benjamins Publishing Company, pp. 169–80.
- Haeri, Niloofar. 1996a. *The Sociolinguistic Market of Cairo: Gender, Class, and Education*. London: Kegan Paul International.
- Haeri, Niloofar. 1996b. “Why do women do this?” Sex and Gender Differences in Speech. In *Towards a Social Science of Language: Papers in Honor of William Labov*. Volume 1: Variation and Change in Language and Society. Amsterdam: John Benjamins Publishing, vol. 127, p. 101.
- Hall, T. Alan. 2000. Typological generalizations concerning secondary palatalization. *Lingua* 110: 1–25. [CrossRef]
- Hall, T. Alan, and Silke Hamann. 2006. Towards a typology of stop assibilation. *Linguistics* 44: 1195–236. [CrossRef]
- Hall, T. Alan, Silke Hamann, and Marzena Zygis. 2006. The phonetic motivation for phonological stop assibilation. *Journal of the International Phonetic Association* 36: 59–81. [CrossRef]
- Harding, Sue, and Georg Meyer. 2003. Changes in the perception of synthetic nasal consonants as a result of vowel formant manipulations. *Speech Communication* 39: 173–89. [CrossRef]
- Hashimoto, Daiki. 2019. Sociolinguistic effects on loanword phonology: Topic in speech and cultural image. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 10: 11. [CrossRef]
- Haugen, Einar. 1950. The analysis of linguistic borrowing. *Language* 26: 210–31. [CrossRef]
- Ibn Jinni, Abū l-Fath cUthmān. 1954. *Sirr Sinaaeat al Ieraab*. Vol. I. Edited by Mustafa al-Saqqa, Muhammad Al-zafzaf, Ibrahim Mustafa and Abdallah Ameen. Cairo: Mustafa al-Bābī al-Halabi.
- Iskarous, Khalil, and Darya Kavitskaya. 2010. The interaction between contrast, prosody, and coarticulation in structuring phonetic variability. *Journal of Phonetics* 38: 625–39. [CrossRef]
- Kerdpol, Karnthida. 2012. Formant Transitions as Effective Cues to Differentiate the Places of Articulation of Ban Pa La-U Sgaw Karen Nasals. *Manusya: Journal of Humanities* 15: 21–38. [CrossRef]
- Kewley-Port, Diane. 1982. Measurement of formant transitions in naturally produced stop consonant–vowel syllables. *The Journal of the Acoustical Society of America* 72: 379–89. [CrossRef] [PubMed]
- Kochetov, Alexei. 2011. Palatalisation. In *Companion to Phonology, 1666–1690*. Edited by Colin Ewen, Elizabeth Hume, Marc van Oostendorp and Keren Rice. Oxford: Wiley Blackwell.
- Kochetov, Alexei. 2017. Acoustics of Russian voiceless sibilant fricatives. International Phonetic Association. *Journal of the International Phonetic Association* 47: 321. [CrossRef]
- Kochetov, Alexei, and John Alderete. 2011. Patterns and scales of expressive palatalization: Experimental evidence from Japanese. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique* 56: 345–76. [CrossRef]
- Ladefoged, Peter, and Ian Maddieson. 1996. *The Sounds of the World’s Languages*. Oxford: Blackwell, vol. 1012.
- Laver, John, and Laver John. 1994. *Principles of Phonetics*. Cambridge: Cambridge University Press.
- Lehiste, Ilse, and Gordon E. Peterson. 1961. Transitions, glides, and diphthongs. *The Journal of the Acoustical Society of America* 33: 268–77. [CrossRef]
- Lindblom, Björn. 1963a. *On Vowel Reduction (Report No. 29)*. Stockholm: The Royal Institute of Technology, Speech Transmission Laboratory.
- Lindblom, Björn. 1963b. Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America* 35: 1773–81. [CrossRef]
- Nichols, Johanna. 1971. Diminutive consonant symbolism in western North America. *Language* 47: 826–48. [CrossRef]
- Ohala, John J. 1994. The Frequency Code Underlies the Sound-Symbolic Use of Voice Pitch. *Sound Symbolism* 2: 325–47.

- Öhman, Sven E. 1966. Coarticulation in VCV utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America* 39: 151–68. [CrossRef] [PubMed]
- Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency. *Frequency and the Emergence of Linguistic Structure* 45: 137–57.
- Poplack, Shana, and David Sankoff. 1984. Borrowing: The Synchrony of Integration. *Linguistics* 22: 99–135. [CrossRef]
- Pruthi, Tarun, and Carol Y. Espy-Wilson. 2007. Acoustic parameters for the automatic detection of vowel nasalization. Paper present at Eighth Annual Conference of the International Speech Communication Association, Antwerp, Belgium, August 27–31.
- Purcell, Edward T. 1979. Formant frequency patterns in Russian VCV utterances. *The Journal of the Acoustical Society of America* 66: 1691–702. [CrossRef] [PubMed]
- R Core Team. 2013. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online: <http://www.R-project.org/> (accessed on 1 November 2021).
- Recasens, Daniel. 1983. Place cues for nasal consonants with special reference to Catalan. *The Journal of the Acoustical Society of America* 73: 1346–53. [CrossRef] [PubMed]
- Recasens, Daniel. 1984. Vowel-to-vowel coarticulation in Catalan VCV sequences. *The Journal of the Acoustical Society of America* 76: 1624–35. [CrossRef] [PubMed]
- Recasens, Daniel. 1999. Lingual Coarticulation. In *Coarticulation: Theory, Data and Techniques*. Edited by William J. Hardcastle and Nigel Hewlett. Cambridge: Cambridge University Press, vol. 24, pp. 80–104.
- Singmann, Henrik, Ben Bolker, Jake Westfall, Frederik Aust, and Mattan S. Ben-Shachar. 2015. Afex: Analysis of Factorial Experiments. R Package Version 0.13-145. Available online: <http://CRAN.R-project.org/package=afex> (accessed on 1 November 2021).
- Sokhey, Navdeep. K. 2015. The Women (and Men) Who Speak Funny: Negotiating Social Meanings through Palatalization of the Cairene nasal. Unpublished Master's Thesis.
- Sussman, Harvey M., Helen A. McCaffrey, and Sandra A. Matthews. 1991. An investigation of locus equations as a source of relational invariance for stop place categorization. *The Journal of the Acoustical Society of America* 90: 1309–25. [CrossRef]
- Youssef, Islam. 2013. Place Assimilation in Arabic: Contrasts, Features, and Constraints. Ph. D. dissertation, University of Tromsø, Tromsø, Norway.
- Youssef, Islam. 2015. Palatalization in educated Cairene Arabic. *Nordlyd* 42: 21–31. [CrossRef]

Article

A Typological Analysis of Cognate Infinitives in Lebanese Arabic Based on Comparative Semitic Evidence

Ana Iriarte Díez

Department of Near Eastern Studies, University of Vienna, 1090 Vienna, Austria; ana.iriarte.diez@univie.ac.at

Abstract: Despite the relatively scarce literature on the topic and the lack of terminological consensus among scholars, Cognate Infinitives (CI) have been identified to share formal and functional characteristics across Semitic. The present study provides a description of the formal features of Cognate Infinitives in Lebanese Arabic (LA) based on the analysis of linguistic data gathered through a participant observation method. The novelty of this description lies in its comparative approach, which has been developed in the light of the Semitic evidence available, gathered through a review of the main literature available on the topic. The results of this comparative analysis reveal that the grammatical features of Cognate Infinitives in Lebanese Arabic seem to be in line with general Semitic trends that do not, however, always find their parallel in prescriptive descriptions of Cognate Infinitives in Classical or Standard Arabic.

Keywords: cognate infinitive; Lebanese Arabic; typology; Semitic languages

Citation: Iriarte Díez, Ana. 2021. A Typological Analysis of Cognate Infinitives in Lebanese Arabic Based on Comparative Semitic Evidence. *Languages* 6: 183. <https://doi.org/10.3390/languages6040183>

Academic Editors: Simone Bettega and Roberta Morano

Received: 24 June 2021

Accepted: 29 October 2021

Published: 4 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The existence of Cognate Infinitives within the Semitic continuum has been noted for centuries by various scholars, on occasion resulting in the creation of seminal studies on the topic (Goldenberg 1971; Kim 2009).

However, these scholars' attempts to describe the formal and functional nature of this linguistic feature did not always lead to a consensus as far as terminology is concerned. Different grammatical approaches brought about many distinct nomenclatures for one single linguistic form: *maʿfʿūl muṭlaq mubham* in Classical Arabic (Al-Zamaxšari 1870, p. 111); Paronymous Complement in Syrian Arabic (Cowell 1964); Unmodified Cognate Complement in Rural Palestinian Arabic (Shachmon and Marmorstein 2018); Tautological Infinitive in Biblical Hebrew (Goldenberg 1971); Infinitive Absolute in Syriac (Nöldeke 2003); Paronomastic Infinitive in Akkadian (Cohen 2004), etc.

Nevertheless—and notwithstanding the lack of agreement in terminology surrounding Cognate Infinitive constructions—if we are able today to group together this myriad of grammatical labels, it is only because both formal and functional characteristics of Cognate Infinitives seem to be clear enough to be described, even across Semitic languages.

At the formal level, a Cognate Infinitive construction (CI) is formed by two essential elements: (1) a finite verbal form that functions as the lexical head of a predicate (or 'cognate head') and (2) an infinitive that depends syntactically on and is cognate with the verbal head and stands indefinite and unqualified ('cognate infinitive'). This makes CIs different from Cognate Object constructions (CO), which, albeit similar, present an "infinitive" that is specified, modified, or qualified in a variety of ways.

The distinction between CIs and COs throughout most of the Semitic literature also extends to functional grounds. While COs are known to modify the verb adverbially, the function of CIs has been often described with vague notions such as 'emphasis', 'asseveration', 'contrast', or 'intensification'.

Within the Arabic grammatical tradition, these two concepts have been traditionally studied as two facets of one grammatical category: *al-maʿfʿūl al-muṭlaq*. The combined

analysis of CIs and COs in Classical Arabic, which strongly influenced subsequent analyses in Modern Standard Arabic (MSA), seems to neglect the abundant Semitic evidence of analogous constructions that draw a clear grammatical line between these two structures.

The present study argues for the separation of these two types of cognate structures following Semitic treatments, and takes a first step in this direction by describing the formal features of Cognate Infinitives in Lebanese Arabic (LA), based on the analysis of abundant linguistic data gathered through a participant observation method during a period of four years. The novelty of this description lies in its comparative approach, for it was developed in light of the Semitic evidence available, gathered through a review of the literature available on the subject.

Given the scarcity of descriptions of CIs in non-Standard Arabic varieties, adopting a traditional approach for the analysis of the data could misleadingly urge us to read the grammatical features of CIs in LA as ‘exceptions’ to or ‘simplifications’ of the Standard norm, since this is the only one that has been thoroughly documented. However, when analyzed along comparative Semitic evidence, the data results reveal that the grammatical features of CIs in LA seem to be in line with general Semitic trends that, interestingly enough, do not always find their parallel in prescriptive descriptions of CIs in Classical or Standard Arabic.

By using the CIs in LA as a case study, this study attempts to shed light on the general benefits of a cross-Semitic analysis in the study of Arabic linguistics. As it was the case with CIs, the comparative Semitic evidence available may very well help researchers elucidate a broader, more inclusive vision of the grammatical nature of specific linguistic features, and assist them in the challenging task of revisiting traditional classifications to ensure accuracy in typological descriptions of Arabic varieties.

2. Materials and Methods

The present paper is based on an extensive study of CIs in LA based on a corpus of 133 recorded instances collected over 4 years of participant observation in Lebanon for the author’s doctoral dissertation ([Iriarte Diez Forthcoming](#))¹ as well as on a thorough review of the literature on CIs in Semitic languages.

Participant observation is the main method used by anthropologists. Fieldwork through this method requires “active looking, improving memory, informal interviewing, writing detailed field notes, and perhaps most importantly, patience” ([DeWalt and DeWalt 2002](#), p. vii). The author’s long stay in Lebanon (almost 10 years) facilitated the ‘prolonged engagement’ ([Lincoln and Guba 1985](#)) necessary to the creation of the CI corpus. This method was chosen for being the only one able to face the methodological challenges that the study of CIs presented, namely, (1) their marked interactional and emotional nature, (2) the variety of linguistic, social, and communicative contexts in which they occur, and (3) the extreme difficulty of eliciting them². Furthermore, as in previous studies, participant observation considerably diminished the effects of the so-called observer’s paradox ([Labov et al. 1968](#); [Milroy 1977](#)).

Comparison is essential for any researcher to acquire a real and profound knowledge of the true nature of a linguistic feature. In fact, comparison is not only useful “to relativize a phenomenon that we tend to consider as outstanding” but also necessary “to understand the role of the specific grammar of a dialect in leading to a type of evolution” ([Ibrahim 2011](#), p. 128). In this spirit, I carried out a thorough review of the relatively scarce literature on CIs. The purpose of this review was to gain awareness of the formal and functional variation that CIs show along the Arabic and Semitic continuums in order to be able to better evaluate (1) the morphological and syntactical factors that effectively differentiated CIs from COs, and (2) the pragmatic and discursive factors that could hold the key to a deeper understanding of the CI’s function.

As for the data presentation, for the purposes of this paper, the LA examples gathered in the corpus will be numbered and preceded by the letters LA (Lebanese Arabic) (i.e., [LA.n]). This system will help the reader differentiate CI and CO examples in LA from

The following subsections will address, in more detail, some of the formal features of CIs in LA in light of Semitic evidence. The morphological features include form and pattern of CIs and pattern correspondence between CIs and CHs. The syntactic features include CIs’ syntactic case, CIs’ position in the sentence, and the presence of enclitics.

3.1. Morphological Features of CIs in LA in the Light of Semitic Evidence

3.1.1. Infinitival Form of CIs

I have previously illustrated the formal differentiation between CIs and COs in LA as far as the choice of the ‘infinitival form’ is concerned. When looking at CI instances in other Semitic languages, the data consistently show a choice of morphological infinitives,⁵ which are formally and functionally distinct from the cognate nouns used in CO constructions (the latter being generally identifiable by an *-a/-e(t)* ending)). Table 2 illustrates this formal distinction. In the table, both CIs and COs are marked in bold and cognate heads are underlined.

Table 2. CI and CO instances in different Semitic varieties.

CI Instances	CO Instances
[SYR.1] Syriac meštaq [CI] <u>šteq</u> ^w ‘They were completely silent’ (Robinson and Coakley 2013, p. 66)	[SYR.2] Syriac <u>mīt</u> [CH] mawtā bišā wa-mṭarpā [CO] ‘He died an evil and painful death’ (Sim. 333, 3 from Nöldeke 2003, p. 237)
[MAL.1] Maltese <u>johrog</u> [CH] hrug [CI] ‘he goes out extensively’ (Maas 2005, p. 416)	[MAL.2] Maltese <u>ghajtu</u> [CH] ghajta ta’ ferh [CO] ‘They shouted a shout of joy’ (Sutcliffe 1960, p. 169)
[MEH.1] Mehri <u>yishōt</u> [CH] hābū saḥt [CI] (lit. he slaughters people slaughter) ‘he absolutely slaughters people [with his prices]’ (Watson 2012, p. 215)	[MEH.2] Mehri <u>kaṭays</u> [CH] mān kaṭāt kannaṭ [CO] (lit. he cut her a little cut) ‘he cut her lightly’ (Rubin 2010, p. 219)
[AKK.1] Akkadian [š]a ta[q]biam/ana PN/ana ešrišu aqbišim-ma/ apālum-ma [CI] <u>ul ṭipulanni</u> [CH] ‘[Wh]at you to[ll]d me I told PN ten times but answer me she did not’ (AbB 10, 8:16-19 from Cohen 2004, p. 107)	[AKK.2] Akkadian mīnam ēpuškāma ḥa-lu-qām ra-bi-a-am [CO] <u>tuḥallaqanni</u> [CH] ‘what have I done to you, that you are completely ruining me’ [lit. *that you are ruining me a big ruining] (Kouwenberg 2017, p. 653)
[BH.1] sāqōl [CI] <u>yissāqel</u> [CH] haššōr ‘the ox shall be stoned’ (Exod. 21:28 from Van der Merwe et al. 1999, p. 159)	[BH.2] way-yeh ^e rad [CH] Yiṣḥāq ḥ^arādā gdolā [CO] ‘And Isaac trembled a very great trembling’ (Gen. 27:33 from Goldenberg 2013, p. 295)

In Biblical Hebrew (BH), not only is a morphological distinction made between CIs and COs, but the infinitive used for CI constructions is also a special form of infinitive called ‘Infinitive Absolute’, which contrasts with the Infinitive Construct⁶. The distinction between these two forms is both morphological and syntactic⁷.

In contrast to BH, Classical and Modern Standard Arabic seem to be the most outstanding exception to the Semitic constant that morphologically distinguishes between infinitival forms for CIs and COs. In these varieties, both COs [CA.1] and CIs [CA.2] may be formed with a *maṣdar* (infinitive), although the use of a noun of single instance (NSI) in CO constructions is also accepted [CA.3]:

[CA.1] Classical Arabic (CI instance)

qumtu	qiyām-an
(PFV-stand.1S)	stand.INF-ACC)

[lit. I stood standing]

(Al-Zamaxšari 1870, p. 111; my glossing)

[CA.2] Classical Arabic (CO instance with *mašdar*)

qumtu	qiyām-an	ṭawīl-an
(PFV-stand.1S)	stand.INF-ACC	long.M.S-ACC)

[lit. I stood a long standing]

(Al-Zamaxšari 1870, p. 111; my glossing)

[CA.3] Classical Arabic (CO instance with NSI)

qaʿada	qaʿda-ta	al-qurfuṣāʿ
(PFV-sit.3MS)	sitting-NSI	DET-squatting position)

[He squatted]

(Sibawayhi: 112; my glossing)

In fact, the shared use of *mašdar* between CIs and COs was in all likelihood one of the main factors (along with syntactic case⁸) that led traditional Arabic grammarians to group these two phenomena under one single grammatical category, at variance with what we normally find in the descriptions of Semitic languages.⁹

3.1.2. Pattern Correspondence between CIs and CHs

In Semitic, there is in general a high degree of correspondence between the verbal pattern of the CH and the corresponding CI. The majority of the studies on the subject (Goldenberg 1971; Cohen 2004, 2006; Kouwenberg 2010) highlight the existence of an exact pattern correspondence between CHs and CIs in Akkadian. Kim (2006, p. 197) corroborates this fact by compiling 228 examples of CIs throughout the different stages of the Akkadian language, and Finet affirms that this is also true as well for the Mari dialect (Finet 1952, pp. 21–22). On the other hand, Syriac and BH follow different pattern correspondences, similar to those of Lebanese Arabic.

As for Classical Arabic, Talmon’s study on *mafʿūl muṭlaq* occurrences in the Qurʾān shows that 61/64 of the *mašādir* appearing in CI constructions¹⁰ share both the root and the pattern with their governing verb, reaching an ‘almost perfect’ pattern correspondence (Talmon 1999).

CIs in Lebanese Arabic generally share their pattern with their CHs when said heads are in pattern I (*faʿal*), II (*faʿʿal*), III (*fāʿal*), and X (*istafʿal*). In the cases of those patterns that carry passive, reflexive, or reciprocal values, such as V (*tfaʿʿal*) VI (*tfāʿal*), VII (*nfaʿal*), or VIII (*ftaʿal*), CHs take the cognate infinitive of their corresponding active pattern. The following examples illustrate pattern correspondence between CHs in patterns VII and V with their CIs in their active counterpart: Patterns I and II, respectively.

[LA.3] Lebanese Arabic

kēn-ət	mēšy-e	l-ʿilēʿa	bas	baʿdēn
was.3FS	walk.PTCP.ACT.-FS	DET-relation	but	afterwards
ma	b-aʿrəf	šu	šār	[pause]
neg	HAB- IPFV.I.know	what	happened.3MS	

ʔənʔaʔaʿ-ət	ʔaʔəʿ
was.cut-3FS [CH in pattern VII]	cut.INF [CI in pattern I]

‘The relationship was going [well] but then, I don’t know what happened ... it [suddenly] broke off’

[LA.4] Lebanese Arabic

ʔəza	ši	nhār	šəfti-h	hōn	ma
if	some	day	PFV.3MS.see-him	here	neg

txāfe b-yəṭmaššā təməšeyē
 IPFV.2FS.fear HAB-IPFV.3MS.stroll [CH in pattern V] stroll.inf [CI in pattern II]
 huwwe w-əl-kaləb tabaʕ-o la-yrēʔəb əl-bnūke bas
 he and-DET-dog GEN-him to-IPFV.3MS.guard DET-banks only
 ‘If you see him some other day, don’t be afraid . . . He just patrols with his dog to guard the banks [of the area], nothing else’

Detailed pattern correspondence between CIs and CHs for the verbs in my LA data is illustrated in Table 3 below.

Table 3. Pattern correspondence between CIs and CHs in LA and percentage of occurrence in the corpus.

CH's Pattern	CI's Pattern	Example	% of Occurrences
I	I	<i>maʔtūšē ʔaṭəš</i>	67.67%
II	II	<i>msatṭaltəsīl</i>	14.29%
III	III	<i>tseʔdūne msēʔade</i>	0.75%
V	II	<i>btəṭlaxxaš təlxīš</i>	2.26%
VI	III	<i>txēna ʔo xnēʔ</i>	0.75%
VII	I	<i>byəṇmašā maše</i>	3.01%
VIII	I	<i>nəštəğəla šəğəl</i>	0.75%
X	X	<i>staʔbalto stiʔbəl</i>	0.75%
QI	QI	<i>farfaṭ farəʔta</i>	2.26%
	QII	<i>mʔarʔad tʔərʔod</i>	2.26%
QII	QI	<i>təhdalət təhdale</i>	0.75%
	QII	<i>təhdal təhdol</i>	1.50%
Exceptions			3.76%

These apparent ‘asymmetries’ seem to contrast with the ‘perfect’ or ‘almost perfect’ correspondence between CIs and CHs that is claimed to exist in certain Semitic languages.

Nevertheless, exceptions to this apparent ‘perfect correspondence’ in patterns between CH and CI have also been documented in Classical Arabic (CA) by grammarians, who do note that some finite verbs in specific patterns might govern a *mašdar* of the same root but a different pattern [CA.4]:

[CA.4] Classical Arabic

wa-tabattal ʔilay-hi **tabtīl-an**
 (and-IMP.2MS.devote[CH in pattern V] to-him devotion.INF-ACC [CI in pattern II])
 ‘And devote thyself to Him whole-heartedly’

(Al-Zamaxšārī 1870, p. 111; my glossing)

Ibn Yaʿīš argues that in these cases, the two forms of the verb carry the same meaning (Al-Zamaxšārī 1870, p. 111). However, Sibawayh and Al-Mubarrad, among other grammarians, explain the lack of pattern correspondence as a consequence of the elision of the verb¹¹.

Interestingly, however, the logic for pattern correspondence followed by LA seems to be in line with that of Biblical Hebrew and Syriac.¹² In these two varieties, like in LA, the CH and the CI generally share the same pattern. However, Syriac passive verbal forms (i.e., *ethpʾel*; *ethpaʾal*; *ettaphʾal*) can take the infinitive of their corresponding active pattern. In the following example [SYR.3], the CH is in *ethpʾel*—passive counterpart of *pʾal*—and the CI is in *pʾal* pattern.

[SYR.3] Syriac

meḥzā [CI] ʾeṭḥāzā [CH] hʷāt leh šūr mtūm

‘il nʾavait jamais vu Tyr’ [he had never seen Tyre]

(Ined. Syr. 2, 14 from Duval 1881, p. 332; English translation mine)

As for BH, Gesenius’ grammar specifies that with a verb of the derived conjugations, not only the infinitive absolute of the pattern can be used, but also the *Qal* pattern as

“the simplest and most general representative of the verbal idea” (Cowley and Kautzsch 1910, p. 345). This is specifically common with verbs in the *Niph'al* pattern, the passive or reflexive form of the *Qal* pattern. In the following example [BH.3], the main verbs are in the *Niph'al* form while the CIs appear in the *Qal* form.

[BH.3] Biblical Hebrew

lo' tiggā' bō yād, kī sāqōl [CI] yissāqel [CH] 'ō yāroh [CI] yiyyārēh [CH]

“They are to be stoned or shot with arrows; not a hand is to be laid on them.

(NIV 2011, Exod. 19:13)

In light of this Semitic evidence, the logic of pattern correspondence between CI and CH in LA can be said to align with systems existing in other geographically adjacent Semitic languages, rather than simply ‘deviating’ from the idealized CA standard of ‘perfect’ correspondence.

3.2. Syntactic Features of CIs in LA in Light of Semitic Evidence

3.2.1. Case Marking on CI

Syntactic case is not marked in Lebanese Arabic; therefore, CIs are not marked with any syntactic case in this variety. This is also the case in all the spoken varieties of Arabic as well as in the great majority of Semitic languages, except for Akkadian, Ugaritic, and Classical Arabic—the only Semitic languages that are known to mark syntactic case with enough attested examples of CIs to render a syntactic analysis possible.¹³

Although syntactic case has no bearing on LA, looking at the case of CIs in case-bearing Semitic languages may provide further evidence on the distinct natures of CIs and COs, and it may shed light on the function of the CI across Semitic.

CIs in Classical Arabic (CA), such as [CA.1] and [CA.4], appear in the accusative case, generally marked with the indefinite accusative ending *-an*. In fact, the relevance of the syntactic case for the CA description of CIs is such that Sibawayhi and Al-Mubarrad decided to name this construction after its syntactic case (i.e., *maṣḍar maṣṣūb*, lit. ‘accusative infinitive’).

Ibn As-Sarrāj observes that the *maṣḍar* used to ‘strengthen’ the meaning of the action has to be in the accusative: “فإذا لم يكن فيها إلا التوكيد نصبت والرفع بعيد جدا” [if its only [function] is emphasis, then it [appears] in the accusative, for the nominative is too improbable] (Ibn As-Sarrāj 1985, p. 168; translation mine).¹⁴ Although Sibawayhi had also stated that CIs are accusative, a more detailed reading of his description reveals that he conceived both accusative and nominative as acceptable options and placed the decision in the speakers’ hands: “وكذلك جميع المصادر ترتفع على أفعالها إذا لم تشغل الفعل غيرها” [in the same way, any *maṣḍar* may be in nominative of their verb if (the verb is) not (syntactically) occupied with another (subject)] (Sibawayhi 1996, p. 229; translation mine).

This flexibility in the case marking of CIs falls in line with the general situation in the Semitic languages. Both *-u(m)/-u(n)*—generally associated with nominative in Semitic—and *-a(m)/-a(n)*—generally associated with accusative in the Semitic languages—appear to mark CIs in case-bearing Semitic languages. However, while accusative seems to be the norm in CA, Akkadian¹⁵ and Ugaritic present quite the opposite situation, as the following examples illustrate:

[AKK. 3] Akkadian

[ša i]štu šeḫrēku lā āmuru/[am]lārum-ma [CI] ātamar [CH]

“[[That wh]ich I have not seen [si]nce I was young I have seen now”

(AbB 11, 34:5-6 from Cohen 2004, p. 108)

[UG.1] Ugaritic¹⁶

l'akm [CI] 'il'ak [CH] [la'aku-ma 'il'aku]

“I will surely send”

(2.30, 19–20 in Sivan 2001, p. 123)

The predominance of the *-u(m)* ending for CIs' analogous forms in Akkadian and Ugaritic has triggered discussion about the ending's function. While some interpret it as a locative-adverbial case (Rosenthal 1942; Pope 1951; Huehnergard 2012), some preferred to think of it as a nominative (Driver 1956; Finet 1952; Kouwenberg 2017, p. 659)¹⁷, and others stayed neutral on the grounds of a lack of sufficient evidence (Goldenberg 1971; Bordreuil and Pardee 2009).

Be that as it may, a wider review of the literature on CI case marking that observes Arabic varieties as part of the Semitic continuum suggests that CIs are syntactically marked as salient entities and bearers of adverbial meaning rather than objects. This supposes additional evidence of the formal and functional differentiation of CIs and COs in Semitic varieties and presents the seemingly consistent accusative marking of CIs in Classical and Standard Arabic as an 'exception' within the Semitic continuum.

3.2.2. Presence of Enclitics

CIs in Lebanese Arabic do not present any kind of enclitics, as is the case in the majority of the other Semitic languages. The review of the literature shows, nonetheless, that in Akkadian and Ugaritic, CIs systematically present the enclitic *-ma/-m*, respectively.

In CI constructions in Akkadian (i.e., what scholars referred to as the *parāsum* (*-ma*) *iprus* type), the enclitic particle *-ma* often appears attached to the infinitives [AKK.1] [AKK.3].¹⁸ However, while the enclitic *-ma* seems to appear very frequently in Akkadian CIs, it is almost non-existent in COs.¹⁹ This 'emphasizing' particle, according to Huehnergard (1997, p. 325), marks the "logical predicate of a sentence" while for Buccellati (1996, p. 387) *-ma* "is more often than not associated with emphasis of limitation"²⁰—both functional features associated with CIs and not to COs.

The situation is similar in Ugaritic, where the so-called pronomastic infinitive (i.e., CI) appears with an enclitic *-m*. Pope (1951, p. 124) suggests that the enclitic *-m* attached to CIs in Ugaritic indicates "merely additional emphasis" and its omission or addition does not affect the meaning perceptibly. Scholars agree that the Ugaritic *-m* is related to the aforementioned Akkadian enclitic *-ma*.

[UG.3] Ugaritic

mtm [CI] 'amt [CH] [mātu-ma/mūtu-ma 'amūtu]

"verily I will die"

(1.17 VI, 38 in Sivan 2001, p. 124)

The presence of enclitics in CI constructions in Akkadian and Ugaritic—two of the oldest documented Semitic languages—points directly to the correlation between the use of CIs and the marking of logical predicates, limitation, and/or focus. Once again, a cross-Semitic analysis of formal features elucidates the functional nature of the CI that appears to be linked to information structure.

3.2.3. CIs' Position in the Sentence

The syntactic position of CIs has been the formal feature probably more widely studied in Semitic languages. Throughout the Semitic continuum, CIs are found both in post-verbal and pre-verbal positions.

The literature suggests that CIs consistently show a post-verbal position in most Arabic varieties, including Lebanese Arabic ([LA.4] [JA.1] [EA.1] [RPA.1] [CA.1] [CA.4]), with the exception of Sason Arabic [SSA.1, SSA.2] (Akkus and Öztürk 2017). Functionally speaking, it is worth mentioning that this position in many varieties of spoken Arabic is oftentimes reserved for focus of contrast (Brustad 2000).

[LA.4] Lebanese Arabic

məš maʕʔül šu b-təšbah-ik...
 NEG possible what HAB-IPFV.3FS.look.like-you.2FS

bəzəʔti-a [CH] **bəzəʔ** [CI]
 PFV.spit.2FS-her spit.INF

‘It’s incredible how much she looks like you . . . like two drops of water! [lit. you spat her spitting]’

[JA.1] Jordanian Arabic

il-bandora . . . ynaššfū-ha [CH] **tanšif** [CI]
 DEF-tomato dry-SBJV.3MP-3FS drying

‘As for tomatoes, they used to dry them properly’

(Personal communication from Bruno Herin)

[EA.1] Egyptian Arabic

nāyim [CH] fī *l-ʕasal* **nōm** [CI]
 sleep.PTCP in the-honey sleep

‘He is sleeping soundly’

(Woidich 2006, p. 269)

[RPA.1] Rural Palestinian Arabic

yixinkūna [CH] **xanīk** [CI]

‘they suffocate us completely’

(Shachmon and Marmorstein 2018, p. 32)

A post-verbal position is also typical of CIs in South Semitic languages such as Mehri [MEH.1] [MEH.3] and Ge’ez [GE.1]:

[MEH.3] Mehri

ḥšūr [CH] ḥābū **ḥāšar** [CI]

‘he wiped the people out’

(Watson 2012, p. 215)

[GE.1] Ge’ez

zabṭəwwo [CH] **zəbṭata** [CI]

‘They whipped him heavily’

(Lipiński 2001, p. 520)

However, this does not seem to be the case in other Semitic languages belonging to the Northwest and East Semitic groups, which tend to show a pre-verbal position, in some occasions accepting CIs both before and after the CH.

Akkadian, Phoenician, and Eblaite, along with Sason Arabic, regularly place the CI in a pre-verbal position. Kim (2006) argues that the pre-verbal position is the most common order in all the extinct Semitic languages while Solà-Solé (1961, p. 191) regards the post-verbal position as the original one. In Akkadian, the CI regularly precedes the main verbal form²¹, following the [CI+CH] order in affirmatives and [CI+neg+CH] in negative utterances. This is also the case in the few documented instances of CI in Eblaite [EB.1], Phoenician [PH.1], and Ugaritic [UG.1] [UG.2]:

[EB.1] Eblaite

pá-kà-ru [CI] a-pá-kà-ru [CH]

‘‘They should join firmly’’

(Lipiński 2001, p. 520)

[PH.1] Phoenician

ʾm **nhl** [CI] tnhl [CH] mgštk ʾlk wmgšt ʾly

“If you shall come into possession of it (the money), your share is yours and my share is mine”

(Krahmalkov 2000, p. 210)

[UG.2] Ugaritic

y^dm [CI] l y^dt [CH] [yadā^cu-ma lā yada^cta]

“verily you (m.s.) knew not”

(2.39, 14 in Sivan 2001, p. 123)

In the realm of currently spoken languages, an interesting case is that of Sason Arabic, where, in contrast to the post-verbal position of CIs in all other Arabic spoken varieties, both CIs [SSA.1] and COs [SSA.2] are canonically placed before the CH²²:

[SSA.1] Sason Arabic

ʃuʃa qar^f [CI] inqar^af [CH]

‘The glass broke a breaking’

[SSA.2] Sason Arabic

babe fadu-ma hedi [CO] infada [CH]

‘The door opened a slow opening’

(Akkuş and Öztürk 2017, p. 3)

However, despite researchers’ interest in classifying CIs according to their position, not all languages have clear preferences for pre-verbal or post-verbal positions—what is more, most of those where CIs are fairly well documented often exhibit both. Some examples of the latter are Biblical Hebrew, Syriac, Mandaic, and Northeastern Neo-Aramaic dialects (NENA)²³ where, according to the literature, CIs may be found both pre- and post-verbally. Examples from some of these languages are shown in Table 4:

Table 4. CIs in both pre-verbal and post-verbal positions in specific Semitic varieties.

Pre-Verbal CIs	Post-Verbal CIs
[SYR.5] Syriac meḥtā [CI] l mānā ḥt-ayt [CH] ‘Why hast thou then [so greatly] sinned?’ (Aphr. 270, 5 in Nöldeke 2003, p. 236)	[SYR.6] Syriac mramrmin-an [CH] mramrāmu [CI] l-āk ‘We extol thee (lit. we exalt you exalting)’ (Psalm 30 in David and Rahmani 1896, p. 424)
[NENA.1] North Eastern Neo-Aramaic ’āna zāla [CI] ḥar-zilōn [CH] biya ‘I have absolutely gone with it! (i.e., I am finished!)’ (Khan 2008, p. 731)	[NENA.2] North Eastern Neo-Aramaic ’εga lanwa briṯa [CH] ’ana braya [CI] ‘At that time I was not even born’ (Khan 2008, p. 732)
[BH.4] Biblical Hebrew ḥe ’akōl [CI] ’akalnū [CH] min hameleḳ ‘Have we eaten at all any of the king’s provisions?’ (2Sam. 19:43. NIV translation. See also Gen. 37:8; Isa. 50.2)	[BH.5] Biblical Hebrew šim ū [CH] šāmōa [CI] w-’al tābīnū, w-r’ū [CH] rā ō [CI] w-’al ted’ū ‘Be ever hearing, but never understanding, be ever seeing, but never perceiving’ (Isa. 6:9. NIV translation)

Given that in BH the pre-verbal (and often clause-initial) position of the CI seems to be the most frequent²⁴ and often regarded as the “basic structure” (Goldenberg 1971, p. 65), there is a vivid scholarly debate as to whether the post-verbal position is syntactically conditioned or not.²⁵ The situation is similar in Syriac (Hoffmann 1827, p. 341; Duval 1881, p. 333; Nöldeke 2003, p. 235)—while some believe that the post-verbal CI expresses a higher degree of emphasis (Nöldeke 2003, p. 236), others find that there is no difference in meaning between both variables (Duval 1881, p. 332 based on Barhebraeus).

Immersed in this process, both groups of scholars seem to have overlooked the possibility that pre-verbal and post-verbal CIs could be, in fact, two separate (although closely related) grammatical forms—with their corresponding functions—rather than two different manifestations of one single grammatical form. In the following section, I discuss how a comparative analysis of CIs’ position across Semitic languages such as the one I propose in this study could, in fact, shed light on the ‘evolution’ of both the formal and functional variation of cognate infinitives across Semitic languages.

4. Discussion

4.1. CI Position across Semitic Varieties: A Practical Discussion

In their attempt to understand CIs’ syntactic variation, many scholars traditionally identified the most frequent position in the varieties they studied, then proceeded to find the formal explanations of what factors may condition the occurrence of the ‘exceptions’ to the rule they had themselves drafted.

It is worth noticing that different approaches in the literature have inevitably resulted from the nature of the available data in each variety, but also from the feeding influences that diverse grammatical schools may have received and, especially, from the authors’ attitudes towards other related Semitic varieties.

Studies on the CI in most Semitic varieties have explored both positions as two different manifestations of the same form (Goldenberg 1971; Cohen 2004, 2006; Mengozzi and Miola 2018). As for Arabic, the few descriptions of CIs in Arabic varieties—where the grammatical tradition establishes the post-verbal position of the *mafʿūl muṭlaq*—hardly ever include examples of the so-called “extraposed” CIs, for they are considered to be, simply, a separate grammatical entity.

One of the few studies that actually includes such examples—Blau’s *Grammar of Christian Arabic*—rules out the possibility of extraposed CIs existing in Christian Arabic, and ascribes their occurrence to a Greek-Aramaic interference resulting, presumably, from poor translations (Blau 1967, p. 605)²⁶:

[CRA. 1] Christian Arabic

سمعاً يسمعون ولن تفهمون وبصراً تبصرون ولن تنظرون

“hear indeed and understand not, and see indeed and perceive not!”

(Blau 1967, p. 604)

Contrary to Blau’s opinion, although extraposed CIs are rarely documented, they seem to be indeed used in Spoken Arabic. The followings are some examples from Lebanese and Najdi Arabic:

[LA.ext1] Lebanese Arabic

wəqraye [ext.CI] kenna neqra qimet sa’a u-nəss ben-nhār bel-qeşaṣ wer-rwāyāt el-ğrāmiye

“Notre travail durait environ une heure et demie par jour et consistait dans la lecture d’histoires amusantes et romans d’amour” [our daily work would last around an hour and a half and it would entail reading entertaining stories and romantic novels]

(Feghali 1935, p. 10; translation mine)

[NA.1] Najdi Arabic

hawāš rebuking.INF hāwaš-t-ih rebuked-I-him

‘As far as rebuking is concerned, I have rebuked him’

(Ingham 1994, p. 43)²⁷

Oftentimes, these extraposed CIs may appear followed by a *-w* that introduces the CH and contributes to the expression of topicalized enumerations in Lebanese and Egyptian Arabic:

[EA.2] Egyptian Arabic

bōs wi-bosti, **hizār** wi-hazzarti, **liḥb** wi-liḥḥbti

As for kissing, you kissed. As for flirting, you flirted. As for playing, you played.²⁸

(Movie: *El nōm ft-l-ʿasal* ('Sleeping in Honey').)

[LA.ext2]

akal		w-akalna	raʿaṣ	w-raʿaṣna
eat.INF		and-PFV.eat.1S	dance.INF	and-PFV.dance.1S
ma	fi	ši	ma	ʕamelnē
NEG	there.is	thing	NEG	PFV.do.1S.it.M.SG

'We ate, we danced . . . there is nothing we did not do!'

As the previous examples show, at least in Spoken Arabic, extraposed CIs seem to function as regular topics. The infinitive in this case is the chosen form for the topicalization of the finite CH.²⁹ Extraposed CIs could thus be simply considered Infinitival Cognate Topics³⁰, as Ingham (1994, p. 43) suggested, which, as infinitives in topic positions “can be used to encode states of affairs as topics” (Maslova and Bernini 2006, p. 83).

In this sense, it is my impression that, at least in Spoken Arabic, these extraposed CIs, for which we adopt Ingham’s term “Infinitival Cognate Topics”, have a more accentuated nominal character than that of (post-verbal) CIs. For this reason, perhaps, it is common to find in extraposition those infinitives that have been almost completely nominalized (e.g., *akal*, *raʿaṣ*, *qrāye*, etc.).³¹ This, along with the function of topic, would differentiate them (but not necessarily isolate them) from the ‘canonical’ CIs described in this study.

Were the case of Spoken Arabic applicable to other Semitic languages, there would be a possibility that the existence of both pre-verbal and post-verbal positions of the CI are simply a manifestation of two closely related grammatical forms. One would be a reduplication of the verb that has been fronted, therefore topicalized, while the other represents the reduplication of the verb that focuses on the event expressed by the CH.

In this case, although the joint analysis of Infinitival Cognate Topics and CIs that has often been adopted in the literature—originally based on an excessive concern for the ‘form’ to the detriment of function—might raise some doubts from a functional perspective, it would be also understandable given the lack of extensive and comprehensive data available in most of the studied varieties.

Another difficulty that might have added to the typological confusion is that the line between the pragmatic notions of topicalization and focus is not only thinner than what it seems, but also, practically imperceptible for scholars working with written texts and consequently deprived from any information regarding the communicative contexts of the utterances in question. With such thin a line, it is not surprising that topics and focus sometimes overlap.

In fact, the function of topicalization (normally assigned to pre-verbal CIs) could have overlapped with that of focus (normally assigned to post-verbal CIs) under the umbrella of CIs in languages such as Old Babylonian (Cohen 2004)—which exhibited two functions but only the pre-verbal extraposed position.

4.2. Data, Ideologies and Their Role in the Creation of Typologies: A Theoretical Discussion

A review of analogous phenomena along the Semitic continuum provides us with invaluable comparative evidence that reveals that the grammatical features of CIs in LA seem to be in line with general Semitic trends that, interestingly enough, do not always find their parallel in prescriptive descriptions of CIs in Classical or Standard Arabic.

Consequently, by taking the analysis of CIs in LA in light of Semitic evidence as a case study, this paper’s results invite the reader to ponder on several theoretical and

methodological questions of relevance for the field of Arabic linguistics in general and for that of typology in Arabic dialects in specific.

4.2.1. Data

Scholars working with written texts, particularly those working with extinct varieties, are often deprived of essential information regarding the communicative contexts of the analyzed utterances (e.g., discursive context, communicative setting, communicative intentions and priorities, speakers' stance, etc.), for the data generally tend to have a performative character, rather than an interactive one.

This leads to the concern that both the amount, but especially the quantity, variety, and contextualization, of linguistic data available in certain varieties may be insufficient for a functional approach. Given that most grammatical descriptions of standard norms rely on this kind of data, typologies may also be of a predominantly form-based nature and therefore neglect important functional aspects.

The joint classification of the functionally distinct categories of CIs and COs in Classical and Standard Arabic is just one example of the potential consequences of an excessive reliance on form over function for grammatical description. As the study of CIs across Semitic languages illustrates, the subsequent grammatical labels that arise from such descriptions and that are systematically attributed to analogous linguistic phenomena across varieties may hamper the typologists' task of identifying functional cross-dialectal and cross-linguistic patterns, compromising the linguistic accuracy of typologies.

Consequently, the aforementioned considerations would compel researchers to reflect on the following methodological question: In reconsidering old typologies, what kind of data should be considered for the creation of new typologies?

4.2.2. Cross-Dialectal and Cross-Semitic Approaches for Descriptive Purposes

More often than not, the amount of available data at our disposal might be neither enough nor have the desired quality. While questioning the representativeness of the data behind the literature remains an academically healthy and necessary exercise, the reality is that oftentimes the available tools are relatively scarce and limited.

This, however, does not mean that researchers cannot optimize the worth of these resources. Adopting a comparative approach, even when the purpose of the research is not comparative *per se*, may be an excellent strategy to broaden the researcher's scope—particularly when facing the scarcity of available data—thus improving the quality of the description.

The study of CIs in light of a comparative analysis of the data available in Semitic varieties has:

- (1) Formally differentiated CIs and COs as different grammatical features and thus questioned traditional Arab grammarians' categorization of CIs and COs under the label of *maf'ūl muṭlaq*.
- (2) Shed light on the ways that the functioning of CIs in LA aligns with that of analogous features in other Semitic varieties.
- (3) Provided us with valuable hints and theoretical leads for a better understanding of the functional nature of CIs in LA—one that goes beyond the overused and linguistically vague notions of 'emphasis' and 'intensification' and points to information structure.
- (4) Given a more holistic understanding of the feature that allows us to build more educated theories regarding the development of this feature.

When considering the Semitic evidence for the study of a linguistic feature in a non-standard Arabic variety, both formal and functional patterns of use become more discernible. Catching sight of these patterns provides researchers with a more holistic understanding of the feature in question and enables them to create linguistic models and descriptions with potential cross-dialectal applications.

4.2.3. Ideologies

According to one characterization of the concept of ‘language ideologies’, these include “speakers’ consciousness of their language and discourse as well as their positionality (in political economic systems) in shaping beliefs, proclamations, and evaluations of linguistic forms and discursive practices” (Kroskrity 2004, p. 498; Kroskrity 2000).

Previous studies have revealed that the ideological biases in linguistic scholarship have tangible effects on practices such as linguistic mapping and/or on the interpretation of historical linguistics (Irvine and Gal 2000).

The thorough reading of the Semitic literature carried out for this study also uncovered two different types of ideologies that may have very possibly affected the accuracy of some of the descriptions of CIs, namely (1) ideologies toward the linguistic feature and (2) underlying ideologies toward linguistic varieties and their grammatical traditions:

- (1) *Ideologies toward the linguistic feature*: Given its reduplicative character, some scholars have often treated CIs as redundant, literally as mere “ornaments” (Guismondi 1913, p. 65) or as a “purely rhetorical” complementation (Krahmalkov 2000, p. 210). These ideological biases have been enough for some to consider CI features not worthy of systematic analysis. This functional stance is also quite present in the underlying implications of other qualifiers that have been traditionally used in the literature to name CIs—such as ‘paronomastic’, which implies some kind of pun or play on words, or *tautological*, which directly implies that this infinitive is not necessary and thus “syntactically and pragmatically insignificant” (Callaham 2006, p. 4).³² These ideologies have had a rather tangible effect on grammatical descriptions on CA and MSA, where, in the name of eloquence, the use of CIs is often said to be appropriate only in cases where the meaning of the action is doubtful or vague. Consequently, expressions such as *أَكَلَ أَكَلًا ?akala ?aklan* (lit. ‘He ate an eating’) or *قَعَدَ قَعْدًا qa?ada qu?ūdan* (lit. ‘he sat a sitting’), although grammatically correct, are considered by some grammarians as ‘rhetorically weak’, since the meaning of the verbs *أَكَلَ ?akala* (to eat) or *قَعَدَ qa?ada* (to sit) are not in a situation of uncertainty or doubt (Hasan 2009, pp. 326–27).

However, in sentences such as *طَارَت السَّمَكَةُ فِي الْحَوِّ طَيْرَانًا ?arati s-samka fi-l-?aww ?ayarānan* (lit. ‘the fish flew a flight in the air’), however, the use of the CI is justified by the bizarreness of the meaning (Hasan 2009, p. 327; translation mine). A fairly quick look at the available data, however, shows that this description of CIs is not usage-based, but rather ideology-based.

- (2) *Ideologies toward linguistic varieties and their grammatical traditions*: In spite of the formal and functional similar nature of CIs all across Semitic languages, certain analyses of CIs across Semitics show traces of ideological biases that can lead to typological inaccuracies.

Goldenberg (1971), for instance, wrote a seminal paper on CIs—which he referred to as Tautological Infinitives—in Biblical Hebrew and presented a classification of this feature with the help of comparative Semitic data. His rather detailed cross-linguistic classification clearly distinguishes “Tautological Infinitives (TI)” (Cognate Infinitives) from “Inner object constructions” (Cognate Object constructions).

Despite the productivity of CIs in different Arabic varieties (illustrated throughout the present study), Goldenberg seems to show some reticence at including Arabic as one of the languages where Tautological Infinitive constructions occur, for he considers the grammatical concept of ‘*mafʿūl muṭlaq*’ a mere synonym of his notion of ‘Inner Object’ (CO).³³ Although, in absolute terms, his description does account (albeit briefly and only in the final pages) for “exceptional” examples of what could be type A and type B Tautological Infinitive constructions in different Arabic varieties, one cannot help but notice that the scanty Arabic data are treated and analyzed with certain skepticism.³⁴

The different treatment of Arabic varieties in this classification—despite the abundant examples of CI available in CA and MSA—could be explained by either (a) an excessive reliance on the joint grammatical traditional label of *mafʿūl muṭlaq*³⁵ or by (b) insufficient research on CIs in the different Arabic varieties (including the spoken varieties) and/or an overgeneralization of the scarce available data.

Be what it may, both of these factors, probably fueled by ideological biases, may have led to an excessive reliance on formal features (in this case, on syntactic order and case) that in turn, resulted in typological inaccuracies.³⁶

The results of this study suggest that the use and functioning of CIs is fairly similar across Semitic languages with clearly identifiable patterns of formal variation. The divergence between CIs in different Semitic varieties lies, to a great extent, in the different ideological approaches used for its analysis rather than in its linguistic nature.

Once we, as Arabic dialectologists, acknowledge the profound ways in which language ideologies can shape presumably “objective” linguistic analyses, it thus becomes imperative to pose the question: In revising the typologies of the grammatical tradition, how can we identify and set aside language ideologies to ensure accuracy in future typological descriptions of Arabic varieties and its shared features?

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of the American University of Beirut (protocol ID: KS1.06).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Acknowledgments: I owe a huge debt of gratitude to Kristen Brustad for her always valuable comments and corrections. I thank also Stephan Procházka and the two anonymous peer reviewers for their close readings and helpful remarks. All remaining errors are my own responsibility.

Conflicts of Interest: The author declares no conflict of interest.

Notes

- ¹ This paper is a concise version of the first chapter of my doctoral dissertation: *The Communicative Grammatical Function of Cognate Infinitives in Lebanese Arabic*, Zaragoza (Spain): University of Zaragoza Press (forthcoming), which aims to elucidate the communicative grammatical function of Cognate Infinitives (CI) in Lebanese Arabic (LA) in light of socio-cognitive and functional-pragmatic linguistic theories.
- ² In his study on *Syntactic Reduplication in Arabic*, Maas already highlighted that the oral and spontaneous character of the Cognate Infinitive makes its elicitation a “troublesome task” (Maas 2005, p. 417).
- ³ Christian Arabic is a main source for the study of Middle Arabic. The majority of the non-standard texts written in Arabic by Christians have been found in the South of Palestine and the Sinai and go back to the 8th century CE. The language of these texts was less influenced by the literary variety. However, we must bear in mind that most of the texts are translations from Greek or Syriac, which might create confusion when trying to discern between interferences from the vernacular and those from the translation’s original language (Versteegh 2014).
- ⁴ The term ‘cognate head’ has been taken from (Bond and Anderson 2014).
- ⁵ The Arabic grammatical notion of *maṣḍar* integrates both the notions of “infinitive” and of “verbal noun” simultaneously, given that a *maṣḍar* has both a verbal and a nominal nature. Regarding this, Talmon (1999) maintains that when the cognate *maṣḍar* is followed by a qualifier—this is, in a CO, construction—the substantival character of the *maṣḍar* comes forward. In contrast, when the cognate *maṣḍar* appears undefined and acts as an emphazier—that is, in a CI construction—the *maṣḍar* shows an infinitival character. It was precisely this verbal infinitival character that the *maṣḍar* shows in CI constructions that motivated me to choose the term “infinitive” over that of “verbal noun” or of “*maṣḍar*”—which would have been, in my opinion, less grammatically accurate and subject to ambiguity as well as potentially less recognizable for non-specialists of Arabic grammar terminology.
- ⁶ The Infinitive Construct has the same form as the masculine singular imperative (e.g., /k-t-b/ktob) while the Infinitive Absolute is characterized by the appearance of a long ‘o’ (e.g., /k-t-b/katob). As for the origin of these forms, Waltke and O’Connor (1990, p. 581) argue that these two forms are “historically distinct and unrelated”—while the *Qal* Infinitive Absolute of BH *qātōl* finds its origins in proto-Semitic *qatāl, the Infinitive Construct developed from the Semitic nominal pattern *qtul or *qutul.

Kim (2006, p. 223), however, argues the opposite: “As far as the evidence goes, in these languages [Semitic] the tautological and non-tautological infinitives share the same form, supporting the view that the Hebrew infinitive absolute and construct developed from a single form” (Kim 2006, p. 23).

7 Interestingly enough, the morphological distinction between these two infinitives seems to be an innovation of BH, given that it cannot be traced back in the Semitic continuum (Waltke and O’Connor 1990, p. 581; Bordreuil and Pardee 2009, p. 56).

8 See Section 3.2.1. of this study.

9 In fact, Sibawayhi (760-796 CE) originally referred to this notion as *mašdar maṣṣūb*, although later this feature became commonly known by the term coined by Ibn As-Sarrāj in the 9th century: *al-maḥṣūl al-muṭlaq*.

10 Talmon, like most traditional grammarians, refers to CI constructions as *maḥṣūl muṭlaq muḥḥam*, which is considered a subcategory of the grammatical category of *maḥṣūl muṭlaq*, along with *maḥṣūl muṭlaq muxtaṣṣ* (which would be equivalent to CO constructions).

11 This process would also explain the label of *maḥṣūl muṭlaq* that traditional Arab grammarians apply to constructions where the verb and the *mašdar* have completely different roots but carry similar meanings (e.g., أَبْغَضْتَهُ كَرَاهَةً? *abğaðtuhu karāhatan* [lit. I despised it/him a hatred] (Al-Zamaxšarī 1870, p. 112).

12 I would go as far as affirming that the patterns are identical if it not for the fact that the documented data in these varieties are, unfortunately, not enough for me to make an empirical claim.

13 Nominative and accusative cases were also attested in Amharic and Tigrinya, and accusative in Ge’ez. Other Semitic languages such as Hebrew and most varieties of Aramaic have a system of *differentielle Objekt markierung*.

14 Categorizing a *maḥṣūl muṭlaq* as such was, in fact, oftentimes exclusively dependent on the syntactic case of the *mašdar*, to the extent that neither he nor other grammarians stipulate that the verb and the *mašdar* should share the same root. As a matter of fact, Arab grammarians provided us with an exhaustive description of various cases where the *mašdar maṣṣūb* stands alone after the elision of the verb. However, this analysis falls out of the scope of our study for those constructions do not show an explicit verbal root repetition.

15 See also [AKK.1].

16 When dealing with Ugaritic, it should be born in mind that the final ending of the infinitive is only discernible in III-?roots (i.e., roots whose last radical is /ʔ/). In these cases, paronomastic infinitives (CIs) show an *-u* ending.

17 While for Kouwenberg CIs appear in nominative because this is functionally the unmarked case in Akkadian, for Finet the use of the *-u(m)* ending in CIs only confirms the marked usage of the nominative to mark a *casus pendens*, i.e., a fronted, topicalized nominal: “Ceci est conforme à l’usage, même abusif, de ce cas pour mettre en valeur le mot important de la phrase” (Finet 1952, p. 22).

18 See examples [AKK.1] and [AKK.3] for instances of CIs with the enclitic *-ma*.

19 Only two instances where *-ma* is attached to an accusative CI (i.e., the *parāsam iprus* type) have been documented. See (Cohen 2006, p. 428).

20 In fact, Buccellatti claims that *-ma* would be precisely the best option to translate the English ‘just’.

21 With the exception of the utterance: *at-ta-kil* [CH] *ta-ka-lu* [CI] “I trusted” (Kim 2006, p. 192; Rapallo 1971, p. 108).

22 Sason Arabic is heavily influenced by non-Semitic languages with typical SOV word order such as Kurdish and Turkish.

23 Moreover, the different Spoken Aramaic dialects are an excellent example of variation on the CI’s position. In Ṭurōyo, we only find pre-verbal CIs. However, in some NENA dialects such as that of Barwar, we can find both pre-verbal and post-verbal CIs, while in others, such as that of Qaraqosh, only post-verbal CIs occur (Mengozzi and Miola 2018).

24 Kim (2009, p. 46) notes that the most frequent order of the pattern is [CI + CH], but also identifies thirteen occurrences (out of 224) that present CH + CI order.

25 Some scholars who agree with the former (Cowley and Kautzsch 1910, p. 342; Van der Merwe et al. 1999, p. 158; Goldenberg 1971, p. 64; Harbour 1999; Kim 2009, p. 46) enumerated a series of syntactic conditions under which the CI cannot precede a verb. Joosten (2009, p. 106) observes that in BH, only postponed infinitives are attested with imperatives and volitive forms. Cf. Hatav (2017, p. 226) and Kim (2009, pp. 46–50) who suggest that the syntactic environment is not solely responsible for the change of order of the constituents in CI constructions.

26 Blau’s reasoning leaves original CRA examples such as *أم مزاح تمزح؟* (“or are you joking?”) unexplained (Blau 1967, p. 605). Moreover, given that similar constructions of topicalized infinitives are also readily available in the spoken varieties of Arabic, Blau’s argument remains, in my opinion, a questionable one.

27 Ingham refers to extraposed CIs as ‘Cognate Topics’.

28 I thank Prof. David Wilmsen for drawing my attention to this instance of Egyptian Arabic through a personal communication.

29 This is, according to Bernini, a common practice in a variety of languages. “Many languages tend to resort to inflected forms with the least amount of specification with respect to the major variables of speech act form and topic time, such as the infinitive forms of Italian, Yiddish, Russian and German.” (Bernini 2009, p. 113). For more examples of extraposed CIs in the world languages see (Mengozzi and Miola 2018, pp. 272–79).

30 In line with type A of CIs in Biblical Hebrew (Goldenberg 1971) and Old Babylonian (Cohen 2004).

- 31 “In Indo-European and in other languages, forms of this kind are removed from the prototype of the verb category and overlap with nouns in many aspects of their behavior” (Bernini 2009, p. 113).
- 32 These views are also representative of the current attitudes of LA native speakers towards the feature, who consider the use of CIs in LA as “a mistake” or a “dialect thing”. The acceptability tests of this feature carried out with LA native speakers revealed that this attitude towards CIs made many of my informants feel ashamed when asked about specific CIs that they had uttered, and sometimes even deny having used CIs at all (Iriarte Díez Forthcoming).
- 33 Goldenberg argues that “it is essential to the accuracy of the description to distinguish constructions with an “inner object” (or “internal”, or “general”, or “absolute”, or “cognate” object, or المفعول المطلق from those with a “tautological infinitive” of either type A, B, or C” (Goldenberg 1971, p. 76).
- 34 On two examples of type A TIs in Classical Arabic and a variety of Palestinian spoken Arabic: “[...] those instances, whose genuineness is above suspicion . . . ” (Goldenberg 1971, p. 77). On the possibility of Arabic having TIs: “When المفعول المطلق is a مصدر مبهم (indefinite infinitive) and implies hardly anything more accurately definable than the vague تأكيد (“strengthening” or “emphasis”) it corresponds apparently to some uses of the inf.-constr. of type C” (Goldenberg 1971, p. 77). Affirming that the claims that regard certain type A TIs in Arabic should “not be regarded as as simple intrusion of a structure completely alien to the nature of the Arabic language” (Goldenberg 1971, p. 78) is as far as Goldenberg goes when discussing the possibility of original type A and B TI examples existing in Arabic varieties.
- 35 I find this possibility unlikely, given that he did not do so with other varieties such as Syriac, whose traditional grammar also shows a joint analysis. Goldenberg (1971) even finds it “regrettable” that certain grammarians in other Semitic varieties (not Arabic) have failed to make the distinction between CI and CO in their works, which shows his ability to recognize CIs and COs under one grammatical label: “It is regrettable that Semitists like Nöldeke, Brockelmann and Reckendorf have failed to make the necessary distinction, and in fact, in their treatment of the relevant constructions confusion prevails.” (Goldenberg 1971, p. 77).
- 36 This would explain why, years later, Goldenberg would include examples of CIs in Written Arabic under the grammatical label of ‘Inner/Cognate Object’ and along examples of CO constructions—of a clearly different functional nature—from a variety of Semitic languages (Goldenberg 2013, p. 167).

References

- Akkus, Faruk, and Balkız Öztürk. 2017. On Cognate Objects in Sason Arabic. *University of Pennsylvania Working Papers in Linguistics* 23: 2.
- Al-Zamaxšari, Al. 1870. *Kitāb al-mufaṣṣal fī n-naḥw*. Edited by Jens Peter Broch. Christiania: Christianiae Libraria P.T. Mallingii.
- Bernini, Giuliano. 2009. Constructions with Preposed Infinitive: Typological and Pragmatic Notes. In *Information Structure and Its Interfaces*. Edited by Lunella Mereu. Berlin: Walter De Gruyter, pp. 105–28.
- Blau, Joshua. 1967. A Grammar of Christian Arabic, Based Mainly on South-Palestinian Texts From The First Millennium. In *Fasc. III: Syntax II*. Louvain: Secrétariat du CorpusSCO, vol. 279, pp. 257–468.
- Bond, Oliver, and Gregory Anderson. 2014. Aspectual and Focal Functions of Cognate Head-Dependent Constructions: Evidence from Africa. *Linguistic Typology* 18: 215–50. [CrossRef]
- Bordreuil, Pierra, and Dennis Pardee. 2009. *A Manual of Ugaritic*. Winona Lake: Eisenbrauns.
- Brustad, Kristen. 2000. *The Syntax of Spoken Arabic: A Comparative Study of Moroccan, Egyptian, Syrian, And Kuwaiti Dialects*. Washington, DC: Georgetown University Press.
- Buccellati, Giorgio. 1996. *A structural Grammar of Babylonian*. Wiesbaden: O. Harrassowitz.
- Callahan, Scott Nelson. 2006. The Modality of the Verbal Infinitive Absolute in Biblical Hebrew. Ph.D. dissertation, Southwestern Baptist Theological Seminary, Fort Worth, TX, USA.
- Cohen, Eran. 2004. Paronomastic Infinitive in Old Babylonian. *Jaarbericht Ex Oriente Lux* 38: 105–12.
- Cohen, Eran. 2006. The Old Babylonian Paronomastic Infinitive in -am. *Journal of the American Oriental Society* 126: 425–32.
- Cowell, Mark W. 1964. *A Reference Grammar of Syrian Arabic*. Washington, DC: Georgetown University Press.
- Cowley, Arthur Ernest, and Emil Kautzsch. 1910. *Gesenius' Hebrew Grammar*. Oxford: Clarendon Press.
- David, Joseph, and Aloysio Rahmani. 1896. *Grammatica Aramaica seu Syriaca*. Mausili: Typis Fratrum Praedicatorum.
- DeWalt, Kathleen, and Billie DeWalt. 2002. *Participant Observation: A Guide for Fieldworkers*. Walnut Creek: AltaMira Press.
- Driver, Godfrey Rolles. 1956. *Canaanite Myths and Legends*. Edinburgh: Old Testament Studies, No.3.
- Duval, Rubens. 1881. *Traité de Grammaire Syriaque*. Paris: Friedrich Vieweg.
- Feghali, Michel T. 1935. *Contes, Légendes, Coutumes Populaires du Liban et de Syrie: Texte Arabe, Transcription, Traduction Et Notes*. Paris: Adrien-Maisonneuve.
- Finet, A. 1952. Sur trois points de syntaxe de la langue des «Archives de Mari». *Revue d'Assyriologie et d'archéologie orientale* 46: 19–24.
- Goldenberg, Gideon. 1971. Tautological Infinitive. *Israel Oriental Studies (Tel-Aviv University)* 1: 36–85.
- Goldenberg, Gideon. 2013. *Semitic languages: Features, Structures, Relations, Processes*. Oxford: Oxford University Press.
- Guismondi, Enrico. 1913. *Linguae Syriacae: Grammatica et Chrestomathia Cum Glossario: Editio Quarta*. Rome: Excudebat carolus de Luigi.
- Harbour, Daniel. 1999. Two Types of Predicate Clefts: Classical Hebrew and Beyond. In *Papers on Morphology and Syntax: Cycle Two*. Edited by Vivian Lin, Cornelia Krause, Benjamin Bruening and Karlos Arregi. Cambridge: MIT Press, pp. 159–75.

- Hasan, Abbas. 2009. *Al-Naḥw Al-Wāfi Ma'a Rabṭihi Bi-L-³asālib Al-Raff'a Wa-L-Ḥayāt Al-Lughawiyya Al-Mutaḡaddida*. Cairo: Dar al-Ma'arif.
- Hataw, Galia. 2017. The Infinitive Absolute and the Topicalization of Events in Biblical Hebrew. In *Advances in Biblical Hebrew Linguistics: Data, Methods And Analyses*. Edited by A. Moshavi and T. Notarius. Winona Lake: Eisenbrauns.
- Hoffmann, Andreas Gottlieb. 1827. *Grammaticae syriacae libri 3: Cum tribus tabulisvaria scripturae aramaicae genera exhibentibus*. Halle: Orphanotropheus.
- Huehnergard, John. 1997. *A Grammar of Akkadian*. Atlanta: Scholars Press, vol. 45.
- Huehnergard, John. 2012. *An Introduction to Ugaritic*. Peabody: Hendrickson Publishers.
- Ibn As-Sarrāj, Abu Naṣr. 1985. *Kitāb al-'uṣūl fī n-naḥw*. Ed. 'Abd al-Ḥusayn al-Fatḥī. Beirut: Mu'assasat ar-Risāla.
- Ibrahim, Awad. 2011. Evolution of Expressive Structures in Egyptian Arabic. In *Modern Trends in Arabic Dialectology*. Edited by Mohamed Embarki and Moha Ennaji. Trenton: The Red Sea Press, pp. 121–38.
- Ingham, Bruce. 1994. *Najdi Arabic. Central Arabian*. London: John Benjamins Publishing.
- Iriarte Diez, Ana. Forthcoming. *The Communicative Grammatical Function of Cognate Infinitives in Lebanese Arabic*. Zaragoza: University of Zaragoza Press.
- Irvine, Judith T., and Susan Gal. 2000. Language Ideology and Linguistic Differentiation. In *Regimes of Language*. Edited by Paul V. Kroskrity. Santa Fe: School of American Research Press, pp. 35–83.
- Joosten, Jan. 2009. Three Remarks on Infinitival Paronomasia in Biblical Hebrew. van Bekkum, W. (2011). In Zaphenath-Paneah: Linguistic Studies Presented to Elisha Qimron on the Occasion of His Sixty-Fifth Birthday. *Journal for the Study of Judaism* 51: 99–113.
- Khan, Geoffrey. 2008. *The Neo-Aramaic Dialect of Barwar*. Leiden: Brill.
- Kim, Yoo-Ki. 2006. The Function of the Tautological Infinitive in Classical Biblical Hebrew. Ph.D. dissertation, John Hopkins University, Baltimore, MD, USA.
- Kim, Yoo-Ki. 2009. The Function of the Tautological Infinitive in Classical Biblical Hebrew. In *Harvard Semitic Studies*. Winona Lake: Eisenbrauns.
- Kouwenberg, Norbertus Johannes Cornelis. 2010. *The Akkadian Verb and its Semitic Background*. Winona Lake: Eisenbrauns.
- Kouwenberg, Norbertus Johannes Cornelis. 2017. *A Grammar of Old Assyrian*. Leiden: Brill.
- Krahmalkov, Charles R. 2000. *A Phoenician-Punic Grammar*, 1st ed. Leiden: Brill.
- Kroskrity, Paul. V. 2000. Regimenting Languages. In *Regimes of Language*. Edited by P. V. Kroskrity. Santa Fe: School of American Research Press, pp. 1–34.
- Kroskrity, Paul V. 2004. Language Ideologies. In *A Companion to Linguistic Anthropology*. Edited by A. Duranti. Oxford: Blackwell Publishing, pp. 496–517.
- Labov, William, Philip Cohen, Clarence Robins, and John Lewis. 1968. A Study of the Non-Standard English of Negro and Puerto-Rican Speakers in New York City. In *Report on Co-Operative Research Project 3288*. New York: Columbia University.
- Lincoln, Yvonna S., and Egon Guba. 1985. *Naturalistic Inquiry*. Beverly Hills: Sage.
- Lipiński, Edward. 2001. *Semitic Languages: Outline of a Comparative Grammar*. Leuven: Peeters Publishers.
- Maas, Utz. 2005. Syntactic Reduplication in Arabic. In *Studies in Reduplication*. Edited by Bernard Hurch. Berlin and Boston: Walter De Gruyter, pp. 395–430.
- Maslova, Elena, and Giuliano Bernini. 2006. Sentence Topics in the Languages of Europe and Beyond. In *Pragmatic Organization of Discourse in the Languages of Europe*. Edited by Giuliano Bernini and Marcia Schwartz. Berlin and New York: Mouton de Gruyter, pp. 67–120.
- Mengozzi, Alessandro, and Emanuele Miola. 2018. Paronomastic Infinitives in North-Eastern Neo-Aramaic: A Typological Approach. *Aramaic Studies* 16: 270–99. [CrossRef]
- Milroy, James. 1977. *The Language of Gerard Manley Hopkins*. London: André Deutsch.
- NIV—New International Version. 2011. *BibleGateway.com*. Available online: <http://www.biblegateway.com/versions/New-International-Version-NIV-Bible/#booklist> (accessed on 30 March 2021).
- Nöldeke, Theodore. 2003. Compendious Syriac Grammar. In *Translated from the Second and Improved German*. Edited by James A. Crichton. Eugene: Wipf and Stock Publishers.
- Pope, Marvin. 1951. Ugaritic Enclitic-m. *Journal of Cuneiform Studies* 5: 123–28. [CrossRef]
- Rapallo, Umberto. 1971. Tipologia dell'infinito paronomastico. *Archivio Glottologico Italiano* 56: 105–27.
- Robinson, Theodore Henry, and James Farwell Coakley. 2013. *Robinson's Paradigms and Exercises in Syriac Grammar*. Oxford: Oxford University Press.
- Rosenthal. 1942. Ugaritic Grammar. The Present Status of the Linguistic Study of the Semitic Alphabetic Texts from Ras Shamra. *Analecta Orientalia* 20: 171–79.
- Rubin, Aaron. 2010. *The Mehri Language of Oman*. Leiden: Brill.
- Shachmon, Ori, and Michal Marmorstein. 2018. badhačak dahič "I'll smash you altogether!" The Unmodified Cognate Complement in Rural Palestinian Arabic. *Zeitschrift für Arabische Linguistik* 68: 31–62. [CrossRef]
- Sibawayhi. 1996. 'Amr ibn 'Uthmān (d. ca. 177/793). In *Kitāb Sibawayh*. Beirut: Dar Al-Jil.
- Sivan, Daniel. 2001. *A Grammar of the Ugaritic Language (2nd impression with corrections)*. Atlanta: Society of Biblical Literature.
- Solà-Solé, José Maria. 1961. *L'infinitif Sémitique*. Paris: Honoré Champion Editeur.

- Sutcliffe, Edmund Felix. 1960. *A Grammar of the Maltese Language: With Chrestomathy and Vocabulary*, 3rd ed. Oxford: Oxford University Press.
- Talmon, Rafael. 1999. The Syntactic Category *Maf'ul Muṭlaq*: A Study in Quranic Syntax. In *Arabic Grammar and Linguistics*. Edited by Yaser Suleiman. Richmond: Curzon, pp. 107–29.
- Van der Merwe, Christo J., Jacobus A. Naudé, and J. H. Kroeze. 1999. *A Biblical Hebrew Reference Grammar*. Sheffield: Sheffield Academic.
- Versteegh, Kees. 2014. *The Arabic Language*. Edinburgh: Edinburgh University Press.
- Waltke, Bruce K., and Michael Patrick O'Connor. 1990. *An Introduction to Biblical Hebrew Syntax*. Winona Lake: Eisenbrauns.
- Watson, Janet. 2012. *The Structure of Mehri*. Wiesbaden: O. Harrassowitz.
- Woidich, Manfred. 2006. *Das Kairenisch-Arabische: Eine Grammatik*. Wiesbaden: O. Harrassowitz.

Article

An Innovative Copula in Maghrebi Arabic and Its Dialectological Repercussions: The Case of Copular *yabda*

Adam Benkato ^{1,*} and Christophe Pereira ²

¹ Department of Middle Eastern Languages and Cultures, University of California, Berkeley, CA 94720, USA

² Département Études Arabes, Institut National des Langues et Civilisations Orientales, 75007 Paris, France; christophe.pereira@inalco.fr

* Correspondence: abenkato@berkeley.edu

Abstract: Research on copulas in Arabic dialects has hitherto largely focused on the pronominal copula, and has also mostly ignored Maghrebi dialects. Drawing on published literature as well as fieldwork-based corpora, this article identifies and analyzes a hitherto undescribed verbal copula in dialects of Tunisian and northwestern Libya deriving from the verb *yabda* (“to begin”). We show that copular *yabda* occurs mostly in predicational copular sentences, with time reference including the habitual present and generic future. It takes nominal, adjectival, and locational predicate types. We also argue for broader inclusion of syntactic isoglosses in Arabic dialectology, and show how copular *yabda* crosses the traditional isogloss lines established on the basis of phonology, morphology, or lexicon, and therefore contradicts established dialect classifications such as Bedouin/sedentary or Tunisian/Libyan.

Keywords: Tunisian Arabic; Libyan Arabic; copulas; syntactic isoglosses; dialect classification

Citation: Benkato, Adam, and Christophe Pereira. 2021. An Innovative Copula in Maghrebi Arabic and Its Dialectological Repercussions: The Case of Copular *yabda*. *Languages* 6: 178. <https://doi.org/10.3390/languages6040178>

Academic Editors: Simone Bettega and Roberta Morano

Received: 17 August 2021
Accepted: 21 October 2021
Published: 26 October 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Arabic dialectology has largely focused until now on understanding the geographic distribution of varieties through socio-historical parameters. The traditional dialectological approach to the Arabic varieties of northern Africa (“Maghrebi” varieties) foregrounds a classification scheme which is organized not only along geographical lines, but one which also depends on ecological categories (“Sedentary” vs. “Bedouin”) as well as socio-historical ones (“pre-Hilali” vs. “Hilali”) (Caubet 2001; Palva 2006; Pereira 2011, 2018). While certain categories used for classifying Maghrebi Arabic varieties have recently been subject to critique from historical perspectives (Kosansky 2016 on “Judeo-Arabic”; Benkato 2019 on “Bedouin”), it has also been shown that the existing linguistic evidence does not necessarily support the utility of other categories.¹ Similarly, it can be pointed out that the existing classifications rely almost exclusively on phonological and morphological isoglosses, and to a lesser extent on lexical ones. Though neglect of morphosyntax for drawing isoglosses is typical of dialectology in general, the problem is particularly acute in Arabic dialectology in northern Africa. This is not only because morphosyntax is almost entirely ignored, but because regional variation in phonology and morphology can often be rather limited, meaning that dialect boundaries drawn on the basis of a handful of such isoglosses are not strong.

The dialectology of Maghrebi Arabic, therefore, could benefit not only from the continued interrogation of the traditional classification system but also from drawing on a broader set of data that includes previously unexamined linguistic features, particularly morphosyntactic ones. This study, by describing a syntactic feature and examining its consequences for dialectology, aims to show how such work has the potential to change the traditional map of Arabic in northern Africa. It opens by giving a brief overview of the copula in Arabic dialects (Section 2), before proceeding to the description of a hitherto unidentified copula in varieties of Tunisia and Libya (Section 3). The study then discusses

the neglect of syntax in Arabic dialectology and shows that syntactic isoglosses may conflict with isoglosses based on other linguistic features (Section 4).

2. Copulas in Arabic

Copula constructions are to be understood as constructions used to encode the identity of two participants and to express group membership, classification, location and the ascription of a range of properties to a participant and the element linking these is a copula. It is common to assume that the copula is lexically-semantically empty (Pustet 2003, p. 5) and that its main role is in semantic composition and in carrying tense/aspect (Roy 2013). The generally-accepted major types of copula construction are predicational, equative, specificational, and identificational (Higgins 1979, pp. 204–93; Mikkelsen 2011). For our purposes, an equative copula construction is one which equates the referents of the two elements besides the copula (Mikkelsen 2011, pp. 1807–8), while a predicational copula construction is one whose subject is referential and whose predicate is some non-verbal element, whether nominal, adjectival, or prepositional (Mikkelsen 2011, pp. 1808–9).

While languages vary greatly in terms of what elements provide copulas, and which constructions require overt copulas, we can state the following regarding how copula constructions in Arabic are typically viewed.² Predicational constructions with present reference usually use a zero copula (1–2). In Tripoli Arabic, copula constructions with zero copula describe facts and express general truths inthetic utterances, serving to present an entity, a proposal or a state of affairs that is new information.³ In such utterances, a state of being (an inherent or permanent characteristic of a being, as in the first example) or a current activity (including location, as in the second example), considered true by the speaker, is expressed.

(1)	Tripoli Arabic (Pereira corpus) ⁴		
šaxšiyt-a	∅		dʕiʕf-a
personality-3SG.M	∅		weak-F
“His personality is weak”			

(2)	Tripoli Arabic (Pereira corpus)		
ʕadnān	∅	f-əṣ-ṣaḥṛa	tawwa
Adnan	∅	in-DEF-desert.F	now
“Adnan is in the desert now”			

In examples (1–2) above, the zero copula is employed in phrases in which the speaker validates the predicative relation. The zero copula thus expresses realis/indicative. Copula constructions of all types which have temporal reference to the non-present, however, require an overt copula, usually provided by a form of the verb *kān/ykūn* “be” (3a). Moreover, if epistemic modality is to be expressed, the overt copula *ykūn* is required (Pereira 2010, pp. 453–67): in copula constructions with *ykūn*, the predicative relation is to some extent uncertain and the construction thus expresses irrealis/potential (3b).

(3)	Tripoli Arabic (Pereira corpus)		
(a)	ʕadnān	kān	f-əṣ-ṣaḥṛa
	Adnan	be\PFV[3 SG.M]	in-DEF-desert
	“Adnan was in the desert”		
(b)	ʕadnān	y-kūn	f-əṣ-ṣaḥṛa
	Adnan	3M-be\IPFV	in-DEF-desert
	“Adnan may be in the desert”		

To explicitly situate the copula construction in the future, the preverb *hā-* precedes the verb *ykūn* (Benmofteh and Pereira 2019). In the following example (4), we can compare the use of the zero copula with a generic present reference and the verb *ykūn* with a future reference.

(4)		Tripoli Arabic (Pereira corpus)				
	<i>ġudwa</i>	∅	<i>yōm</i>	<i>ždīd:</i>	<i>hā-y-kūn</i>	<i>kull-a</i>
	tomorrow	∅	day	new	FUT-3M- be\IPFV	every- 3SG.M
	<i>našāf</i>	<i>u</i>	<i>bidāy-a</i>	<i>ždīd-a,</i>	<i>ya</i>	<i>ṛabb!</i>
	energy	and	beginning-F	new-F	VOC	Lord

“Tomorrow is a new day: everything will be energy and a new beginning, oh Lord!”

Moreover, many dialects make use of copula forms in addition to the zero copula and *kān/ykūn* copula. For example, in some dialects, such as those in Egypt or Lebanon, present-tense equative constructions in which the complement is a definite noun phrase optionally use a copula based on the 3rd-person independent pronoun (Choueiri 2016) (example 5). Peripheral Arabic dialects go farther and employ the full range of the independent personal pronouns in these constructions (Akkuş 2018, pp. 459–62)

(5)		Lebanese Arabic (Choueiri 2016, p. 102)		
	<i>Sami</i>	<i>huwwe</i>	<i>mudīr</i>	<i>l-madras-e</i>
	Sami	COP\3SG.M	director	DEF-school-F

“Sami is the director of the school”

It is worth pointing out that essentially all literature on the copula in Arabic, theoretical or descriptive, has been devoted to either the “typical” copula situation or to the pronominal copula.⁵ Other types of copulas in Arabic dialects, especially ones which derive from verbs, have hardly been described. Only very recently have scholars begun to address the existence of other copulas, in particular the use of *gāʿīd*, formally the active participle of “to sit/to stay”, as a present-tense predicational copula in varieties such as Maltese and others (Camilleri and Sadler 2019, 2020) (example 6).

(6)		Maltese (Camilleri and Sadler 2019, p. 10)	
	<i>omm-i</i>	<i>qieghd-a</i>	<i>d-dar</i>
	mother-1SG	COP\PTCP.ACT-F	DEF-house.F

“My mother is at home”

Here, we describe for the first time the existence of an additional copula occurring in Arabic varieties of Tunisia and northwest Libya. This copula, supplied by the verb *yabda* (lexically “to begin”) occurs in certain types of predicational constructions. In Section 3, we will analyze copular *yabda* on the basis of representative examples from the well-documented varieties of Tunis (northern Tunisia), Douz (southern Tunisia), and Tripoli (northwest Libya). Since the goal of our study is dialectological in nature, deeper discussion of the grammaticalization path undergone by *yabda* to become a copula will be left aside, and we will concentrate on describing and comparing its function in these three dialects.⁶ The geographical range of copular *yabda* and its importance for dialectology will then be discussed in Section 4.

3. Copular *yabda* in Maghrebi Dialects

A copular element consisting of the verb *yabda* in the imperfective conjugation occurs in predicational constructions, mainly those where an overt copula is required. Copular *yabda* mainly occurs in narrative or descriptive contexts to refer to a habitual action, event, or description. It can also be used to refer to a present state and to describe an event which is happening at the moment of speech. Moreover, *yabda* can have a future value. Finally, it is used in addition to *ykūn* as the auxiliary of the future perfect. So far as we can tell, *yabda* never has a past reference, that is, in the perfective conjugation it is used only as a lexical verb and not as a copula. From a modal point of view, copular *yabda* seems to be used, as opposed to *ykūn*, when the speaker considers the states and the situations to be true or when the speaker believes that the content of the interrogative sentence can be validated by the interlocutor.

While the dialects under discussion all have parallels in the syntax of *yabda*, it should be noted that the phonological or morphological particularities of each dialect do apply

to copular *yabda*, without affecting its meaning. For example, the variety of Douz marks gender in the plural verb while that of Tripoli does not; hence Douz has both a 3PL.M *yabdu* and a 3PL.F *yabdan* while Tripoli has only 3PL *yabdu*. Or, the morphophonology of the 3PL may differ: *yabdāw* in Tunis but *yabdu* in Tripoli.

3.1. Habitual Present

In the majority of our examples, copular *yabda* indicates the usual occurrence of a state or situation. In all the following utterances, copular *yabda* has a habitual present value. It can also be used to express a general truth. It appears in main clauses as well as temporal clauses and can occur with adjectival, nominal, or locational predicates.⁷

3.1.1. Adjectival Predicate

The predicate can be adjectival (including passive participles). In the following examples, *yabda* refers to actions or events that take place habitually: indeed, in the first example, it describes an event that takes place every year because of the change of seasons; in the second one, every time a meat dish is cooked according to a particular method; finally, in the third example, every time the family gets together.

(7) Tripoli (Pereira 2010, p. 361)⁸
f-faʃl *ər-rbīʕ* *ta-bda* *d-dənya* *xadʕa*
 in-season DEF-spring 3F-COP\IPFV DEF-world.F green.F
 “In spring (i.e., every spring), nature is green”

(8) Tripoli (Pereira corpus)
u *ya-bda* *l-ḥam* *hādākāy* *mawʒūd*
 and 3M-COP\IPFV DEF-meat DEM.DIST.M exist\PTCP.PASS.SG.M
f-əs-silvər *u* *y-tlawwalḥ* *f-ūʃt* *əl-ḥufr-a*
 in-DEF- and 3M-be_thrown in-middle DEF-hole-F
 aluminum
 “And (each time you cook it) that meat is present in the aluminum foil and is thrown into the middle of the hole”

(9) Tripoli (Pereira corpus)
u *tabʕan* *ta-bda* *l-ʕəl-a* *malmūm-a*
 and of_course 3F-COP\IPFV DEF-family-F gather\PTCP.PASS-F
u *na-bd-u* *n-hadʕr-z-u* *fī* *mʷwādʕ* *ʕāmm-a*
 and 1-begin\IPFV-PL 1-talk\IPFV-PL in topic\PL general-F
u *mʷwādʕ* *t-xuʃʃ* *əl-ʕəl-a* *u* *kāda*
 and topic\PL 3F-concern\IPFV DEF-family-F and so
 “And of course (each time we go to my grandfather’s house) the family is reunited and we begin to talk about general topics and topics concerning the family and so forth”

It is also the case in Tunis Arabic where *yabda* allows the expression of habitual facts. Without *yabda*, examples (10–11) would have an actual present meaning.⁹

(10) Tunis (Singer 1984, p. 317)
el-farʃ *yə-bda* *ʕāli* *ʕa-l-aṛḍ*
 DEF-bed 3M-COP\IPFV high above-DEF-ground.F
 “The bed is high above the ground”

(11) Tunis (Singer 1984, p. 317)
tə-bda *d-dənya* *bard-a*
 3F-COP\IPFV DEF-world.F cold-F
 “Nature (i.e., the weather) is cold”

Copular *yabda* also appears in temporal clauses with a habitual present value. The conjunctions (*lamma*, *kif* and *kī*) refer not to a single, but rather to the habitual, occurrence of an event. The three representative dialects agree in this usage.

- (12) Tripoli (Pereira corpus)
aḷḷāh *lamma* *ya-bda* *xrayyəf* *kif*
 god when 3M-COP\IPFV lamb\DIM just
madbūh *aḷḷāh*
 slaughter\PTCP.PASS.SG.M god
 “Oh god (i.e., what a delight) when a lamb is just slaughtered!”
- (13) Tripoli (Pereira corpus)
lamma *l-bāb* *ya-bda* *hākki* *abyad* *w* *ʿaḷlē-ha*
 when DEF-door 3M- COP\IPFV so white on-3SG.F and
l-akwān *hādu* *maʿnā-ha* *ṣāḥəb* *əl-hōš* *māši* *l-əl-hažž*
 DEF- DEM.PROX. meaning- owner DEF-house go\PTCP. to-DEF-
 color\PL PL, 3SG.F ACT.SG.M pilgrimage
 “When the door is like this white and (with) these colors on it, it means that the owner has gone to the pilgrimage”
- (14) Douz (Ritt-Benmimoun 2014, pp. 67, 195)
tawwin *ya-bd-an* *f-əblāš-a* *mā-hī-š* *milēh-a*
 when 3-COP\IPFV-PL.F in-place-F NEG-3SG.F- NEG good-F
 “When they are in a place which isn’t good”
- (15) Douz (Ritt-Benmimoun 2014, p. 203)
iš-šgīr *kī* *ya-bda* *māzāl* *šgīr* *māzāl*
 DEF-small when 3M- COP\IPFV still small still
kī *tkawwan* *y-tīh*
 just develop\pfv[3SG.M] 3M- fall\IPFV
 “When the child (lit. small one) is still small, still just developed, it is miscarried (lit. falls)”
- (16) Douz (Ritt-Benmimoun 2014, p. 203)
ʕanəz *kī* *ta-bda* *simh-a* *y-dīr-u* *fā-ha*
 goat.F when 3F- COP\IPFV pretty-F 3-do\IPFV-PL in-
kammūn *aswad* 3SG.F
 cumin black
 “When a goat is pretty, they put black cumin on it”
- (17) Tunis (Singer 1984, p. 317)
el-kār *kīf* *tə-bda* *məlyān-a* *ma-ʕād-š*
 DEF-bus.F when 3F- COP\IPFV full-F no_longer
t-āqef *bəlkull*
 3F-stop\IPFV at_all
 “When the bus is full, it no longer stops at all”

3.1.2. Nominal Predicate

The predicate can also be a nominal phrase and copular *yabda* allows to provide a comment on an event or a fact as they habitually occur.

(18)		Douz (Ritt-Benmimoun 2014, p. 189)				
<i>amma</i>	<i>kīf</i>	<i>ya-bda</i>	<i>ʕām</i>	<i>xēr,</i>	<i>il-filā</i>	<i>tu-kṛuf-ha</i>
		3M-COP\ IPFV	year	good	DEF-plain	3F- sniff\ IPFV- 3SG.F
<i>b-xašəm-ha</i>	<i>in-nāg-a</i>	<i>walla</i>	<i>ž-žimal</i>			
by-nose- 3SG.F	DEF- camel-F	or	DEF-camel			

“But when it is a good year, the male or female camel sniffs the plain with its nose”

(19)		Tripoli (Pereira 2010, p. 302)				
<i>hūwa</i>	<i>ya-bda</i>	<i>wāzīr</i>	<i>w</i>	<i>əl- buwwāb</i>	<i>ʔmtāʕ-ah</i>	
3SG.M	3F-COP\ IPFV	minister	and	DEF- doorman	of-3SG.M	
<i>y-gaʕmz-u</i>	<i>f-waḡʔt</i>	<i>lə-ḡde</i>	<i>yə-tḡədd-u</i>		<i>mʔʕābaʕd-hum</i>	
3-sit\ IPFV- PL	in-time	DEF-lunch	3-eat_lunch\ IPFV-PL		together-3PL	

“He is a minister, and his doorman, they sit at lunchtime and eat lunch together”

(20)		Tripoli (Marwa Benschshin, p.c.)				
<i>marʔāt</i>	<i>ta-rʕaʕ</i>	<i>bə-ḡāʔ</i>		<i>u</i>	<i>ta-bda</i>	
sometimes	3F- return\ IPFV	by-small\ PL		and	3F-COP\ IPFV	
<i>hādi</i>	<i>l-muškil-a</i>	<i>lə-kbīr-a</i>			<i>t-tāny-a</i>	
DEM.PROX.F	DEF- problem-F	DEF-big.F			DEF-second-F	

“Sometimes she (viz. a divorced woman) is back with children and this is the other big problem”

In the following utterance, contrary to the previous examples, copular *yabda* is used in Tunis Arabic in an equative construction.

(21)		Tunis (Sellami corpus)				
<i>fi</i>	<i>š-šmāl</i>	<i>əl-məsfūf</i>	<i>yə-bda</i>	<i>kisiksi</i>	<i>žwayyəd</i>	<i>abyaḡd</i>
in	DEF-north	DEF-masfūf	3M-COP\ IPFV	couscous	fine	white

“In the north, masfūf is fine white couscous”

This example shows that *yabda* is required because there is a semantic constraint, in this case the circumstantial *fi š-šmāl* “in the north”, which limits the applicability of the claim about what *masfūf* is to a particular region. Otherwise, the equative construction with no overt copula would be used: *əl-məsfūf ʔ kisiksi* “*masfūf* (is) couscous”.

3.1.3. Locational Predication

Locational predication can also be expressed with copular *yabda*. In this case, the copula complement consists of a prepositional phrase or a locational adverb. The locational predication can have a habitual value (22–23).

(22)		Tripoli (Pereira corpus)				
<i>l-ḥarār-a</i>	<i>mā-t-žī-šši</i>	<i>mən</i>	<i>žih-a</i>		<i>wāḥd-a,</i>	
DEF-heat-F	NEG-3F-come\ IPFV- NEG.3SG.M	from	direction-F		one-F	
<i>ta-bda</i>	<i>mən</i>	<i>žamīʕ</i>	<i>əl-žih-āt</i>			
3F-COP\ IPFV	from	totality	DEF-direction-PL.F			

“(Each time you cook it) the heat doesn’t come (to the meat) from one direction, it is from all directions”.

(23)		Douz (Ritt-Benmimoun 2014, p. 117)				
<i>il-liḡ-a</i>	<i>ta-bda</i>	<i>fi</i>	<i>ḥuḡr-a</i>	<i>lōḡa</i>	<i>bārd-a</i>	
DEF-viper-F	3F-COP\ IPFV	in	hole-F	below	cold-F	

“The viper is in a cold hole below”

In the following example from Tunis, copular *yabda* expresses locational predication in a temporal clause.

(24)	Tunis (Singer 1984, p. 317)			
<i>kullma</i>	<i>yə-bda</i>	<i>f-əd-dār</i>	<i>i-walli</i>	
every_time	3M-COP\IPFV	in-DEF-house.F	3M-return\IPFV	
<i>i-šārek</i>	<i>fī</i>	<i>mart-u</i>		
3M-argue\IPFV	in	wife-3SG.M		

“Every time he is at home, he keeps criticizing his wife”.

3.2. Future

Depending on the context, copular *yabda* situates an event or a state in the future, whether it is a question about location or state or a wish or hope about a situation. The three representative dialects agree in this usage.

(25)	Tripoli (Pereira corpus)			
<i>amta?</i>	<i>lamma</i>	<i>na-bd-u</i>	<i>ʕaʕle</i>	<i>ʕaʕkākiz?</i>
when	when	1-COP\IPFV-PL	on	cane\PL

“When? When we will be on (i.e., walking with) canes?”

(26)	Tunis (Singer 1984, p. 317)			
<i>fīn</i>	<i>tə-bda</i>	<i>ʕand</i>	<i>l-ūwəl?</i>	
where	2-COP\IPFV	at	DEF-one	

“Where will you be at noon?”

(27)	Tunis (Sellami corpus)			
<i>škūn</i>	<i>yə-bda</i>	<i>hūni</i>	<i>fī</i>	<i>ūt?</i>
who	3M-COP\IPFV	here	in	august

“Who will be here in August?”

(28)	Douz (Ritt-Benmimoun 2014, p. 401)					
<i>šālla</i>	<i>ta-bda</i>	<i>ħdā-na</i>	<i>w</i>	<i>taww</i>	<i>ti-tfarʕaž</i>	<i>taww</i>
God_willing	2-COP\IPFV	near-1PL	and	FUT	2-watch\IPFV	FUT
<i>ta-šbaħ</i>						
2-see\IPFV						

“God-willing you will be near us and you will watch, you will see”

3.3. Future Perfect

Finally, followed by a verb in the perfective, *yabda* is also used as the auxiliary verb of the future perfect, indicating a state or situation that is expected or planned to occur in the future. Here, however, *yabda* and *ykūn* can both be used with a variation in meaning that requires further study. This usage only exists in two of the representative dialects: Tripoli and Tunis.

(29)	Tripoli (Pereira corpus)					
<i>na-bd-u</i>	<i>mšē-na</i>	<i>l-ħoš</i>	<i>žədd-i</i>	<i>matalan</i>		
1-COP\IPFV-PL	go\PFV-1PL	to-house	grandfather-1SG	for_example		
<i>u</i>	<i>malmūm-īn</i>	<i>mʕa-baʕd</i>	<i>fā</i>	<i>n-ākl-u</i>	<i>mʕaʕābaʕd</i>	<i>bāzīm</i>
and	gather\PTCP.PASS-PL.M	together so	1-eat\IPFV-PL	together	bazin	

“We will have gone to my grandfather’s house, for example, so gathered together we eat bazin together”

(30) Tripoli (Pereira corpus)
 ṡādatan maṡa s-sāṡa tnīn n-kūn-u ṡawwəḥ-na mən
 normally with DEF-hour two 1-be\IPFV-1PL return\PFV-1PL from
 əṡ-ṡle maṡnā-ha maṡa t-tlāta u ṡubəṡ n-kūn-u
 DEF-prayer meaning-3SG.F with DEF-three and quarter 1-be\IPFV-1PL
kəmməl-na
 finish\PFV-1PL
 “Normally around 2 pm we will have returned home from prayer, so around 3:15 pm we will have finished (eating lunch)”.

(31) Tunis (Sellami corpus)
 lā 21 mā-n-nəṡṡəm-ṡ bāṡ nə-bda rawwəḥ-t
 no 21 NEG-1-be_able\IPFV-NEG FUT 1-COP\IPFV return\PFV-1SG
 “On the 21st I can’t I will have gone back home”.

(32) Tunis (Sellami corpus)
 ṡḥāb-ək l-qdom l-koll y-kūn-u ṡarrs-u w
 friend\PL-2SG DEF-old\PL DEF-all 3-be\IPFV-3PL marry\PFV-3PL and
 ənti lā
 2SG.F no
 “All of your old friends will have gotten married but not you”.

In Douz neither *yabda* nor *ykūn* can be used for the future perfect, but instead *ywalli* (lexically “to become”) is used.

(33) Douz (Ritt-Benmimoun, p.c.)
 n-walli rawwəḥ-t min il-qṡāya
 1-become\IPFV return\PFV-1SG from DEF-study
 “I will have returned home from school”.

In Tripoli, the use of *yabda* or *ykūn* for the future perfect seems to break down along the following lines: *yabda* is used when the speaker considers the future state or situation as certain to occur, while *ykūn* in contrast allows for the addition of modality, expressing a supposition or a fictional or probable hypothesis. This aligns, in fact, with the use of *ykūn* for expressing epistemic modality in the present (Pereira 2010, pp. 453–67).

(34) Tripoli (Pereira 2010, p. 460)
 (a) y-kūn bne ḥōṡ-a
 3M-be\IPFV build\PFV[3SG.M] house-3SG.M
 “He will have built his house”¹⁰
 (b) mumkən y-kūn bne ḥōṡ-a tawwa
 maybe 3M-be\IPFV build\PFV[3SG.M] house-3SG.M now
 “He may have built his house now”

The distinction between *yabda* or *ykūn* seems to be similar in Tunis as well, though this requires further study.

4. Copular *yabda* as Isogloss and the Problem of Syntactic Isoglosses

As shown in the preceding section, copular *yabda* exists in both “northern” and “southern” Tunisian varieties, as typified for this study by the areas of Tunis and Douz, respectively. More generally, according to Tunisian colleagues and colleagues working on other Tunisian varieties, it can be considered a pan-Tunisian feature.¹¹ In Libya, the

only location where copular *yabda* has been documented is Tripoli, though it would be unsurprising if other varieties of northwestern Libya, about which there is little published, also had the feature. The total geographic extent of copular *yabda* is not yet known; but it does not exist in Benghazi or eastern Libyan varieties generally, and there is essentially no documentation of eastern Algerian varieties available for comparison. It is unknown in areas of central coastal Algeria, such as Algiers or Dellys, however.¹² According to the existing information, therefore, it is a shared feature of the varieties of Tunisia and Tripoli (see Table 1).¹³

Table 1. Domains of copular *yabda*.

Dialect	Habitual Present	Future	Fut. Perf.
Tunis	Yes	Yes	Yes (<i>yabda/ykūn</i>)
Douz	Yes	Yes	No
Tripoli	Yes	Yes	Yes (<i>yabda/ykūn</i>)

That these dialects share a linguistic feature, in particular an innovation, is unexpected given the categories and isoglosses typically used in Arabic dialectology. Copular *yabda* crosses not only national boundaries (Tunisia/Libya) but also the pseudo-typological ones most prominent in Arabic dialectology, in particular the categories of “pre-Hilali/Hilali” or “sedentary/Bedouin”. Besides the fact that these categories are outdated and problematic from a socio-historical point of view, it must also be pointed out that the collection of features on which they are based almost never includes syntactic features. In Arabic dialectology, syntax plays very little role in discussion of dialect classification. For example in a recent handbook, the authors note that “syntax will, and we do not constitute an exception in so doing, only be taken into account in a restricted manner, although in this area too significant differences between dialects are present” (Behnstedt and Woidich 2005, p. 68). More generally, recent large projects of regional dialectology, such as the *Wortatlas der arabischen Dialekte* (Behnstedt and Woidich 2011–2021), include phonology, morphology, and lexicon, but not syntax. Even the most recent overviews of Maghrebi dialects (e.g., Aguadé 2018) do not treat syntax. Syntax has received slightly more attention from sociolinguists and contact linguists, but is typically not used as the basis for regionally-organized dialect groupings nor has it been studied as part of intra-dialect variation in ways comparable to phonology or morphology.¹⁴

Syntax seems to be neglected in dialectology in general regardless of language. Even recently, scholars have gone as far as stating that “there is no doubt that syntax has been the most neglected linguistic subsystem in classical dialectology” (Berger et al. 2012, p. 93). On one hand, this goes back to the fact that traditional dialectological methods, such as the word list and questionnaires, can be unsuitable for describing syntax; on the other, syntax does not necessarily fit the diachronic documentation goal of traditional dialectology, which concentrated on phonological and lexical criteria (Glaser 1996; Werlen 1994). However, this state of affairs has changed quite significantly in certain fields, such as Germanic and Romance dialectology (Kortmann 2010; Berger et al. 2012; Glaser 2013).

Arabic dialectology has largely shared the traditional dialectological emphasis on uncovering archaisms, partially due to its goal of answering questions about the historical origin of Arabic dialects. As with other languages, Arabic dialect groupings have been made primarily on the basis of phonological, morphological, and lexical isoglosses.¹⁵ For example, of the 73 isoglosses used by De Jong (2000, pp. 39–48) to group the Arabic dialects of the Sinai peninsula, only 4 can potentially be characterized as (morpho-)syntactic. Meanwhile, some of these traditional non-syntactic isoglosses may not withstand scrutiny: Embarki (2008) argues, for example, that some of the isoglosses traditionally considered to be strong markers of dialect type, such as the interdental consonants, exhibit too much variation within a single dialect to really be useful discriminants (and see again Guerrero, forthcoming).

This being the case, attention to syntax as part of dialectology has the potential to complexify and even complicate the typical dialect groupings. Indeed, it has been noted that syntactic isoglosses often cross and contradict the established isoglosses based on phonology or lexicon (Poletto 2013). As Glaser (2013, p. 204) puts it, “that geographically conditioned syntactic variation indisputably exists does not entail, however, that the distribution of syntactic variants is identical to the distribution of phonological or lexical variants”. For the Arabic varieties under discussion here, this crossing and contradiction can easily be illustrated with a quick look at only a few isoglosses (Table 2).

Table 2. Selected Isoglosses in Tunis, Douz, Tripoli Arabic varieties.

Feature	Tunis	Douz	Tripoli
1. q	q	g	g
2. Interdentals	yes	yes	no
3. Final -ā	-ā	-ē	-ē
4. Gender in PL	no	yes	no
5. 3SG.M suffix	-u	-a	-a
6. “do”	yaʿmal	ydir	ydir
7. “want”	yħabb	ydawwir	yəbbi
8. “a lot”	barša	yāsər	hālba
9. COP <i>yabda</i>	yes	yes	yes

The above table considers three phonological, two morphological, three lexical, and one syntactic variable. Each of these categories yields different isogloss lines: in some cases Douz and Tripoli agree (nos. 1, 3, 5, 6), in other cases Tunis and Douz agree (no. 2). An isogloss grouping Tunis and Tripoli can even be found, namely the lack of gender marking on plural verbs (no. 4). Of course, many of these features are shared with dialects beyond these three and so only serve to connect two of the three with each other, but not to separate them out from surrounding dialects. Copular *yabda* not only is an isogloss connecting Tunis, Douz, and Tripoli, but also one which separates them out from other Maghrebi dialects.

This raises the question of how much weight a syntactic isogloss should have as part of a group of multiple different isoglosses. While phonological and lexical isoglosses are typically more valued by dialectologists, and more frequently available in the published literature, Chambers and Trudgill (2004, pp. 96–100) note that there is evidence that “grammatical variables stratify speech communities much more sharply than do phonological and lexical variables”, suggesting that regions delimited by grammatical isoglosses will be more strongly regarded as different dialect areas than regions separated by mostly phonological and lexical ones.¹⁶ Moreover, there seems to be agreement that grammatical isoglosses delimit larger areas than phonological or lexical ones. In this regard, one would think that copular *yabda* and other syntactic isoglosses should actually have a fair amount of weight when it comes to drawing up-to-date subgroups of Maghrebi Arabic.

Proponents of the traditional dialectological view might note that *yabda* is relatively new in the history of the Arabic varieties in question and that, as an innovation, only represents the spread of a particular feature in very recent history and therefore does not affect the traditional classification. But we would counter that copular *yabda* is not necessarily all that new, as it is already attested in Tripoli at the end of the 19th century.¹⁷ Moreover, an innovative feature that is well-attested in a fairly significant region should be the concern of dialectologists and future research should attempt to account for its history and present distribution. For example, did copular *yabda* jump between urban areas, slowly spreading into the rural areas between them? Or did it radiate out from a particular area where it was first innovated? Why has it, seemingly, not been accompanied by the spread of non-syntactic features?

If we are dealing with the spread of a syntactic innovation in the Arabic varieties of a particular region, then we indeed have to think less about the traditional classifications, which attempt to explain how the distribution of Arabic may have looked centuries ago, and more about processes of inter-dialectal contact and diffusion. And it is here that copular *yabda* may also make a contribution, since studies of inter-dialectal contact in Arabic have typically focused on what happens when different dialects come into contact in urban environments, rather than looking at the diffusion of a feature over a large region. These studies also typically focus on phonological and morphological variables, rather than syntactic ones. Meanwhile, general studies of convergence do typically focus on morpho-syntax, though in most cases they deal with totally different *languages* rather than different *varieties* of a language. Copular *yabda* may represent a case of a syntactic innovation being spread through dialect contact over a large region, giving rise to a dialectal version of a “linguistic area”, that is, the “outcome of diffusion of structural ‘patterns’ across language boundaries” (Matras 2011, p. 146). In that case, it may be one example of area formation in Arabic dialects, and indeed one that does not follow national boundaries but instead crosses them. And again here, syntax is important, since, as is clear from Table 2 above, the diffusion of copular *yabda* seems, so far as can be seen, not to have been accompanied by the diffusion of other linguistic features. It thus speaks to interaction *between* Maghrebi dialects that can not be seen simply by looking at areas like phonology or lexicon. Future research should therefore look to morphosyntax in search of other features which (unexpectedly) link Tunisia and northwest Libya, or characterize other dialect areas in general.

5. Conclusions

In this study, we have attempted to describe the occurrence of a verbal copula in certain types of predicational, and less frequently equative, constructions in dialects of Tunisia and northwest Libya. This copula is provided by the verb *yabda*, lexically meaning “to begin”, and occurs in predicational constructions which require an overt copula, both in the present and future, including constructions with temporal or modal implication. This can be illustrated succinctly with a final example, taken from social media, where the generic predicational construction with zero copula (35a) contrasts with the temporal construction requiring an overt copula (35b) which is supplied with a form of *yabda*.

(35)		Tunis Arabic		
(a)	<i>umm-ək</i>	∅	<i>tūnsīy-a</i>	
	mother-2SG	∅	Tunisian-F	
	“Your mother is Tunisian”			
(b)	<i>kī</i>	<i>ta-bda</i>	<i>umm-ək</i>	<i>tūnsīy-a</i>
	when	3F-COP\IPFV	mother-2SG	Tunisian-F
	“When your mother is Tunisian...”			

Importantly, however, the *yabda* copula is attested in a number of dialects, including three dialects—Tunis, Douz, and Tripoli—which are not closely linked in the traditional dialectological classifications. As a syntactic isogloss, *yabda* crosses the isoglosses drawn from other linguistic levels, ignoring national and typological boundaries, exhibiting behavior seen in syntactic isoglosses more generally. While our study has only been able to use currently existing material to suggest what the rough area contained by the *yabda* isogloss may be, additional data from locales in between these three representative locations may be able to help us define that area more precisely, and, in addition, potentially show if there are transitional areas as well. More importantly, copular *yabda* requires explanations that do not draw on the traditional historical classifications for Arabic dialects, but look to diffusion, area formation, and above all contact. We suggest that syntactic features should play a larger role in Arabic dialectology, and including more of them in the lists of isoglosses drawn on for classification has the potential to complexify and even reshape our understanding of the distribution of Arabic dialects and the processes which continue to shape them.

Author Contributions: Both authors wrote and revised the article together and have joint responsibility for all sections. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors would like to thank Zeineb Sellami for sharing and discussing unpublished data with us, as well as Marwa Benschshin and the anonymous reviewers for their valuable suggestions and corrections.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

- ¹ For example, [Taine-Cheikh \(2017\)](#) argues that linguistic criteria do not necessarily support the existence or utility of the subdivision of “Bedouin” into “Hilali”, “Sulaymi”, and “Ma’qil” subgroups. [Guerrero \(forthcoming\)](#) argues that one of the traditional isoglosses separating sedentary/bedouin or pre-Hilali/Hilali dialects, the presence or absence of the interdental consonants (*t̪*, *d̪*), is not well-founded. In addition, the applicability and characteristics of the category of “village” or “rural” dialects has continued to provoke discussion ([Mion 2015](#); [Guerrero 2018](#)).
- ² Linguistic examples in this paper are drawn from the published literature as indicated. We have supplemented these with unpublished examples kindly provided by Zeineb Sellami from their personal corpus of Tunisian Arabic (these are marked ‘Sellami corpus’) and with examples drawn from the personal corpus of Christophe Pereira (marked ‘Pereira corpus’). In addition, the TUNICO corpus, representing the Tunis area, includes a number of examples of copular *yabda* which can be viewed at the following address: https://tunico.acdh.oeaw.ac.at/corpus.html?query=*ibd*&startRecord=51 (accessed on 21 October 2021). For published examples we use the published transcription with modifications to diacritics and segmentation for glossing as needed.
- ³ As opposed, for instance, to utterances where the pragmatic marker *rā* is used to focus the predicate or the entire predicative relation ([Caron et al. 2015](#), pp. 105–7), such as *šaxšiy-t-a rā-hi dšif-a* “his personality, it really is weak” and *šadnān rā-hu f-əš-šahra* “Adnan, he is indeed in the desert”.
- ⁴ The syntax of glosses is largely inspired by the *Moroccan and Libyan Arabic list of glosses* developed by Dominique Caubet, Ángeles Vicente, Alexandrine Barontini, and Christophe Pereira for the CorpAfroAs project. This list of glosses can be viewed at the following address: http://corpafroas.tge-adonis.fr/fichiers/Moroccan_Libyan_Arabic_Glosses_final.pdf (accessed on 21 October 2021).
- ⁵ Such as ([Aoun et al. 2010](#), pp. 35–44; [Alharbi 2017](#); [Alotaibi 2018](#), and many others).
- ⁶ The grammaticalization of “begin” into a copula seems unusual, and is not mentioned in the typological literature on grammaticalization or non-verbal predication (e.g., [Heine and Kuteva 2002](#); [Hengenveld 1992](#); [Pustet 2003](#)).
- ⁷ The copular forms of *yabda* will be glossed as COP rather than etymologically as “begin” in order to avoid confusion with the lexical usage of *yabda*.
- ⁸ Example 7 can be considered as a general truth and this precisely is why the use of *yabda* is required. To express a change taking place, the verb *ywəlli* “become” would have to be employed.
- ⁹ [Singer \(1984, p. 317\)](#) provides an additional number of examples which confirm the use of *yabda* with adjective predicates, but which are hard to discuss because they do not include any context besides the copula and predicate. His translations (e.g., *təbda rīda* “sie ist breit” or *yibdaḥw fārḥin* “sie sind froh”) suggest that *yibda* is to be understood as a normal present copula, but this is not the case as far as we can tell. In contrast, it should be noted that the examples from ([Ritt-Benmimoun 2014](#)) are based on her published text corpus ([Ritt-Benmimoun 2011](#)), and the context can always be checked.
- ¹⁰ Note that this sentence also means “he has built his house” in Tripoli Arabic. The use of *yabda* instead of *ykūn* avoids confusion since *yabda* + suffixed conjugation only expresses the future perfect.
- ¹¹ Zeineb Sellami, p.c. It is also attested in the areas of Chebba (Luca D’Anna, p.c.), Mahadhba ([Marçais and Viré 1981](#), p. 375), and Susa ([Talmoudi 1984](#), p. 63), though because the published attestations are very few, it is not known if these areas align completely with the Tunis or Douz usages or not.
- ¹² Lameen Souag, p.c.
- ¹³ Interestingly, copular *yabda* is attested in the materials from the Fezzan collected mainly by Philippe Marçais in the 1950s and published posthumously ([Marçais 2001](#)). These materials are problematic, however, as for the most part it is not clear where or from whom any given text was collected (materials were gathered not only in the Fezzan, but also in Tripoli and in Algiers, and only one text has the name or any personal details of an informant). There is thus no way to be sure that *yabda* in these materials represents an actual usage of a Fezzani dialect rather than the usage, say, of someone who was originally from the Fezzan but had been living in Tripoli for some time. What also casts some doubt on a copular *yabda* in the Fezzan is that of the nine attestations, seven are attributed to the “Gwayda tribe”, suggesting that they may all come from a single informant (the other two have no

attribution). We have therefore left these attestations out of our analysis. This is not to suggest, however, that copular *yabda* has not now spread to some regions of southern Libya; but there are unfortunately no studies which can confirm this as of yet.

¹⁴ See the recent handbook chapters (Camilleri 2019; Choueiri 2019).

¹⁵ See (Magidow 2016) for an example of dialect classification based on morphological forms. For some broad comparative studies of syntax, see Ritt-Benmimoun (2017, pp. 324–32) on object and aspect marking with *ft*, or Bettega (2019) on gender agreement.

¹⁶ See also Behnstedt and Woidich (2005, pp. 83–92) for Arabic specifically. Note also that whether or not a given isogloss affects mutual intelligibility is a separate question, and the answer may not actually correspond to the isoglosses valued by dialectologists, see (Čeplö et al. 2016).

¹⁷ (Stumme 1898, p. 65); also attested in the mid-20th century in Douz (Boris 1951, pp. 212–5).

References

- Aguadé, Jordi. 2018. The Maghrebi dialects of Arabic. In *Arabic Historical Dialectology: Linguistic and Sociolinguistic Approaches*. Edited by Clive Holes. Oxford: Oxford University Press, pp. 29–63.
- Akkus, Faruk. 2018. Peripheral Arabic Dialects. In *The Routledge Handbook of Arabic Linguistics*. Edited by Elabbas Benmamoun and Reem Bassiouney. London and New York: Routledge, pp. 454–71.
- Alharbi, Bader Y. 2017. The Syntax of Copular Clauses in Arabic. Unpublished Ph.D. thesis, University of Wisconsin-Milwaukee, Milwaukee, WI, USA.
- Alotaibi, Ahmad S. 2018. The Copula in Arabic: Description and Analysis. Unpublished Ph.D. thesis, University of Essex, Colchester, UK.
- Aoun, Joseph E., Elabbas Benmamoun, and Lina Choueiri. 2010. *The Syntax of Arabic*. Cambridge: Cambridge University Press.
- Behnstedt, Peter, and Manfred Woidich. 2005. *Arabische Dialektgeographie: Eine Einführung*. Leiden: Brill.
- Behnstedt, Peter, and Manfred Woidich. 2011–2021. *Wortatlas der Arabischen Dialekte*. 4 vols. Leiden: Brill.
- Benkato, Adam. 2019. From medieval tribes to modern dialects: On the afterlives of colonial knowledge in Arabic dialectology. *Philological Encounters* 4: 2–25. [CrossRef]
- Benmofthah, Najah, and Christophe Pereira. 2019. Les futurs en arabe de Tripoli (Libye): Temporalité, aspectualité et modalités. In *Studies on Arabic Dialectology and Sociolinguistics: Proceedings of the 12th International Conference of AIDA*. Edited by Catherine Miller, Alexandrine Barontini, Marie-Aimée Germanos, Jairo Guerrero and Christophe Pereira. Aix-en-Provence: IREMAM.
- Berger, Claudia Bucheli, Elvira Glaser, and Guido Seiler. 2012. Is a syntactic dialectology possible? Contributions from Swiss German. In *Methods in Contemporary Linguistics*. Edited by Ender Andrea, Leemann Adrian and Wälchli Bernhard. Berlin: De Gruyter, pp. 93–120.
- Bettega, Simone. 2019. Rethinking Agreement in Spoken Arabic: The Question of Gender. *Annali Sezione Orientale* 79: 126–56. [CrossRef]
- Boris, Gilbert. 1951. *Documents Linguistiques et Ethnographiques sur une Région du sud Tunisien (Nefzaoua)*. Paris: Maisonneuve.
- Camilleri, Maris, and Louisa Sadler. 2019. The grammaticalisation of a copula in vernacular Arabic. *Glossa: A Journal of General Linguistics* 4: 1–33. [CrossRef]
- Camilleri, Maris, and Louisa Sadler. 2020. The grammaticalisation of an auxiliary and a copula: The Arabic “sit” participle. *Journal of Historical Syntax* 4: 1–60. [CrossRef]
- Camilleri, Maris. 2019. Morphosyntactic variation: Focus on Maltese and other Western varieties. In *The Routledge Handbook of Arabic Sociolinguistics*. Edited by Enam Al-Wer and Uri Horesh. London: Routledge, pp. 214–26.
- Caron, Bernard, Cécile Lux, Stefano Manfredi, and Christophe Pereira. 2015. The intonation of Topic and Focus: Zaar (Nigeria), Tamasheq (Niger), Juba Arabic (Sudan) and Tripoli Arabic (Libya). In *Corpus-based Studies of Lesser-described Languages. The CorpAfroAs Corpus of Spoken AfroAsiatic Languages*. Edited by Amina Mettouchi, Martine Vanhove and Dominique Caubet. Amsterdam: John Benjamins, pp. 63–115.
- Caubet, Dominique. 2001. Questionnaire de dialectologie du Maghreb (d’après les travaux de W. Marçais, M. Cohen, G. S. Colin, J. Cantineau, D. Cohen, Ph. Marçais, S. Lévy, etc.). *Estudios de Dialectología Norteafricana y Andalusí* 5: 73–92.
- Čeplö, Slavomír, Ján Batora, Adam Benkato, Jiří Milička, Christophe Pereira, and Petr Zemánek. 2016. Mutual Intelligibility of Spoken Maltese, Libyan Arabic and Tunisian Arabic Functionally Tested: A Pilot Study. *Folia Linguistica* 50: 583–628. [CrossRef]
- Chambers, J. K., and Peter Trudgill. 2004. *Dialectology*. Cambridge: Cambridge University Press.
- Choueiri, Lina. 2016. The Pronominal Copula in Arabic. *Brill’s Journal of Afroasiatic Languages and Linguistics* 8: 101–35. [CrossRef]
- Choueiri, Lina. 2019. Syntactic Variation. In *The Routledge Handbook of Arabic Sociolinguistics*. Edited by Enam Al-Wer and Uri Horesh. London: Routledge, pp. 185–200.
- De Jong, Rudolf. 2000. *A Grammar of the Bedouin Dialects of the Northern Sinai Littoral: Bridging the Linguistic Gap between the Eastern and Western Arab World*. Leiden: Brill.
- Embarki, Mohamed. 2008. Les dialectes arabes modernes: État et nouvelles perspectives pour la classification géo-sociologique. *Arabica* 55: 583–604. [CrossRef]
- Glaser, Elvira. 1996. Dialektsyntax: Eine Forschungsaufgabe. *Schweizerdeutsches Wörterbuch. Bericht über das Jahr 1997*: 11–31.

- Glaser, Elvira. 2013. Area formation in morphosyntax. In *Space in Language and Linguistics: Geographical, Interactional, and Cognitive Perspectives*. Edited by Peter v. Auer, Martin Hilpert, Anja Stukenbrock and Benedikt Szmrecsany. Berlin: De Gruyter, pp. 195–221. [[CrossRef](#)]
- Guerrero, Jairo. 2018. Les parlers jbalá-villageois: Étude grammaticale d'une Typologie rurale de l'arabe dialectal Maghrébin. *Dialectologia* 20: 85–105.
- Guerrero, Jairo. On Interdental Fricatives in the First-Layer Dialects of Maghrebi Arabic. forthcoming. *Brill's Journal of Afroasiatic Languages and Linguistics*.
- Heine, Bernd, and Tania Kuteva. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.
- Hengenveld, Kees. 1992. *Non-Verbal Predication: Theory, Typology, Diachrony*. Berlin: Mouton de Gruyter.
- Higgins, Roger F. 1979. *The Pseudo-Cleft Construction in English*. New York: Garland.
- Kortmann, Bernd. 2010. Areal variation in syntax. In *Language and Space. Vol 1: Theories and Methods*. Edited by Peter Auer and Jürgen Erich Schmidt. Berlin: De Gruyter, pp. 837–64.
- Kosansky, Oren. 2016. When Jews Speak Arabic: Dialectology and Difference in Colonial Morocco. *Comparative Studies in Society and History* 58: 5–39. [[CrossRef](#)]
- Magidow, Alexander. 2016. Diachronic Dialect Classification with Demonstratives. *Al-Arabiyya* 49: 91–115.
- Marçais, Philippe, and F. Viré. 1981. Gazelles et outardes en Tunisie: Reportage en parler arabe de la tribu des Mahàdhba. *Arabica* 28: 369–87. [[CrossRef](#)]
- Marçais, Philippe. 2001. *Parlers Arabes du Fezzân: Textes, Traductions et Éléments de Morphologie Rassemblés et Présentés par Dominique Caubet, Aubert Martin et Laurence Denooz*. Geneva: Librairie Droz.
- Matras, Yaron. 2011. Explaining Convergence and the Formation of Linguistic Areas. In *Geographical Typology and Linguistic Area*. Edited by Osamu Hieda, Christa König and Hiroshi Nakagawa. Tokyo: Tokyo University of Foreign Studies, pp. 143–60.
- Mikkelsen, Line. 2011. Copular clauses. In *Semantics: An International Handbook of Natural Language Meaning*. Edited by Claudia Maienborn, Klaus von Heusinger and Paul Portner. Berlin: De Gruyter, vol. 2, pp. 1805–29.
- Mion, Giuliano. 2015. Réflexions sur la catégorie des «parlers villageois» en arabe tunisien. *Romano-Arabica* 15: 269–79.
- Palva, Heiki. 2006. Dialects: Classification. In *Encyclopedia of Arabic Language and Linguistics 1*. Leiden: Brill, pp. 604–13.
- Pereira, Christophe. 2010. *Le Parler Arabe de Tripoli (Libye)*. Zaragoza: Instituto de Estudios Islámicos y del Oriente Próximo (IEIOP).
- Pereira, Christophe. 2011. Arabic in the North African region. In *The Semitic Languages: An International Handbook*. Edited by Stefan Weninger, Janet C. E. Watson, Geoffrey Khan and Michael P. Streck. Berlin: De Gruyter Mouton, pp. 954–69.
- Pereira, Christophe. 2018. Waves of Arabization and the Vernaculars of North Africa: An annotated bibliography. In *The Routledge Handbook of Arabic Linguistics*. New York: Routledge, pp. 488–503.
- Poletto, Cecilia. 2013. Leopard spot variation: What dialects have to say about variation, change and acquisition. *Studia Linguistica* 67: 165–83. [[CrossRef](#)]
- Pustet, Regina. 2003. *Copulas: Universals in the Categorization of the Lexicon*. Oxford: Oxford University Press.
- Ritt-Benmimoun, Veronika. 2011. *Texte im Arabischen Beduinendialekt der Region Douz (Südtunesien)*. Wiesbaden: Harrassowitz.
- Ritt-Benmimoun, Veronika. 2014. *Grammatik des Arabischen Beduinendialekts der Region Douz (Südtunesien)*. Wiesbaden: Harrassowitz.
- Ritt-Benmimoun, Veronika. 2017. The preposition *fi-* as an object and aspect marker in a south Tunisian bedouin dialect. *Journal of Semitic Studies* 62: 319–55. [[CrossRef](#)]
- Roy, Isabelle. 2013. *Nonverbal Predication: Copular Sentences at the Syntax-Semantics Interface*. Oxford: Oxford University Press.
- Singer, Hans-Rudolf. 1984. *Grammatik der Arabischen Mundart der Medina von Tunis*. Berlin: Walter de Gruyter.
- Stumme, Hans. 1898. *Märchen und Gedichte aus der Stadt Tripolis in Nordafrika*. Leipzig: J.C. Hinrichs'sche.
- Taine-Cheikh, Catherine. 2017. La classification des parlers bédouins du Maghreb: Revisiter le classement traditionnel. In *Tunisian and Libyan Arabic Dialects: Common Trends—Recent Developments—Diachronic Aspects*. Edited by Veronika Ritt-Benmimoun. Zaragoza: PUZ, pp. 15–42.
- Talmoudi, Fathi. 1984. Notes on the Syntax of the Arabic Dialect of Sūsa. *Zeitschrift für Arabische Linguistik* 12: 48–85.
- Werlen, Iwar. 1994. Neuere Fragestellungen in der Erforschung der Syntax deutscher Dialekte. In *Dialektologie des Deutschen. Forschungsstand und Entwicklungstendenzen (Germanistische Linguistik 147)*. Edited by Klaus Mattheier and Peter Wiesinger. Tübingen: Niemeyer, pp. 49–75.

Article

Connecting the Lines between Old (Epigraphic) Arabic and the Modern Vernaculars

Ahmad Al-Jallad

Faculty of Theology, University of Groningen, 9712 CP Groningen, The Netherlands; a.m.al-jallad@rug.nl

Abstract: This paper investigates three linguistic features—wawation, the 1CS genitive clitic pronoun, and the relative pronoun—that are shared between the ancient epigraphic forms of Arabic and modern dialects, to the exclusion of Classical Arabic. I suggest that these features represent the earliest linguistic layer of the modern dialects.

Keywords: historical linguistics; Arabic dialectology; Arabic epigraphy

1. Introduction

It has been widely recognized that the diverse forms of spoken Arabic today do not descend in a linear manner from the literary Arabic of medieval prose and poetry—conventionally termed Classical Arabic—or the language of the Quranic Consonantal Text (QCT), Old Ḥigāzī (for the most recent appraisal, see [Holes 2018a](#), pp. 1–28; [Al-Jallad 2020b](#), chps. 4 and 5). Indeed, when viewed through the lens of the comparative method, many modern Arabic vernaculars exhibit features that are more archaic than their Classical Arabic counterparts. Na’ama [Pat-El \(2017\)](#) has skillfully identified a number of such features in her 2017 article “Neo-Arabic and Comparative Semitics”. Clive Holes has also done pioneering work on pre-Islamic relics in the modern vernaculars of the Gulf, especially in the realm of the lexicon ([Holes 2018b](#), pp. 112–32). [Van Putten and Benkato \(2017\)](#) isolated relics of an earlier stratum of Arabic in loans in Awjila Berber that is distinct from the present-day dialects of Libya. And I have suggested that the phonology of the emphatics of pre-Hilalian Maghrebian Arabic may be connected to the pre-Islamic dialects of the Levant ([Al-Jallad 2015](#)). The existence of these features implies that an unidentified stratum of Arabic that failed to achieve written form in the early Islamic period contributed to the formation of modern vernaculars.

This essay explores the possibility that such ancestors may be attested in the pre-Islamic epigraphic record. Before approaching this question, however, it is important to recognize two things. The modern vernaculars never existed in a vacuum; they have experienced considerable contact with the literary register, which has contributed significantly to their lexicons and to their grammatical structure. In addition to this, interdialectal contact has led to an amalgamation of grammatical features in living speech, ones that originate in different times and places. An obvious example of this is the verb *šāf* “he saw”, which is nearly pan-Arabic today. *šāf*, although presently widespread in the Maghreb, was likely a late introduction through inter-dialectal contact ([Aguadé 2018](#), p. 57). It is absent in Maltese, which became isolated from the Arabic sprachraum by the 13th century, and is not used in several pre-Hilalian dialects. These only know *ra*. The same applies to the Levant. There, *šāf* is the primary verb used to express “to see” in Lebanon, yet Cypriot Arabic, which originates on the Levantine coast and became isolated from the Arabic-speaking world by the 13th c. CE, does not use this etymon. Instead, it employs two verbs for “to see”—*ra* (Proto-Arabic *ra’aya; Classical Arabic ra’ā; [Borg 2004](#), p. 214) and *kiš’e* (Qeltu qaša^ʕ; [Borg 2004](#), p. 388). The latter is fossilized as a presentative in Damascene Arabic, *šā^ʕ* ([Souag 2016](#)). While it is clear that Cypriot Arabic shares a common ancestor with the

Citation: Al-Jallad, Ahmad. 2021. Connecting the Lines between Old (Epigraphic) Arabic and the Modern Vernaculars. *Languages* 6: 173. <https://doi.org/10.3390/languages6040173>

Academic Editors: Simone Bettiga and Roberta Morano

Received: 15 September 2021

Accepted: 15 October 2021

Published: 20 October 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

dialects of the Levant, in the intervening centuries since its isolation, a new verb for “to see” spread as a result of contact with other dialects, in this case perhaps northern Arabian ones.

Likewise, Cypriot Arabic does not know the pseudo-verb *bVdd*- “to want” and instead makes use of a verb derived from the root *rwḍ, piri* (< *birīd “he wants”; Borg 2004, p. 256; cf. Classical Arabic *yurīdu*). This it shares in common with the Qəltu dialects, while the modern dialects of the Levantine coast employ *bVdd*. The latter may also find its source in the North Arabian dialects, where “to want” can be expressed with the prepositional phrase, (*i*)*b-widd*-PN, or simply with *widd*-PN, literally meaning “in PN’s wish” and “PN’s wish”, respectively. If we employ an archaeological metaphor, a dialect area, such as the Levant, can be regarded as an archaeological section. The layers would reflect different chronological strata of contact-based features and local innovations. While *šāf* and *bVdd* may reflect relatively late layers, this paper is interested in identifying the very earliest linguistic strata in the modern vernaculars.

Almost all who have discussed Arabic’s past begin its historical period with the Quran and the nearly contemporary oral poems, passed on traditionally from *rāwī* to *rāwī* until achieving written form in the 8th–9th centuries at the earliest. The Quran itself is far from a linguistic unity. It minimally comprises a consonantal text, *rasm*, which reflects the local dialect of the Ḥigāz, while the reading traditions imposed upon it draw on various 7th and 8th c. varieties. The combination of these two linguistic types sometimes produces features that may never have been used in spoken language (Van Putten 2021, §3.4; Al-Jallad 2020b, pp. 57–72). Likewise, the oral poems can provide us with a glimpse of the performance language of that particular tradition, but we cannot know how much the odes changed over time as they were passed from generation to generation. Finally, their linguistic unity is little more than an assumption rather than a demonstrable fact. No one has yet, as far as I know, engaged in a truly comparative examination of the poetic tradition’s language on its own terms.

Another corpus suitable for comparison exists: pre-Islamic epigraphy.¹ These texts, which are carved in nearly half a dozen scripts, offer both advantages and disadvantages. To begin with the latter, the inscriptions do not belong to a living tradition. While the researcher has the work of early Islamic philologists to rely upon when approaching the *Qaṣīdah* odes and the Quran, the meaning of the pre-Islamic inscriptions must be reconstructed. However, with a proper comparative approach, and with due attention to archaeological and historical contexts, one can be confident about the meaning and grammar of a large part of the corpus. Nevertheless, the consonantal Semitic scripts that encode these ancient Arabic vernaculars provide us with a very limited view of their phonologies and morphology.

These materials come with advantages as well. We can be sure that their language was not filtered through later, Classicizing traditions. They reflect a register of Arabic used at the time they were produced, and since many are simple graffiti, they likely reflect something close the vernacular of their writers. The pre-Islamic inscriptions, moreover, stretch much further into the past than the pre-Islamic odes, as far back as the middle of the first millennium BCE if not earlier, and cover a wider geographic area, spanning from the Syrian desert to the Yemeni frontier.

As such, how can this corpus aid in the understanding of the linguistic history of the Arabic vernaculars? The answer is not straightforward. In some cases, we may posit a direct developmental trajectory between a phenomenon attested in the ancient sources, but in others, similarities may point towards parallel developments in the history of the language. The following pages will identify three features that the modern dialects share with the ancient epigraphy to the exclusion of normative Classical Arabic. I would suggest that these are reflective of the earliest linguistic layer of present-day vernacular Arabic.

2. Wawation

Proto-Arabic inherited the Proto-Semitic case system with only a few changes, including the emergence of a new declension (Huehnergard 2017; Al-Jallad and van Putten 2017;

Al-Jallad forthcoming), but the case system began to disappear in several ancient dialects of Arabic at approximately the turn of our era, mainly concentrated in the Nabataean realm (Corriente 1976; Blau 2006). The first stage of this process appears to have been the loss of final short vowels and then the loss of *nunation* (tanwīn), which resulted in a new set of final vowels in triptotic nouns. While a couple of inscriptions attest a functional declensional system in this state, the majority situation generalizes the nominative ending in all syntactic positions.² This feature—conventionally termed *wawation*—is encountered not only in the Nabataean inscriptions, but wherever one finds triptotic Arabic names in the Aramaic inscriptions of the first millennium BCE and the first half of the first millennium CE. Perhaps the earliest attestation of this feature in the Aramaic script is found in the 5th c. BCE votive inscription of Qaynu son of Guṣam king of Qaydar at Tell Maskhūṭah, Egypt (Rabinowitz 1956). Wawation is attested continuously throughout the centuries in northern Arabic dialects, appearing on the anthroponyms and tribal names in the Namārah inscription and even in 6th c. CE Arabic inscriptions from Syria and North Arabia (Al-Jallad forthcoming).

Tell Maskhūṭah (5th c. BCE)

C zy *qynw br gšm mlk qdr qrb l-hnʾlt*

“That which **Qaynu** son of Guṣam has offered to han-ʾĪlat (the goddess)”

Namārah inscription (S. Syria) (328 CE)

w-mlk ʾl-šryn w-nzrw w-mlwk-hm w-ḥrb mdḥgw

“He ruled the two Syrias and **Nizāru** and their kings and waged war upon **Madḥigu**”

Ḥarrān inscription (S. Syria) (568 CE)

ʾn šrḥyl br ʾlmw

“I am Ṣaraḥīl son of **Zālimu**”

The distribution of ancient *wawation* is as follows: with a few exceptions, it appears on triptotic anthroponyms and on Arabic proper nouns. It does not attach to names terminating with the feminine ending *-at*, nor does it attach to diptotic names belonging to patterns such as *fuʿal*, *ʾafʿal*, and *fʿlān* or names defined by the article. It is reasonable to assume that this distribution applied to nouns as well, although it is impossible to prove as there are so few examples of Arabic prose written in the Classical Nabataean script. JSNab 17, an Arabic inscription carved in the Nabataean script from Madāʾin Šāliḥ (dated 267 CE; Fiema et al. 2015), marks all triptotic nouns with wawation, including definite forms: *ʾlḥgrw* = ʾal-Ḥiḡr, the ancient name of Madāʾin Šāliḥ, *ʾlqbrw* = ʾal-qabru ‘the grave’ (Fiema et al. 2015). While wawation does not apply to anthroponyms with the definite article—for example, the name *marʾalqays* (=imruʾulqays) is always written *mrʾlqys* and never *mrʾlqysw*—its application appears to have been extended in the realm of nominal morphology, at least in some varieties.

The *u* termination is also encountered in the modern Arabic vernaculars of southwest Arabia, concentrated in the Yemeni Tihāmah, extending as far north as the dialect of Balqarn (Behnstedt 2016, p. 81; Greenman 1979; Alqahtani 2015). Nouns terminating in a non-etymological *u* have a distribution virtually identical to anthroponyms terminating in *waw* in the ancient inscriptions: it is restricted to triptotic nouns and does not occur on nouns with the feminine ending *-at*. The striking congruence of both of these systems motivated Blau (2006) to compare them directly. While he stops short of suggesting a genealogical relationship between the dialects of Southwest Arabia and the ancient North Arabian dialects, the particular sequences of changes required to produce a nearly identical distribution at both ends of the ancient Arabic sprachraum does suggest that the feature may share a common ancestor.

The Southwest Arabian dialects, however, attest an important difference. There are some dialects where wawation is in complementary distribution with tanwīn. The former appears in pause and the latter in context. Nöldeke was the first to hypothesize

that the Nabataean *w* had developed from *-un*, but in these Tihāmī dialects we see the process in action. The asymmetric situation is rare, isolated to a few dialects of the ʿAsīr (Behnstedt 2016, p. 81). Rather, most dialects of the area have generalized one form. Those on the Tihāmī coast have generalized *u* while most in the ʿAsīr have only the nunated ending, either *un* or *in*. Thus, as Blau (2006) suggested, the following relative chronology appears secure (Figure 1):

Stage 1	<i>al-baytu</i>	<i>baytun</i>
Stage 2	<i>al-bayt</i>	<i>baytun</i>
Stage 3	<i>al-bayt</i>	<i>baytu (pause) – baytun (context)</i>
Stage 4	<i>al-bayt - baytu (generalized)</i>	<i>al-bayt - baytun/in (generalized)</i>

Figure 1. Stages in the development of wawation.

Those dialects exhibiting the *baytu/baytVn* opposition appear to be more archaic than the Nabataean situation at first glance, but this may simply be an accident of attestation. Since most of the nouns attested in Nabataean occur in an Aramaic linguistic setting, it may be the case that their attested forms are pausal. While there is no direct evidence for the preservation of nunation in Nabataean inscriptions, a clue might be found in the Nahal Hever papyri, which are first c. CE legal documents from the Dead Sea area. The Arabic noun for “contract” is attested with an otiose final *nūn*, ʿ*qdn*. Although Yardeni (2014) suggested that this could possibly be a first person pronominal suffix, it would make little sense in this context. Rather, one could carefully hypothesize that it be interpreted as the ad hoc writing of context form, with nunation. An even earlier example of functional nunation is attested in a widely known yet unpublished inscription from the Taymāʿ area. The text—carved in an oasis North Arabian alphabet—was authored by the king of Dūmat (mod. Dawmat al-Jandal) and can be dated to the middle of the 6th c. BCE based on its reference to the Babylonian king Nabonidus. All non-pausal, non-construct, and non-diptotic nouns terminate in a *nun*.³

The Bsrn inscription

ʿ*n* : *bsrn* : ʿ*bd* : *nbwn*ʿ*d* : *mlk* : *bbl* : *nzrt* : *h-ḡnm* : *b-m*ʿ*tn* : *frsn* : *w-m*ʿ*tn* : *rkb* : ʿ*bl*

ʿI am Bsrn servant of Nabonidus king of Babylon; I have guarded the spoils with a cavalry unit and a unit of cameleersʿ

The phrase *m*ʿ*t frs* “cavalry unit” is widely attested in the Safaitic inscriptions, which are about half a century later (Macdonald 2014). The appearance of *nūns* in this inscription suggest that the two words do not form a genitive construction but rather a noun and adverb, *bi-miʿatin farasan*. The final word of the inscription, ʿ*bl*, lacks a *nūn*, perhaps suggesting that it is a pausal form.

This distribution could indicate that both the ancient northern Arabic dialects and those of southwest Arabia share a common ancestor that had undergone the changes described above. Over the passage of time, each group altered the asymmetric *pausal vs. context* distribution by generalizing one form. The *u* termination was eventually favored in Nabataean and the Tihāmāh while the nunated form was favored elsewhere. Some varieties of Nabataean further generalized wawated forms to the definite declension as well, producing the situation we find in JSNab 17.

If the genealogical connection between these two dialect groups is correct, then it may suggest that an ancient dialect of Arabic similar to what is attested in the Bsrn inscription moved south sometime in the first millennium CE and replaced the pre-Arabic languages of the ʿAsīr and Tihāmāh.⁴ We should further note that Nabataean Arabic and the dialects of southwest Arabia differ in the form of the definite article, *al* and *am* respectively. Thus, it is possible that the definite article of the ancestral dialect to both was *han-*, as attested in the Tell al-Maskhūṭah inscription. This morpheme split into ʿ*al-* in the north and ʿ*am* in the south (on the chronology of the Arabic article, see Al-Jallad 2021) (Figure 2).

	Proto-Dialect	<i>han-baytu baytu</i>			
Northern	<i>(h)al-bayt baytVn – baytu</i>		‘Asīrī	<i>am-bayt baytVn – baytu</i>	
Nabataean	<i>‘al-bayt(u) baytu</i>		Tihāmī	<i>am-bayt baytu</i>	

Figure 2. Evolution of wawation in Nabataean, ‘Asīrī, and Tihāmī Arabic.

Wawation is today not only attested in southwest Arabia. It is also found in the Levant and Mesopotamia, where it is realized as *u* or *o*, depending on the dialect. It has a much more restricted distribution: the feature is found on high frequency kinship terms, such as Levantine Arabic *‘ammu* “paternal uncle”, *ḥālu* “maternal uncle”, *sīdu* “grandfather”, *ḡaddu* “idem.”, and on feminine nouns, *ḥāltu* “maternal aunt”, etc. In northern Mesopotamia the *u/o*-termination applies only to masculine kinship terms, while feminine nouns terminate in *-a*; in Mardin, feminine vocative nouns terminate in *-e*. This distribution speaks against viewing the suffix as a third person masculine singular clitic; there would be no reason that it should be restricted to masculine nouns. Grigore (2007, p. 203) suggested that, at least for the dialect of Mardin, the termination could have a Kurdish source, but Procházka favors a Semitic origin as its distribution extends far beyond the areas in which Persian or Kurdish influence would seem possible (Procházka 2020, pp. 95–96). If I may go further, I would suggest, given the broader Arabic context, that the *u/o*-termination is a reflex of wawation as attested in Nabataean and in the southwestern Arabic dialects. The distribution in the Mesopotamian dialects matches the situation in Nabataean—it does not apply to nouns terminating in the feminine ending. The etymology of the feminine *-a* remains unclear. Perhaps Grigore (2007, p. 203) is correct to see a connection with Kurdish. While the masculine wawated form would have had an Arabic origin, speakers could have understood it as the same morpheme as the Kurdish vocative ending in a bilingual setting. The absence of any marking on feminine kinship terms perhaps motivated the borrowing of the Kurdish feminine ending to produce an etymologically mixed paradigm nearly identical with the Kurdish vocative paradigm.

The Levantine dialects appear to have extended the domain of wawation through analogy, appending the suffix to the female counterparts of male kinship terms; a similar extension of nunation occurred in Classical Arabic as Van Putten (2017) convincingly reconstructs the feminine ending as diptotic in Proto-Arabic.

The Levantine situation may, therefore, reflect a continuation of ancient Nabataean-type wawation, which survived marginally while the rest of the nominal system shifted—either through contact or through internal development—to favor the non-wawated paradigm. The early 6th century CE Arabic inscription from Jebel Usays⁵ already demonstrates that the local Levantine dialects of Arabic had dispensed with wawation on personal names and nouns; thus, it is already possible at this point that the feature was restricted to kinship terms. It is not surprising that kinship terms would preserve older layers of morphology, and so this solution, if correct, would provide a unified analysis of wawation across Arabic.

To conclude, the linguistic stratum of wawation in the Levantine and northern Mesopotamian dialects, the ancient dialects of the southern Levant, and the modern Tihāmī and ‘Asīrī dialects would appear to share a non-Classical Arabic common ancestor with this distinct declensional profile.

3. 1CS Genitive Clitic Pronoun

The next feature I would like to consider is the 1CS genitive clitic pronoun. In all forms of Arabic, the shape of this pronoun is dependent upon the termination of the noun to which it attaches, as in other Semitic languages, but its distribution can vary from dialect to dialect. The pronoun has two allomorphs: *-ī* and *-ya*.

*ī	<p>Classical Arabic: conditioned—following short vowels or consonants: <i>kitāb-ī</i> Ugaritic: conditioned—∅ = /ī/ on nominative singular + fem. pl. nouns Phoenician: conditioned—∅ = /ī/, nominative + accusative</p>
*ya	<p>Classical Arabic: conditioned—following long vowels and diphthongs: <i>‘alay-ya</i> G^oʿaz: unconditioned—<i>hagarē-ya</i> Ugaritic: conditioned—y = /ya/, gen + acc singular, and other nouns; on prepositions Phoenician: conditioned—y = */ya/, genitive nouns</p>

Some contemporary Arabic dialects, most notably those spoken in North Africa, employ the **ya* allomorph following certain prepositions: Maghrebian *liya* “to, for me”; *biya* “in/by me”, in contrast to normative Classical Arabic *lī* and *bī*, respectively. This distribution may in fact not be innovative. Various Quranic reading traditions produce such forms, but perhaps more importantly, the *rasm* itself demonstrates that this allomorph was in existence and had a much wider distribution.

Quran

69:19

فَأَمَّا مَنْ أُوْتِيَ كِتَابَهُ يَمِينِهِ فَيَقُولُ هَذَا مَا أَدْرَأْتُ وَآؤُا كِتَابِيهِ

fa`ammā man`ūtiya kitāba-hū bi-yamīni-hī fa-yaqūlu hā`umu qra`ū kitāb-iyah
 “and whosoever has received his record in his right hand will exclaim—Behold! Read aloud **my record**”

69:20

إِنِّي ظَنَنْتُ أَنِّي مُلْكٌ حَمَائِيهِ

`innī ḡanantu`annī mulāqin ḡisāb-iyah
 “I had thought that I would surely face **my doom**”

69:28

مَا أَغْنَىٰ عَنِّي مَالِيهِ

mā`aḡnā`annī mā-l-iyah
 “**My wealth** has not availed me”

In Sūrat al-Ḥāqqah, the termination *iyah*, where the final *h* should be understood as *hā`u s-sakt*, i.e., a pausal *h* following a short vowel, is used on nouns that are syntactically nominative (*māliyah*) and accusative (*kitābiyah* and *ḡisābiyah*). The employment of the *ya* allomorph in these contexts is certainly motivated by rhyme, but there are other places in the Quran that demonstrate that its conditioning environment was slightly different from normative Classical Arabic. The vocative expression in Quran 12:84, *ناسعى*, is read by Ḥafṣ as *yā`asafā* and by al-Kisā`ī as *yā`asafē*, translated as “woe to me” (lit. O my woe). Q 5:31 attests a similar construction, *بولى*, Ḥafṣ *yā waylatā*, al-Kisā`ī *yā waylatē*. The *alif maqṣūrah*, read by Ḥafṣ as *ā* and al-Kisā`ī as *ē*, reflects the outcome of an original triphthong, **yā`asafa-ya* > *yā`asafē* (Old Ḥigāzī; al-Kisā`ī) and *yā`asafā* (Ḥafṣ) (Al-Jallad and van Putten 2017, pp. 113–14). Thus, these expressions preserve a situation where Arabic deployed the *ya* suffix following a short /a/, the accusative. Finally, in agreement with the modern North African varieties, the first person clitic following the preposition *li-* is sometimes realized as *ya*, depending on the reading tradition. Ḥafṣ reads *لى* *liya*, for example, in Q 36:22.

The pre-Islamic Arabic inscriptions also attest a different distribution of the *-ya* allomorph. The Safaitic inscription BES15 799 attests a construction that is identical to the Quranic use of the *-ya* allomorph in the vocative.⁶

BES15 799

wgd sfr bny f tql 'l-bny w ql ḡbly

“he found the inscription of Bonayy and was weighed down (by grief) on account of Bonayy and said: woe to me (*ḡabla-ya* lit. O my woe)”

The use of the *-ya* allomorph following the short high vowel /i/ is also attested in the pre-Islamic corpus. A Thamudic D inscription from the northern Ḥiǧāz attests this allomorph following the preposition *bi*.⁷

UdhThamD 1 = JSTham 213

rbt šq by ʾll kn ʾmt škrn

‘There is much longing in me (*biya*) for Kn the maidservant of škrn.’

Finally, the Dumaitic inscription WDum 3 = WTI 23 attests the *-ya* allomorph on a noun which is syntactically in the genitive case. Its presence implies that the genitive ending was still productive in this stage of the language.⁸

WDum 3; WTI 23

h rđw w nhy w ʿtrsm sʿd-n ʿl-wdd-y

‘O Ruṣaw and Nuhay and ʿAttarsamē, help me in the matter of my wish (*widādiya*)’

The combination of these facts indicates that the Proto-Arabic distribution of the *i* and *ya* allomorphs of the 1CS genitive pronoun was different from normative Classical Arabic. Rather, its appearance following the accusative in vocatives /a/, and short /i/, following prepositions like *li* and *bi*, and the genitive in Dumaitic, indicates a distribution similar to Ugaritic. Thus, we can reconstruct the Proto-Arabic situation as such:

Nouns

Nom: **gamal-ī*

Gen: **gamali-ya* (Attested: Dumaitic; relics: QCT)

Acc: **gamala-ya* (Relics: vocative in QCT and Safaitic)

i-vowel prepositions:

**li-ya*

**bi-ya*

Long vowels + diphthongs

**alay-ya*

**yadā-ya*

In this light, modern vernaculars that exhibit forms such as *biya* and *liya* continue the ancient situation, while Classical Arabic is innovative in its generalizing of the *-ī* ending to these prepositions. As one reviewer of this paper pointed out to me, the quality of the vowel of the preposition in the Maghrebian varieties suggests that its immediate ancestor was long, *liya* < **liya*. Maghrebian Arabic generally loses etymologically short vowels, except in unstressed function words, where they are reanalyzed as long, e.g., the third masculine plural pronoun *hūma* < *hum*. Thus, an original **liya* would have plausibly yielded *liya*; the same applies to the form *biya*.

The vocative form may also be attested in some modern dialects. In some Levantine dialects, the expression *yābāye* is used in situations of distress. It translates literally as “O my father.” If the expression goes back to **yā ʾabā-yah*, with *hāʾu s-sakt*, then it would parallel similar constructions in the Quran and Safaitic.

Hāʾu s-sakt must be reconstructed for the ancestor of the forms *liya* and *biya* as well. The presence of a final *a* in these cases is anomalous, as final-short vowels, including *a*, have generally been lost in the modern vernaculars (Figure 3).

Proto-West Semitic	Classical Arabic	Egyptian
* <i>kʾatala</i>	<i>qatala</i>	ʾatal
* <i>tahta</i>	<i>tahta</i>	taht
* <i>sāmi ʾina</i>	<i>sāmi ʾina</i>	sam ʾin

Figure 3. Loss of final *a* in Modern Egyptian.

Thus, the survival of the vowel suggests the presence of a final *h*, protecting it from apocope. In other words, the antecedent of dialectal *biya* was not **biya* but rather **biyah*, as attested in Sūrat al-Ḥāqqaḥ.

To conclude, both the distribution and form of the 1CS genitive clitic pronoun in the modern dialects speaks against a Classical Arabic origin, but should rather be connected with phenomena attested marginally in the QCT and in the ancient inscriptions.

4. Relative Pronoun

The relative pronoun ^ʾ*alladī* is restricted to southwest Arabia today (Behnstedt 2016, p. 74), but in former times it was much more widely distributed (Holes 2018a, p. 13). It is the primary form attested in Middle Arabic texts, even those that are quite close to the vernacular. It is attested in the Damascus Psalm Fragment as ελλεδι (8th–early 9th c.; Al-Jallad 2020b, p. 26). If this form was common in medieval vernaculars, it has today given way to the virtually pan-Arabic relative pronoun *^ʾ*alli* (Stokes 2018). Yet *alladī* seems to have spread at the expense of an earlier relative pronoun *dV*:. To the Arabic Grammarians, *dV*: was characteristic of the dialects of southwest Arabia, where it can still be heard today, and the Najdi dialect of Ṭayyiʾ (Rabin 1951, chps. 3 and 14). In the modern dialects, *d*-base relatives are common in Southwest Arabia (Behnstedt 2016, p. 74) and in the Maghreb (Aguadé 2018, p. 54). The genitive particles *dīl* and *dēl* (lit. “that which is for”) in the Qeltu dialects and marginally in the Levant also suggest that at one point the relative pronoun of those dialects was a simple *d*-base form (Procházka 2018, p. 280; Lentin 2018, p. 195).

The relative *dV*: is attested across the pre-Islamic Arabic Sprachraum (Figure 4)—indeed, the form ^ʾ*alladī* has not yet appeared in the pre-Islamic epigraphic record, although its feminine counterpart ^ʾ*allatī* has been attested once in the Ḥigāz.

Script	Location	Form
Hismaic	S. Jordan/N.W. Saudi Arabia	<i>d</i> (ms); <i>dʾt</i> (fs)
Safaitic	Syro-Arabian <i>harrāh</i>	<i>d</i> (ms); <i>dʾt</i> , <i>dʾt</i> (fs); <i>dʷ</i> (mpl)
Nabataean	Southern Levant–North Arabia	<i>dʷ</i> (generalized?)
Dadanitic	Al-ʿUlā/NW Arabia	^ʾ <i>it</i> (fs)
Thamudic D	Medina	<i>d</i> (ms)
Ancient South Arabian musnad	Qaryat al-Fāw	<i>dʷ</i> (mpl)

Figure 4. Distribution of relative pronouns in the epigraphic record; data from Al-Jallad (2018).

In at least Safaitic and Hismaic it seems to inflect for case, gender, and number, with the plural form appearing as *dʷ* /*dawū*/ or /*dawī*/). Even as far south as Qaryat al-Faw, in the linguistically mixed inscription from the site, the Rbbl bin Hf m grave inscription, the plural form is attested as *dʷ* (Beeston 1979; Al-Jallad 2014). In Safaitic the relative may rarely agree in definiteness with its antecedent, producing *hd* /*hadḏī*/.

The presence of the *d*-base relative pronoun in all other branches of Semitic permits its secure reconstruction to Proto-Arabic, although there is not enough information to determine the details of its inflectional paradigm (Huehnergard 2017, pp. 16–17). This in turn indicates that the *^ʾ*alladī* and later *^ʾ*alli* forms are innovative, and spread at a later period, similar to *šāf* and *bVdd* discussed in the introduction.

Since *dV*: is an archaism it cannot be used to argue for a shared genealogical relationship between the dialects that preserve traces of it. It does, however, demonstrate that these dialects do not descend linearly from Classical Arabic, which had replaced this form with the *alladī*-type relative. Moreover, its presence throughout pre-Islamic Arabic prevents us from assuming that the *d*-base relative pronoun in the modern vernaculars is a result of “South Arabian” influence, as has been previously suggested (Corriente 2007). The relative was not bound to a single geographic area in pre-Islamic times, but was in use from Yemen to Syria. Rather, it was the *alladī*-type relative that appears to have had a specific geographic distribution, restricted to the Ḥigāz. Today’s dialect geography reflects a reversal of the pre-Islamic situation. The *alladī*-type relative, including **alli*, has spread at the expense of the older *d*-type, which is today restricted to the periphery of the Arabic sprachraum.

5. Concluding Remarks

The features discussed here are but a small sample of possible Old Arabic relics strewn throughout modern Arabic vernaculars. They nevertheless motivate one to think in terms of a three-dimensional dialect continuum, extending not only geographically but also chronologically. Interdialectal contact, substrate contributions from the pre-Arabic languages of all regions to which Arabic spread, and the heavy superstrate influence of Classical Arabic prevent us from regarding any dialect as a monogenetic descendent of a pre-Islamic variety. Yet there can be no doubt that pre-Islamic phonological and morphological features absent in Classical Arabic contributed to the formation of the modern vernaculars.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A



Figure A1. Safaitic Inscription BES15 799 (courtesy OCIANA).

BS 821: *l mğyr bn msk bn ʿmd bn mlk w wgd sfr bny f ʿql ʿl bny w ql ḥbly w ḍkr rgl f ʿdn*
 {ʿll rgl

“By Moğayyer son of Māsek son of ʿamīd son of Mālek and he found the writing of Bonayy and was weighed down by grief for Bonayy and said “O my woe” and he remembered Rāgel and was debased (by grief) for Rāgel”

Commentary:

This text was discovered in the Jordanian Ḥarrah at 32.43341; 37.270460, during the 2015 campaign of the Badia Epigraphic Survey project. The author produced three other Safaitic inscriptions KRS 38, 1885, and 1886, in the same general region.

wgd sfr: “he found the writing”, a common inscriptional genre produced upon the finding of the inscription’s of distant or deceased loved ones.

ʿql: “he was weighed down”, cf. Classical Arabic *taqula*. The verb is only attested in grieving contexts and so should be construed as a metaphor for worry and grief.

ql ḥbly: “he said: woe to me!” The meaning of this line was discussed in section three of this paper. A similar expression is attested in KRS 941: *w ql ḥbl-h trḥ* “sorrow afflicted him”.
w ḏkr rgl w ʾdn: “he remembered Rgl (likely vocalized as Rāgel) and was debased. ʾdn, the causative of *danna* “to make lowly” should be construed as a passive here with an unexpressed agent, namely, grief.

Notes

- 1 For a summary and linguistic classification of these texts, see Al-Jallad (2018) and Macdonald (2004).
- 2 The inscriptions that continue to exhibit a living case system are the En Avdat inscription (see Macdonald’s contribution to Fiema et al. 2015); the inflection of Nabataean theophoric names such as *ʿbdʾlbʿly* /*ʿabdu-ʾal-baʿli*/; and the Jebel Ramm Hismaic inscription (Macdonald 2018a, 2018b; Al-Jallad 2020a).
- 3 The text was published on Twitter by its discoverer, mr. Aqla al-Rabiah: <https://twitter.com/aqlaalrbeah/status/1293867413197520896> (accessed on 18 October 2021). A preliminary edition of the text can be found here: <https://safaitic.blogspot.com/2021/06/king-of-ancient-dumat-addumatu.html> (accessed on 18 October 2021).
- 4 The movement from north to south is assured by the chronology of the inscriptions. Wawation of this sort appears to be in place as early as the 5th c. BCE in the north, while at the same time Southwestern Arabia was dominated by the Ancient South Arabian languages. The pre-Arabic situation in the ʿAsir is so far unclear, but the existence of a number of texts from the region that defy interpretation indicates considerable linguistic diversity before Arabic dominated the region; see <https://safaitic.blogspot.com/2021/08/more-pre-arabic-texts-from-asir.html> (accessed on 18 October 2021). It is unclear when the process of Arabization began, but the first appearance of Arabic-like features in the inscriptional record dates to the turn of the era, the same time when groups called *ʿrb* appear in the inscriptions.
- 5 This inscription begins with the author’s name *ʾnʾ rḳym bn mʾrf ʾl-ʾwsy* “I am Ruḳaym son of Muʿarrif the Aws-ite”. Wawation would be expected on both Rḳym and Mʾrf according to its normal distribution in the 6th c. Arabic inscriptions and in Nabataean. On this text, see Macdonald’s contribution to Fiema et al. (2015).
- 6 This inscription had not been previously edited. See Appendix A for the edition. Note that *y* only has a consonantal value in Safaitic and cannot indicate word-final *i*.
- 7 This is my interpretation of the text, based on parallels in other Thamudic D inscriptions. The *editio princeps* differs from my reading. See Macdonald (2018a, 2018b).
- 8 Note that *matres lectiones* are not used in the orthography of the Oasis North Arabian scripts to indicate final long vowels, as shown with the verb *sʾdn* /*sāʾidu-nī*/ “help me”. The final *y* of the inscription must therefore be consonantal.

References

- Aguadé, Jordi. 2018. The Maghrebi dialects of Arabic. In *Arabic Historical Dialectology*. Edited by Clive Holes. Oxford: Oxford University Press, pp. 29–63.
- Al-Jallad, Ahmad, and Marijn van Putten. 2017. The Case for Proto-Semitic and Proto-Arabic Case: A reply to Jonathan Owens. *Romano-Arabica* 17: 87–117.
- Al-Jallad, Ahmad. 2014. On the genetic background of the Rbbl bn Hfʾm grave inscription at Qaryat al- Fāw. *Bulletin of the School of Oriental and Asian Studies* 77: 445–65. [CrossRef]
- Al-Jallad, Ahmad. 2015. On the voiceless reflex of *ʃand *ṭin pre-Hilalian Maghrebian Arabic. *Zeitschrift für Arabische Linguistik* 62: 88–95.
- Al-Jallad, Ahmad. 2018. What is Ancient North Arabian? In *Re-Engaging Comparative Semitic and Arabic Studies*. Edited by Daniel Birnstiel and Naʾama Pat-El. Wiesbaden: Harrassowitz, pp. 1–45.
- Al-Jallad, Ahmad. 2020a. Notes on the language of the Hismaic Inscriptions and a re-reading of Line 1 of the Madaba Hismaic Inscription. *Journal of the Royal Asiatic Society* 30: 561–69. [CrossRef]
- Al-Jallad, Ahmad. 2020b. *The Damascus Psalm Fragment: Middle Arabic and the Legacy of Old Ḥigāzī*. With a contribution by Ronny Vollandt. Chicago: Oriental Institute.
- Al-Jallad, Ahmad. 2021. The History of the am-Definite Article—South Arabian or Arabic? *Zeitschrift für Arabische Linguistik* 73: 53–70. [CrossRef]
- Al-Jallad, Ahmad. forthcoming a. “Night” in Proto-Semitic and the emergence of a new nominal declension in Arabic. *Studia Orientalia Electronica*.
- Al-Jallad, Ahmad. forthcoming b. One wāw to rule them all: The origin and fate of wawation in Arabic. In *Scripts and Scripture*. Edited by Fred Donner and Rebecca Hasselbach. Chicago: Oriental Institute.
- Alqahtani, Khairieh. 2015. A Sociolinguistic Study of the Tihami Qahtani Dialect in Asir, Southern Arabia. Ph.D. thesis, University of Essex, Essex, UK.
- Beeston, Alfred F. L. 1979. Namarah and Fāw. *Bulletin of the School of Oriental and African Studies* 42 1: 1–6.
- Behnstedt, Peter. 2016. *Dialect Atlas of North Yemen and Adjacent Areas*. Leiden: Brill.

- Blau, Joshua. 2006. Problems of Noun Inflection in Arabic: Reflections on the Diptote Declension. In *Biblical Hebrew in Its Northwest Semitic Setting: Typological and Historical Perspectives*. Edited by Steven E. Fassberg and Avi M. Hurvitz. Jerusalem: Magnes Press, Winona Lake: Eisenbrauns, pp. 27–32.
- Borg, Alexander. 2004. *A Comparative Glossary of Cypriot Maronite Arabic (Arabic-English): With an Introductory Essay*. Leiden: Brill.
- Corriente, Frederico. 1976. From Old Arabic to Classical Arabic through the Pre-Islamic Koiné: Some Notes on the Native Grammarians' Sources, Attitudes and Goals. *Journal of Semitic Studies* 21: 62–98. [CrossRef]
- Corriente, Frederico. 2007. On the prehistory of the Arabic language. *Aula Orientalis* 25: 141–53.
- Fiema, Zbigniew T., Ahmad Al-Jallad, Michael C. A. Macdonald, and Laïla Nehmé. 2015. Provincia Arabia: Nabataea, the Emergence of Arabic as a Written Language, and Graeco-Arabica. In *Arabs and Empires before Islam*. Edited by Greg Fisher. Oxford: Oxford University Press, pp. 373–433.
- Greenman, Joseph. 1979. A Sketch of the Arabic Dialect of the Central Yamani Tihamah. *Zeitschrift für Arabische Linguistik* 3: 289–304.
- Grigore, George. 2007. *L'arabe parlé à Mardin: Monographie d'un parler arabe "périphérique"*. Bucharest: Editura Universității din București.
- Holes, Clive. 2018a. Introduction. In *Arabic Historical Dialectology*. Edited by Clive Holes. Oxford: Oxford University Press, pp. 1–28.
- Holes, Clive. 2018b. The Arabic Dialects of the Gulf: Aspects of their historical and sociotic development. In *Arabic Historical Dialectology*. Edited by Clive Holes. Oxford: Oxford University Press, pp. 112–48.
- Huehnergard, John. 2017. Arabic in its Semitic Context. In *Arabic in Context*. Edited by Ahmad Al-Jallad. Leiden: Brill, pp. 3–34.
- Lentin, Jérôme. 2018. The Levant. In *Arabic Historical Dialectology*. Edited by Clive Holes. Oxford: Oxford University Press, pp. 170–205.
- Macdonald, Michael C. A. 2004. Ancient North Arabian. In *The Cambridge Encyclopedia of the World's Ancient Languages*. Edited by Roger D. Woodard. Cambridge: Cambridge University Press, pp. 488–533.
- Macdonald, Michael C. A. 2014. Romans go Home? Rome and other 'outsiders' as viewed from the Syro-Arabian Desert. In *Inside and Out. Interactions between Rome and the Peoples on the Arabian and Egyptian Frontiers in Late Antiquity*. Edited by Jitse H. F. Dijkstra and Greg Fisher. Leuven: Peeters, pp. 145–64.
- Macdonald, Michael C. A. 2018a. Clues to How a Nabataean May Have Spoken from a Hismaic Inscription. In *Near Eastern and Arabian Essays Studies in Honour of John F. Healey*. Edited by George J. Brooke, Adrian H. W. Curtis, Muntasir al-Hamad and G. Rex Smith. Oxford: Oxford University Press, pp. 231–39.
- Macdonald, Michael C. A. 2018b. The Ancient North Arabian and Ancient South Arabian Inscriptions. In *The Darb al-Bakrah*. Edited by Laïla Nehmé. Riyadh: Saudi Commission for Tourism and National Heritage, pp. 227–84.
- Pat-El, Na'ama. 2017. Digging up Archaic Features in "Neo-Arabic" and Comparative Semitic in the Quest for Proto Arabic. In *Arabic in Context*. Edited by Ahmad Al-Jallad. Leiden: Brill, pp. 441–75.
- Procházka, Stephan. 2018. The Northern Fertile Crescent. In *Arabic Historical Dialectology*. Edited by Clive Holes. Oxford: Oxford University Press, pp. 257–92.
- Procházka, Stephan. 2020. Arabic in Iraq, Syria, and southern Turkey. In *Arabic and Contact-Induced Change*. Edited by Christopher Lucas and Stefano Manfredi. Berlin: Language Science Press, pp. 83–114.
- Rabin, Chaim. 1951. *Ancient West-Arabian*. London: Taylor's Foreign Press.
- Rabinowitz, Isaac. 1956. Aramaic Inscriptions of the Fifth Century B.C.E. from a North-Arab Shrine in Egypt. *Journal of Near Eastern Studies* 15: 1–9. [CrossRef]
- Souag, Lameen. 2016. Book Review: Werner Diem. *Negation in Arabic: A Study in Linguistic History*. Wiesbaden: Harrassowitz, 2014. 119 pp. ISBN 978-3-447-10117-2; David Wilmsen. *Arabic Indefinites, Interrogatives, and Negators: A Linguistic History of Western Dialects*. Oxford Studies in Diachronic and Historical Linguistics. Oxford: Oxford University Press, 2014, 245 pp. ISBN 978-0-19-871812-3. *Linguistics* 54: 223–29.
- Stokes, Phillip. 2018. The Plural Demonstrative Based on *VI in Arabic and the Origin of Dialectal illi. In *Re-Engaging Comparative Semitic and Arabic Studies*. Edited by Daniel Birnstiel and Na'ama Pat-El. Wiesbaden: Harrassowitz, pp. 131–54.
- Van Putten, Marijn. 2017. The Feminine Ending -at as a Diptote in the Qur'anic Consonantal Text and Its Implications for Proto-Arabic and Proto-Semitic. *Arabica* 64: 695–705. [CrossRef]
- Van Putten, Marijn. 2021. *Quranic Arabic: From Its Hijazi Origins to Its Classical Reading Traditions*. Leiden: Brill.
- Van Putten, Marijn, and Adam Benkato. 2017. The Arabic Strata in Awjila Berber. In *Arabic in Context*. Edited by Ahmad Al-Jallad. Leiden: Brill, pp. 476–502.
- Yardeni, Ada. 2014. A List of the Arabic Words Appearing in Nabataean and Aramaic Legal Documents from the Judaeian Desert. In *Scripta Classica Israelica: Rome, Judaea, and Its Neighbours: Special Issue in Honor of Hannah M. Cotton*. Edited by Werner Eck. Jerusalem: The Israeli Society for the Promotion of Classics, pp. 301–24.

Article

The Old and the New: Considerations in Arabic Historical Dialectology

Alexander Magidow

Modern and Classical Languages and Literatures, University of Rhode Island, Kingston, RI 02881, USA; amagidow@uri.edu

Abstract: Arabic historical dialectology has long been based on a historical methodology, one which seeks to link historical population movements with modern linguistic behavior. This article argues that a nexus of interrelated issues, centered around a general theme of “oldness,” has impaired this work, and proposes basic principles to avoid the misinterpretation of linguistic data. This article argues that there is a strong tendency to essentialize the idea of linguistic conservatism and attribute it to the groups that have archaic features. Against this view, it proposes that linguistic conservatism should be seen as a failure to participate in otherwise widespread innovations. It critiques the assumption that the modern dialect distribution is directly derived from the earliest settlements established during the Islamic conquests in the seventh century, arguing instead that long-term linguistic durability is unlikely. The article further challenges the assumption that highly conservative dialects such as those of Yemen are ancestral to modern dialects in a meaningful way, arguing instead that either more proximate ancestors or wave-like diffusion had a greater impact on the development of modern dialects. Finally, the paper suggests that a heuristic approach based on typical processes of language diffusion and human migration offers a more productive approach to understanding the history of Arabic dialects than a model based on historical events; many of the existing linguistic classifications may be directly derived from this heuristic.

Citation: Magidow, Alexander. 2021. The Old and the New: Considerations in Arabic Historical Dialectology. *Languages* 6: 163. <https://doi.org/10.3390/languages6040163>

Keywords: Arabic dialects; dialectology; historical dialectology; nomadism; methodology; geography; dialect geography

Academic Editors: Simone Bettega and Roberta Morano

Received: 8 July 2021

Accepted: 28 September 2021

Published: 9 October 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In his introduction to the recent volume *Arabic Historical Dialectology: Linguistic and Sociolinguistic Approaches*, Clive Holes, an expert on Bahraini and Gulf Arabic dialects, relates the following anecdote to illustrate the similarity of so-called “Bedouin” type dialects:

[W]hen, in the mid-1970s, my employer transferred me from Kuwait to Algeria, a distance of several thousand miles, I had no difficulty, if I spoke in Gulf Arabic, in making myself understood to (and in understanding) ordinary Algerians in southern oasis towns such as Ourgla and Touggourt, even though most of them had never left Algeria in their lives: we were all speaking ‘bedouin’ dialects. But the Arabic of the city of Algiers, only a few hundred miles to the north, and where I was based, is of North African ‘sedentary’ type and was so incomprehensible to me (as was my Gulf Arabic to the Algérois) that throughout my two-year residence there I found it easier to speak French. (Holes 2018a, p. 22)

This example comes in a section on Bedouin dialects, and is intended to illustrate how, despite issues with the use of “Bedouin” as a classification, it still has value as a category of analysis.¹ However, the exact mechanism by which these two far-flung dialects remain mutually intelligible—but unintelligible compared to the nearby coastal sedentary dialects—remains uninterrogated. How exactly does their status as “Bedouin” dialects render them similar? The explanation becomes simply that they are Bedouin dialects, and by existing in the same classification, they are expected to be similar, without invoking

either history or linguistics to delve deeper into that similarity. If a linguist from outside the field was presented with this example, they would almost certainly explain it simply as a matter of time-depth—clearly the Gulf dialects and the southern Algerian dialects are simply the result of a relatively recent divergence, rather than due to a vaguely defined typological similarity.²

This article will argue that it is precisely due to an accretion of traditional approaches to Arabic linguistic and linguistic history that the situation Holes describes is seen as anomalous or explainable only through categorization. Instead, if we reassess some of those existing views, this situation falls out naturally from basic linguistic and historical processes, without the need to invoke specific social categories like “Bedouin” as explanatory. The issue here is not that the category of “Bedouin dialect” itself is invalid, but rather that both the category, and the linguistic evidence, are not sufficiently interrogated. This article aims to suggest ways in which our traditional approaches to link Arabic dialectology and the social history of Arabic-speaking peoples can be profitably reconsidered, investigating a nexus of interrelated issues that center around a general theme of “oldness”: how dialectologists interpret conservative linguistic features, how they conceive of earlier versus later layers of movement and population, and what conservative dialects mean for the genealogy of modern dialects.

This paper is, at its heart, a historiographical exploration, looking at the narratives that surround the history of Arabic as much as the linguistic data itself, and how these narratives shape our conception of that history. Since it is difficult to survey the entire field in a meaningful or systematic manner in a paper of this length, much of the attention is focused on the recent survey that the quote above comes from, Holes (2018a). This seminal work is a well-researched, elegantly conceived volume which makes the historical dialectological work performed until now both easily accessible, and easily comparable, and it is doubtful this paper could have been easily written prior to the publication of that volume. Indeed, the high quality of many of the essays therein make the larger critique in this paper difficult at times, as many of the authors have indeed begun to move beyond the assumptions critiqued here. However, as argued here, those assumptions continue to influence the research on a less conscious level.

This paper is not intended to be polemical, nor is it intended as a broader criticism of the work in the field of Arabic dialectology and linguistics. The incredible work performed by scholars for the past several centuries has and continues to be of immense value, and none of the critiques laid out here could even be articulated without that work. The goal, rather, is to offer constructive suggestions to improve the depth and accuracy of the work in the field of Arabic historical dialectology. When a scholar is quoted in the process of identifying a common theme in the literature, these quotes are intended to represent a larger narrative within the field, and certainly not to criticize the author quoted or their work more generally. Indeed, much of the criticism is focused on the author’s own earlier work.

The paper also seeks to suggest concrete, actionable ways to avoid common issues in historical dialectological work. Section 5 presents a novel heuristic approach to considering how linguistic and population movements are intertwined in the South-West Asia and North Africa region based on general linguistic and social principles. Similarly, the conclusion presents several specific recommendations for future historical dialectology.

Prior to beginning, it is important to step back and consider the goals of what we seek to determine from Arabic historical dialectology research. The closest statement that Holes (2018a) makes on this topic is in a footnote: “Our main purpose in writing this book is to show how a more historically and socially grounded linguistic approach, despite the gaps in the record, can help trace the long-term dynamics and some detail of what happened [in the development of Arabic dialects today] (note 7).” In a sense, what historical Arabic dialectology seeks to do is to take the various snapshots we have of the Arabic dialects—their modern distribution and the glimpses of the dialects we find in historical records—and interpolate them to develop a narrative of how the dialects developed. We seek to be able to somehow “rewind” the historical development of the

language to see how it came to be. It is, of course, the “gaps in the historical record” that are the primary difficulty in this endeavor, but this paper will also argue that it is *how* we view the available historical and linguistic record that can, at times, hamper our progress toward that larger goal of understanding the history of Arabic dialects.

2. Conservative Features and Conservative Dialects

One key issue in Arabic dialectology is how we interpret the data that is available to us. As a preliminary, it is important here to be clear in differentiating several levels of linguistic analysis that operate at difference scales.³ The lowest level of analysis is a linguistic feature, a particular way of using language, such as the use of a certain reflex of a proto-phoneme, a certain word used to mean “to go,” an intonation pattern in declarative sentences, etc.⁴ At a higher scale is the location of that linguistic feature in space, i.e., its dialect geography, which also implies a certain point in time as well, analogous to what we find on the pages of a dialect atlas. Finally, for the purposes here, we have a dialect (or more awkwardly, “feature bundle”), which is the total collection of linguistic features spoken at a bounded area in time and space.⁵

It is important to differentiate these levels of analysis as there is a significant difference between a given feature being old and long attested, and its presence in a particular space and time being long-term. This too is different from presence alongside a cluster of other features being long-term in a particular area or among a particular speech community. The first of these can be quite easy to prove, even in Arabic where diglossia muddies the waters considerably—if we can find an early attestation, we can prove that a feature is quite old. Of course, the latter is rarely true—if we cannot prove the antiquity of a feature, it does not necessarily mean it is new or old. More difficult is to prove that a feature has been in the same location over a long period of time. Certainly Occam’s razor suggests that if we find a feature in, e.g., early Levantine Middle Arabic documents, and also today, it must have been resident in that place the entire time. However, it is easy to imagine a scenario in which two waves of movement and replacement occurred, such that the feature ceased to exist in the area, and then was replaced by a dialect that again had the feature. Finally, it is most difficult to establish the long-term durability of a dialect or cluster of features. Among other challenges in establishing dialect durability is at which threshold one considers that bundle of features to be fundamentally altered, such that a declaration of continuity, or of change, can be declared a kind of “Ship of Theseus” problem.⁶

A significant issue in the dialectology literature is in how we interpret dialects which have a preponderance of archaic features, as opposed to dialects with many innovative features. There is a strong tendency to associate conservative features with a kind of “originalism,” a primordial state that is often taken to imply long-term residency in an area, or some kind of genetic priority to more developed dialects.⁷ This is a form of essentialism in which the linguistic conservatism becomes linked to a larger conservatism that is seen as an inseparable characteristic of the dialects which have those linguistic features.

We see many examples of this conflation in the literature. [Behnstedt and Woidich \(2018, p. 81\)](#) list a variety of migrations into the Fayyum, up to and including the eleventh century, but consider the Fayyum dialect to represent “the earliest linguistic stratum” in Egypt based on conservative linguistic features. For a dialect area only a short distance from Cairo (certainly half the distance from Cairo to Alexandria), and certainly in a close relationship of trade with that city, what is remarkable is precisely that the Fayyum somehow resisted those assimilatory pressures to which Alexandria was subjected, as detailed in Section 3.1. Similarly they argue that a conservative syllable structure reflects earlier migrations to upper Egypt, while less conservative syllable structures represent later or continuing migration, independent of historical data (p. 84). [Procházka \(2018\)](#) formally distinguishes between inherited and innovative traits in his discussion of the Northern Fertile Crescent dialects, a welcome division given how often these are conflated as distinguishing features of dialects. However, he considers the inherited traits to be “‘archaic’ or ‘pre-diasporic’, i.e., going back to dialects spoken in Arabia before Islam (p. 262),” when

by definition these traits should be found in any dialect that has not innovated a new form, regardless of when it migrated into or out of an area. Similar arguments regarding linguistic conservatism as evidence of longer settlement or a vague sense of “oldness” are found throughout the volume (see pp. 1, 57, 71, 81, 136, 162–63, 264, 298, 304).

Indeed, this idea of ‘old features’ as ‘conservative’ is fundamental to the differentiation between sedentary and Bedouin dialects, and the related tendency to consider Bedouin dialects as themselves ‘conservative’ by extension. Quoting [Rosenhouse \(2006, p. 259\)](#), [Holes \(2018a, p. 20\)](#) notes that Bedouin dialects are seen as “more conservative” since they “retain many ‘Classical’ features lost elsewhere,” though even without considering Classical Arabic, many characteristically Bedouin features such as the retention of interdental are certainly retentive with respect to most nearby sedentary dialects. Lists of purportedly Bedouin features are rarely more than lists of retentions, rather than innovations, with the only innovation that commonly can be said to unite all Bedouin dialects being the use of a voiced reflex of the (Q) variable ([Palva 2006](#)).

Though Holes and many other modern authors have developed more detailed understandings of the distinctions between Bedouin and sedentary dialects, they are still viewed as fundamentally distinct from one another and form a key category in the linguistic analyses in the field. In the Holes volume, under the larger category of “major areal and typological” distinctions, there are 35 distinct entries in the index under “Bedouin vs. sedentary,” totaling over 60 pages, while related patterns such as the pre-Hilali vs. Hilali and *qultu* vs. *gilit* distinctions have a further 18 and 17 entries each. The only other categories listed under this heading are “Maghrebi vs. Mashreqi”, “peripheral vs. heartland” (a total of 50 entries) and “urban vs. rural” which often tends to functionally be a “Bedouin” vs. “sedentary” distinction, especially in the chapter on the Maghreb. The Bedouin vs. sedentary distinction is by far the most ubiquitous distinction in this volume, and one predicated primarily on the apparent conservatism of Bedouin dialects (but more accurately, the features of those dialects).

Indeed, it is notable that Bedouin dialects do not appear to be more or less innovative in general than sedentary dialects. Rather, they have participated, by and large, in different innovations than sedentary dialects. [Magidow \(forthcoming\)](#), in a sample of 52 dialects across the Arab world, divided them into Bedouin dialects if they had a voiced realization of (Q), and sedentary dialects if they did not. Out of a pool of 59 total possible innovations, Bedouin dialects showed an average of 13.1 innovations, versus sedentary dialects with an average of 13.9 innovations. As expected, there was no significant statistical difference between these groups—the two groups are effectively equally innovative.

This focus on conservatism tends to miss a key point, which is that conservatism vis-a-vis dominant linguistic features in the area is simply the result of a group failing to participate in an innovation, not necessarily a deeper statement about the history of that dialect. The fundamental observation of historical linguistics is that only successful sharing of a feature is indicative of shared history or participation in a common speech community ([Hetzron 1976](#); [Magidow 2017](#)). Sharing of linguistic features implies connection between the dialects sharing the features, while linguistic conservatism implies a lack of sufficient connection. There are only a few ways that connection can occur, and in that sense all happy linguistic families are alike in that they share many innovations, reflecting a shared past of contact.⁸

However, unhappy linguistic families, those without a connection, are often uniquely different. A failure to participate in an innovation can be caused by a wide range of factors. Dialects may simply be too far apart, unable to be exposed to a particular feature, or the speakers of two nearby dialects by virtue of their lifestyles never come into contact. Social rather than geographical barriers may play a role—a group of speakers can resist aligning themselves linguistically with their neighbors, even if they live one neighborhood over—this is precisely what happens with communal and sectarian dialects ([Blanc 1953, 1964](#); [Walters 2006](#)). There may indeed be some kind of influence from the social structure of the speakers of a dialect—the work by Lesley and James Milroy has long shown that

dense social networks can inhibit the diffusion of innovations in comparison to looser social networks (Milroy 2008).⁹ The role of network density needs to be investigated in greater depth for Arabic, but in most cases it is likely the lack of contact—a lack of frequent linguistic interactions between two populations—that explains most of the disparity between dialects.

Indeed, when there is a connection and interaction between groups over sufficiently long periods of time and little reason to resist change, we would expect to find that linguistic features would diffuse across the entire population. From this perspective, what is remarkable about Bedouin dialects is their lack of participation in innovations. If we discard the essentializing notion that they are in some fundamental, unchangeable way conservative, the most logical explanation for the deviance of Bedouin dialects from sedentary dialects is not that Bedouin dialects are somehow “old,” but rather that they are relatively new arrivals to an area. This is where the distinction between features—which may indeed be old, archaic, or non-innovative, relative to other features found in a language—and feature position in space is key. The individual features of a Bedouin dialect may be archaic, but their deviance from the features in the surrounding dialect geography is almost certainly indicative of a relatively recent movement into that area.

Another issue arises here, which is that even as a nomadic dialect might contain a variety of conservative features, the “feature-bundle” of that dialect may or may not be continuous, any more than in a sedentary dialect. The way that tribes are imagined in the dialectology literature are as relatively unchanging familial groups, but the reality is that tribal groupings are in actuality political rather than genetic entities (Hoyland 2009, p. 390). They can divide and recombine, even as traditional tribal names might be retained (Magidow 2013, pp. 119–22). Today’s Ma’qil tribe is not necessarily yesterday’s Ma’qil (or Ibn Khaldoun’s), either in terms of the genetic makeup of its members or its linguistic behavior. Here again, the distinction between conservative features and conservative dialects is key. The presence of any number of conservative features in a dialect should not necessarily be understood to mean that a specific combination of features has been used together for a long period of time. Nor does it imply that the group which uses those conservative features has had long-term cohesion and durability. It is possible that this was the case, that there has been continuity, but it should not be assumed.

Moreover, Bedouin dialects are imagined as being traditionally spoken nomadic groups subsisting in resource-poor areas with quite low population density. From the perspective of a dialect map, it takes relatively little human movement for a given space in a low-population area to change its linguistic behavior. Even a dialect map drawn at different seasons might show significant changes as groups move to summer and winter camps within these marginal areas. Contrast this with the dialect maps of high population sedentary areas where a massive catastrophe would be required to cause a migration sufficient to change the overall linguistic landscape—in these areas, one would instead expect the linguistic features to diffuse across the landscape, while the speakers themselves remained stationary.

Magidow (2013, pp. 133–34) refers to this contrast between linguistic conservatism, but recent migration, as the “*Bedouin paradox*”:

Nomadic speakers generally do not always participate in the spread of innovations among settled groups, and therefore they appear to retain archaic linguistic features in comparison with their settled neighbors. However, their extreme mobility and the ease of replacing indigenous nomadic groups means that these ‘archaic’ speakers may be newcomers to an area in comparison with settled groups.

This idea helps explain Holes’ observation about the Bedouin dialects in Algeria. The dialects that Holes reports being able to understand so well are not magically “Bedouin” in nature. Rather, they likely have a much shallower historical branching from the dialects that he was already familiar with, and had moved into southern Algeria relatively recently in comparison with the sedentary dialects of the coasts. The conservatism of these dialects

(which nonetheless have acquired “Maghrebi” features from nearby settled areas) reflects their relatively recent arrival on the dialectological scene.

3. Early Layers and Later Layers

Another key idea linked to the idea of “oldness” as deterministic in the history of Arabic dialect is the “big-bang” model of the expansion of Arabic. This model holds that it is the initial expansion of Arabic in the early Islamic period that is at essence responsible for much, if not all, of the modern geographical distribution of Arabic dialect features.¹⁰ This model has a genetic component—it is these old dialects, first distributed across what is now the Arabic-speaking world, that are the direct ancestors of the modern Arabic dialects, with changes within those dialects due to contact, urbanization or similar processes.

This concept is a foundational in Arabic dialectology. The strongest modern proponent of the idea, Jonathan Owens, explicitly designed his monography, *A Linguistic History of Arabic*, around the goal of reconstructing the Arabic of the period from 630 to 790 (2006, pp. 2–5), and continues to use a similar methodology in more recent papers (Owens 2018). My own earlier work, Magidow (2013) followed this basic assumption quite closely as well, and it is a common underlying assumption throughout the Holes volume, where the introduction focuses on what “language . . . the conquerors spoke” (Holes 2018a, p. 7). The *Encyclopedia of Arabic Language and Linguistics* article on “Dialects: Genesis” quite explicitly states that “by the 10th century, dialect areas were already shaped” in essentially their present distribution (Abboud-Hagggar 2006, p. 620). Jastrow (2002) divides between Zone I dialects, those in the Arabian Peninsula, against Zone II dialects, those “colonial” dialects that are a results of the early Islamic expansions. The idea was also key in earlier work. Ferguson’s famous idea of an Arabic *koine*, the ancestor of modern sedentary dialects, assumes that “its spread coincided roughly with the spread of urban Arabo-Islamic culture (Ferguson 1959, p. 618), and the same is essential true of Versteegh’s pidginization hypothesis (Versteegh 1984, 2004).¹¹ Even earlier approaches which assume linear descent of the Arabic dialects directly from Classical Arabic are, at their heart, assuming a diffusion of relatively similar speakers at the time of the conquests, with later developments occurring in-situ, with many of these ideas going back to even the very early grammarian traditions that spoke of Bedouin informants and dialects becoming corrupted by sedentarization (Blau 1977; Fück 1950; Garbell 1958; Versteegh 2014, p. 138).

The big-bang phenomenon also has a related phenomenon in the study of North Africa, what could be called the “little bang.” The first big-bang is shared with the rest of the Arab world, as Arab armies lead the conquest of North Africa and Andalusia in the 7th and 8th centuries. This is believed to have laid down an initial layer of Arabic, known as the “pre-Hilalian” variety of Arabic (Marçais 1938). Following this era, another major linguistic expansion occurred in the movement of tribes from the lineage of the Banū Hilāl, supposedly from the Arabian Peninsula (by way of a brief stopover in southern Egypt) beginning in the early eleventh century and ending by the fourteenth in the typical accounts. This group is said to be responsible for the “Hilalian” dialects of North Africa, a group of dialects primarily spoken by Bedouin, rural or recently urbanized populations.

If true, the big-bang idea would be extremely convenient for the historical dialectology of Arabic. Researchers would be able to ignore the complex histories that follow the time periods in which these “bangs” occurred, and instead focus on the vast, early historical tradition which reports many of the early population movements in and out of the Arabian peninsula. This would allow us to reduce the enormity of the task of Arabic historical dialectology, and to focus on linking those historical reports to the modern distribution of dialects (Aguadé 2018; Behnstedt and Woidich 2018; Magidow 2013; Procházka 2018). Unfortunately, the big-bang model appears untenable for three primary reasons. The first is that it is not clear how durable dialect geography is over time. Second, given the lack of durability of features-in-space over time, it is important to pay attention to the significant evidence that major population movements occurred well after the Islamic conquests. Finally, the model (in either the big-bang or little bang versions) simply does not

make effective linguistic predictions, with the actual linguistic features of modern dialects contradicting the predictions that these models would make.

3.1. Durability of Linguistic Material over Time

In general, there is an unarticulated assumption that linguistic features, once in-place, will generally persist over time. On a basic level, this is often true, but the Arabic speaking world is a crossroads of civilizations, with both long-term, continuously inhabited cities and vast areas of quite low population density. Indeed, the disappearance of the many languages other than Arabic following the Arab conquests gives lie to this theory, for clearly this linguistic inertia can be interrupted and once dominant languages driven extinct, like Coptic, or into a very marginal status, as with Aramaic.

There is plentiful evidence from sociolinguistics that language change can proceed extremely rapidly. Miller (2005) found that within a generation of arrival, many Upper Egyptian migrants to Cairo had assimilated to a wide variety of different Cairene linguistic features. The koineization of the Amman dialect appears to have happened within three generations, and has significantly changed the linguistic repertoire of the newly created city (Al-Wer 2003, 2007). The migration of ‘Arab dialects to Bahrain, though hailing perhaps from the 18th century, accelerated after the 1930s and so, in spite of sectarian differences, endogamous marriage within groups and other barriers, by 1995 there was already a developing areal koine (Holes 1995), and by the late 2000s, even in rural areas the old village Baharna dialects “have now all but disappeared (Holes 2015, p. 475).” One notes also that for several key variables, including the shift of (Q)/q/>/P/, Behnstedt (1997, map 9) differentiates between the oldest and youngest generations, showing change in within three generations. These kinds of changes, well attested in the sociolinguistics literature more generally, typically occur on timescales of 3–4 generations, equivalent to approximately one century (Trudgill 1986). To expect any significant linguistic durability of features-in-space, or even of dialect bundles, across longer timespans seems wildly optimistic.

Some accounts for the “big-bang” approach have attempted to formalize the idea of linguistic inertia. For example, Owens (2018, p. 209) suggests that in the framework of Dixon (1997), the Islamic conquests represent a “punctuated phase” in a larger linguistic equilibrium. Even leaving aside the many criticisms of Dixon’s model (Bowern 2006), and that it is clearly meant for longer time-periods than treated here, it is unclear how this is the only punctuated phase that is meaningful in the history of Arabic, or how long the phase lasted exactly. Arabicization took centuries in most places, and is still incomplete in many others, such as North Africa. Going back to Dixon (1997, esp. Chapter 6), virtually every form of punctuation he discusses—natural causes (e.g., plague), material innovations (especially of weapons), “development of aggressive tendencies”—happened repeatedly since the early Islamic conquests.

Magidow (2013) adopted a different approach, attempting to formalize this model of persistence using a concept from geography, adopted by Labov for sociolinguistics, the “principle of first effective settlement.” This principle states:

Whenever an empty territory undergoes settlement or an earlier population is dislodged by invaders, the specific characteristics of the first group able to effect a viable, self-perpetuating society are of crucial significance for the later social and cultural geography of the area, no matter how tiny the initial band of settlers may have been. (Zelinsky 1992, p. 13)

To which Labov (2001, p. 504) adds:

In any one generation, if the numbers of immigrants rise to an order of magnitude greater than the extant population, the doctrine may be overthrown, with quantitative changes in the general speech pattern.

Though this principle does indeed seem to match with the role of population density in acting as a barrier to linguistic change (Magidow 2013, 99ff; Ostler 2005), it is frustratingly vague, and again we simply do not have sufficient access to the complete history of the

places in question. Magidow (2013), drawing heavily on Conrad (1981), makes much of the plagues occurring immediately around the time of the Islamic conquests. However, there were clearly many subsequent plagues, including the Black Death that devastated the entire western hemisphere in the 14th century (Dols 1974). Between plague, conquest, migrations, deurbanization and urbanization due to changes in trade, climate, and other facts, Labov's "order of magnitude" criterion must have been regularly fulfilled in the millenium after the Islamic conquests.¹²

3.2. Later Population Movements

Indeed, we find that when we do look at the history of the Arab world, there are often many examples of later movements and changes that clearly post-date the early Islamic conquests, and which have significant implications for the linguistic history of a region. Even if we restrict ourselves to the chapters in Holes (2018a), we find numerous examples where the current distribution of linguistic features in space clearly are a result of post-conquest population movements.

In Egypt, Behnstedt and Woidich (2018) find many examples where the dialectological situation owes its distribution to much later phenomena. Against Owens (2003), they argue that "the constant return of Maghrebi tribes to Egypt" reinforced the use of the *niktub-niktubu* verb paradigm, and that for certain regions of the Delta these forms are "at least partly due to later Maghrebization from the fourteenth century onwards (p. 76)." For the city of Alexandria, older linguistic layers have apparently been erased. Alexandria has a long and storied history, previously having many Maghrebi features. By the time the French arrived it had only 7000 inhabitants, growing again only in the nineteenth century under Muhammad Ali Pasha. By the end of the 19th century, it continued to have non-Cairene dialect features, many of which were also Maghrebi, such as /d³/ for the (j) variable versus Cairene /9/, and common use of the *niktub-niktubu* verbal paradigm. By the 1970's it was "a 'one foot in the grave dialect' (p. 79)" effectively replaced by the Cairene dialect in middle and upper-class speech in younger people, and with older people preserving only some of the original features. These are changes which largely took place only in the last three centuries, such that the pre-modern dialect has little relationship to the current one, and that original dialect may be difficult if not impossible to reconstruct.

Indeed, while we generally would expect that large cities would be the most stable across time given high population density and durability of location, the situation of Alexandria is surprisingly common. Many cities witnessed periods of intense depopulation and repopulation, particularly in the 20th century when urban areas underwent spectacular growth (Miller 2007). Many modern cities in the Arab world are virtually ex nihilo creations, such as Casablanca (25,000 in 1900 to millions today), Amman (effectively founded in 1923) and Nouakchott (founded 1957). However, even older cities had surprisingly low populations until recent times. Table 1.1 in Miller (2007) shows the vast growth in many of the major cities of the Arab world, most of which have grown at least 4-fold in 20th century—and given the rates of urbanization that have also grown in that time, from 14.5% in 1900 to 59.7% in 2005, this is almost certainly a massive movement of rural inhabitants into urban spaces.

All of this occurred only in the last century. Going further back, the internal histories of many of these cities are replete with cycles of growth and decline, and so it is quite difficult to be sure that an urban space is going to continue earlier linguistic behavior. Baghdad, once one of the largest cities in the world, is said to have had as few as 15,000 inhabitant in the 1650s (Palva 2009, p. 31). Even rural populations may have had recent depopulating and repopulating events—Behnstedt and Woidich (2018) suggest that upper Egypt was depopulated multiple times, and that in some Upper Egyptian dialects "one has to suppose that the immigration from the Hijāz lasted right up until the present era, and that some of the village dialects evolved only in recent times (p. 84)".

In the Levant, an area for which we have some of the oldest clear examples of Arabic language in the form of late Nabatean and Safaitic, many modern dialects and their features

certainly seem to be more recent, certainly more recent than the Islamic conquests. Lentin (2018, p. 175) quotes Ayalon as saying that near the end of the 15th century, the area between Latakia and modern Biredjik was Turkish, rather than Arabic-speaking (these form a line that passes northeast from Latakia, to approximately 50km north of modern Manbij). While certainly not entirely surprising, given that much of this area is Turkish speaking today, it pushes the development (or perhaps, “deployment”) of the Cilician dialects later than might be imagined.

In the fertile crescent and the Syrian-Iraqi desert, Procházka (2018) begins the history of the region in the pre-Islamic era, with the desert hinterlands said to already have contained Arabic-speaking tribes. His primary claim is that the late tenth century is the *terminus ante quem* for the features he describes as characteristic of relatively more sedentary dialects in this region, but ascribes the Bedouin features in others only to the era following the Mongol conquests in the 13th century. The Shāwi Bedouin dialects he considers an early stratum, but one he links only to the 11th century, while he also notes frequent movement even into the 20th century. The camel-breeding Shammar and ‘Anaza are said to have come only in the 19th century, the time he gives for similar migrations to the Cilician Plain, building on Lentin’s account above about the Arabicization of north-west Syria. It is also notable that folk accounts put the major migrations into Tillo, near Siirt, Turkey by Arabic speakers to ca. 1300 and 1600 in two waves (Procházka 2018, n. 5). While Procházka does note features which are attested early, and still present in the region, such as the shift of /*r/> [y], attested in Al-Jahiz (d. 869 CE), this is only a report of a feature, with some marginal spacial information (p. 270). No data exists to determine whether the current linguistic situation reflects continuous inhabitation by the same dialect group.

In the Gulf, Holes (2018b) states that only in the past century, beginning in the 1930s to the 1970s, “the Gulf dialects as a whole (with the partial exception of Oman) underwent a number of ‘reductional’ changes in their morphology (p. 134, emphasis original).” These changes include a loss of gender distinction in plural verbs, loss of the internal passive, loss of the dialectal tanwin, and the innovation or increased use of analytical genitive markers. All of these features are typical of the features used in dialectology for classification and historical reconstruction. If they can spread in a few decades, then we must be cautious about reconstructing even further back. Holes’ “tribal Arab” dialects are said to be primarily due to 18th century migrations, with an earlier stratum of unknown chronology (pp. 134–35). Though his evidence for the antiquity of his B strand of dialects as an early layer is quite compelling, he rests some of his argument on the isolated nature of Oman, noting that significant changes have occurred since the ascension of Sultan Qaboos in the 1970s, but that many speakers he worked with in Oman showed extremely limited mobility. However, one wonders whether earlier periods in Oman’s history, such as the Ya’rubid dynasty (1624–1742) and Omani Empire periods (1710–1783), when Oman was major regional power, might not have had a similar impact on the language to the present growth of the Omani state.¹³

While all of the authors in this volume, and in Arabic dialectology more generally, are clearly aware of these later changes, it is still difficult for them to pull entirely away from the “big-bang” model. Behnstedt and Woidich (2018) search hard for the “first layer” in Egypt, drawing on early accounts of the tribal affiliations of the migrants, even as they acknowledge later strands of migration. Holes (2018b, p. 133) attributes the *-imm-* infix’s distribution in the Arabian Peninsula to the era of the early Islamic conquests, though he notes that its movement into Egypt, Sudan and the West Sudanic area was probably later. Owens (2018), as noted previously, simply assumes a big-bang model which leads quite directly to suspect historical reconstruction. He argues that the *b-* prefix in modern dialects come from a single source simply as a result of his historical model, “the spread of *b-* described did occur in some regions very early, and indeed has existed in the forms which will be reconstructed since at least the earliest Islamic period, if not in the pre-Islamic era (p. 212),” which leads him to treat the Yemeni *b-* prefixes, which are quite transparently derived from **bayna(ma)* (Behnstedt 2016, p. 213) as equivalent to other *b-* prefixes which

are likely from other sources (Owens himself reconstructs them as being from *yabgā* > *yaba*). Procházka also tends to use the construct of “pre-diasporic Arabic” (p. 267) and focuses on the early tribal conquests and settlement, even as he acknowledges the later population movements.

The other side of the big bang equation also seems lacking at times—there simply is not always a compelling record of Arabicization in the earliest periods. The Arabicization of Egypt probably did not begin in earnest until the Fatimid era, with complaints about the shift from Coptic to Arabic reported into the eleventh century CE (Magidow 2013, pp. 220–22; Papaconstantinou 2012). Aramaic took time to be supplanted even in the Levant and Iraq, with neo-Aramaic dialects surviving to this day in Syria and Iraq.

The area where early Arabicization is most unlikely is North Africa, where vast portions of the area remain either un-Arabicized or show significant bilingualism between Berber languages and Arabic, even after 13 centuries of Arabic presence in the region. The chronology outlined in Aguadé (2018) does not seem likely to have produced significant Arabic penetration prior to the tenth century (the “pre-Hilalian” period). He repeats reports that Arabs formed only a part of the population in many Tunisian towns into the twelfth century CE, while Qayrawan in his telling only developed into a major regional center by the ninth century. Even if it was, as claimed with very little evidence “the origin of the spread of all pre-Hilali Maghrebi dialects” the process that would have resulted in significant Arabicization would have taken centuries. Even in the traditional French chronologies, Tunisia itself, a mostly flat and accessible area immediately surrounding Qayrawan, only is completely Arabicized by the 15th century, suggesting that the spread of Arabic must have been quite slow (Aguadé 2018, p. 42). Fes is said to have been “surrounded by Arab tribes” in the twelfth century, while the supposedly Arab settlements of Başra and Nakūr disappeared by the 11th century, though it seems likely Berber was still the dominant language. There is little in this history to suggest a strong early Arabicizing trend in North Africa that would allow us to clearly attribute the supposedly “Pre-Hilalian” layer to the earliest era of settlement. We do have evidence that a dialect similar to modern Moroccan Arabic existed by the twelfth century, but it still shows some differences, such as the use of *mtaːʕ* ‘of’ instead of *dyāl* now more common in Morocco (Vicente 2012).

Indeed, there certainly seem to have existed an even earlier layer of dialects than the Pre-Hilalian ones. For example, in Ajwila Berber in Libya, the words for ‘Friday’ and ‘heaven’, contain a reflex of the (j) variable that must be originally a palatal-velar [j] or [dʒ], which goes against the “pre-Hilalian” [ʃ] and Egyptian (and likely Proto-Arabic) [ʒ] (van Putten and Benkato 2017). Borrowed words with the feminine ending have *-at*, not the current *-a* used outside of construct state, and this is the case in Berber dialects across North Africa (Kossmann 2013, pp. 209–14). All of this is suggestive of a very minor early level of Arabicization, that was almost certainly swept away by later layers, many of which probably came into the area well after the early Islamic conquest period.

The big bang view would perhaps presume that these later migrations were simply additive, building on an already-established dialect geography. However, the poor level of initial penetration of Arabic into the conquered areas, the high likelihood of extreme depopulating events occurring between those initial settlements and the present, and the relatively rapid pace of linguistic change all suggest that the big bang model of linguistic history is far too simplistic, and that the convenience of the model is not something that we as a field are lucky enough to enjoy.

3.3. Effectiveness of the Big Bang Approach as a Linguistic Model

The big bang approach is also problematic simply because it does not appear to be a strongly predictive linguistic model. As Behnstedt and Woidich (2013, sec. 15.5.2.2) succinctly note, “Neither of these two approaches [Jastrow 2002’s zones, Owens 2006 ‘pre-diasporic Arabic’] is convincing for linguistic subgrouping, because they cannot be related to linguistic variables which would justify them.” This is in part because of the difficulty, noted in Section 2, of linking not just individual features, but features clustered with one another within a dialect across nearly two millennia of history. Attempts to do so starting

in the mid-twentieth century were largely unsuccessful, failing to identify clear features of the earliest layer of conquest (Blau 1977; Cohen 1962; Ferguson 1959; Miller 1986; Versteegh 1984), while both Magidow (2013) and Owens (2006) should both be re-evaluated in light of the arguments presented here.

There are certainly features shared by all modern dialects that contrast with earlier layers of Arabic—the development of vowels in hollow verbs, for example, suffuses all modern Arabic dialects, even though there was a clear historical memory of the earlier situation, where the glides in hollow verbs were retained, also attested in the pre-Islamic Safaitic inscriptions (van Putten 2017a). Similarly, virtually all modern dialects now have *-aya > /a:/ for the *alif maqsūra*, though the /e:/ reflex is still attested in Classical and Quranic Arabic (van Putten 2017a). However, the fact that these features date to around the time of the Islamic conquests does not tell us that these features directly hail from that era, since any later migrations almost certainly would also have brought these features to those areas.

The little bang theory makes more specific predictions than the big-bang theory: there should be a clear division between the dialects that settled North Africa during the “pre-Hilalian” era and those that came later. Of course, even if this prediction was proven true, it does not prove the specific historical claims of the model—it simply proves that there have been multiple waves of migration, a situation that exists in basically all Arabic-speaking regions. However, the linguistic evidence still does not support a simple binary bifurcation of the dialects in this region.

Aguadé’s (2018) article in the Holes volume has done a monumental job of listing all the claimed isoglosses between pre-Hilalian and Hilalian dialects, which previously were largely scattered across dozens of publications. This allows us, however, to note how contradictory many of these isoglosses are, often crossing the supposed boundary between the two layers. For example, the phonemes /b/ and /m/ replace each other in both types of dialects; interdental sibilants are generally merged with dentals in pre-Hilalian dialects, but many exceptions exist, even in dialects like Cherchell that are traditionally considered key examples of pre-Hilalian dialects (p. 44). Short vowels are lost in open syllables in both types of dialects in Morocco (p. 47), though this is an urban feature further east. Monophthongization of historical diphthongs similar cross the boundary (p. 48). Unconditioned *imāla* cuts across all Maghrebi dialects (p. 49). The classic *niktub-niktubu* isoglosses characterizes all North African dialects, while the loss (or retention) of gender distinctions has occurred in both Bedouin and sedentary dialects (p. 55). While there are indeed some remaining isoglosses which do distinguish these groups, for example the realization of (Q) and greater use of analytic genitives in sedentary dialects, it remains quite possible that these are later than the alleged migrations (see also below in this section).

Research outside of the traditionally highlighted isoglosses again supports a sub-grouping of North African dialects which cuts across the divide. Magidow (forthcoming) analyzes the personal pronouns, demonstratives, and interrogatives in over 80 Arabic dialects to find isoglosses which can be used to classify the entire Arabic-speaking region. In contrast to many classification schemes, where the historical model often drives the selection of isoglosses, the isoglosses in this study were derived directly from the linguistic data without reference to a historical model. Isoglosses in this study are only those in which a clear innovation has occurred, so retentive features (e.g., retention of the interdentals) that are so often mentioned in the dialectology literature are except for comparison with traditional classifications.

One striking result from this analysis is that North African dialects, regardless of where they fall on the Hilalian divide, tend to cluster quite strongly. The study identified 9 isoglosses that were much more strongly represented in North African dialects than in the mainstream of Arabic dialects, identified in Table 1, which shows the isoglosses, their prevalence among the dialect sample as a whole, their prevalence in North African dialects, and their prevalence in dialects which have voiceless and voiced realizations of the variable (Q), used to distinguish between putatively pre-Hilalian and Hilalian dialects.¹⁴

Table 1. Innovations found primarily in North African dialects, compared with other dialects and divided according to “Bedouin” or “sedentary” features.

Isogloss	% of All Dialects (n = 80)	% of North African ¹ (n = 22)	% of NA with Voiceless Q (n = 13)	% of NA with Voiced Q (n = 9)
3mp <i>hum(n)a</i>	51%	95%	100%	88%
1cs < \int a:na < * \int ana	33%	86%	93%	75%
(ha:)Øu pl. dem	24%	86%	86%	88%
2ms + 2fs merged suffix pronouns	23%	81%	100%	50%
2cp \int <i>intu:ma</i>	19%	68%	78%	50%
* \int a \int <i>ku:n</i> for ‘who’	19%	68%	78%	50%
- <i>ayya</i> suffix on some pronouns	11%	41%	35%	50%
<i>waqta: f</i> for ‘when’	15%	36%	29%	50%
2s \int <i>inti:na</i>	8%	22%	35%	0%

¹ Note that this sample group here includes Maltese as well as Andalusian Arabic as they are often included with North African dialects in the literature.

It is clear from Table 1 that, not only are North African dialects strongly linked by the isoglosses, but that these innovations cut across the Hilalian divide. Many of the features are common in both sets of dialects, even as they are rare elsewhere. There is a great deal more homogeneity, and a larger number of isoglosses which distinguish this group in contrast to the other dialect groupings found in the few other dialect groupings found in (Magidow forthcoming). The “Peninsular Bedouin” group, found in the western Arabian peninsula, is only distinguished by 3–5 isoglosses, while the “Sedentary Levantine” group has 8 isoglosses.

This is not to say that North African dialects do not show some differentiation along the lines of the supposed Hilalian divide, but this tends to be the exception in the sample and data here rather than the rule. The \int *inti:na* innovation appears to characterize only dialects which fit the pre-Hilalian mold—none have voiced Q, all have merged interdentals. These 5 dialects (Anjra, Morocco; Larache, Morocco; Fez, Morocco; Tlemcen, Algeria; Djidjelli, Algeria)¹⁵ share almost all of the major North African isoglosses, though some lack -*ayya* suffixes, \int a \int *ku:n* or *waqta: f*. They also share another isogloss, a ‘when’ form based on **fi: \int ay waqt*.

The ‘where’ interrogative *fayn* is never found in dialects with voiced Q, and almost all dialects (75%) with that form have merged interdentals. From the other side, only voiced Q dialects have -*a(h)* in the 3ms suffix pronouns. Interrogatives for ‘when’ derived from \int *ayy mata:* are primarily found in the voiced Q group, but this is also found in two putatively pre-Hilalian dialects, that is Marrakesh and the Jewish dialect of Tripoli, Libya.

Figure 1 is a heatmap of the number of these features, showing all dialects with two or more features from Table 1. The size of the circles corresponds to the number of isoglosses from the table found in each dialect. Overlaid on this are symbols for whether the dialects have merged the interdentals with the dental consonants (taken typically as an indication of ‘sedentary’ dialects) or if they have a voiced realization of (Q). What is evident from this figure is that these features are found in both sedentary and Bedouin dialects. The number of features appears to form an east–west cline, rather than a division between Bedouin and sedentary dialects. As one moves west, one tends to find more of these features, while further east there are fewer—to the point where Benghazi’s dialect shows none of these isoglosses at all. There are of course some dialects outside of North Africa which have some of these isoglosses as well, though not in the areas of Arabia often held to be the source of the North African dialects (Magidow 2013, p. 236 shows the presumed homeland of the Banū Hilāl and Sulaym as western Arabia).

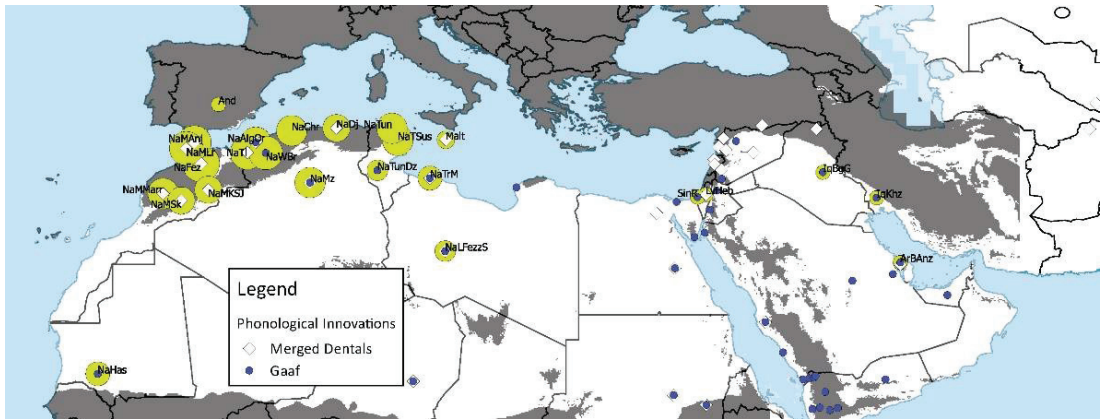


Figure 1. Map of dialects which show “North African” features, with dialects having at least 2 such features with a yellow circle sized proportionate to the number of innovations found in each dialect. Overlaid on this are markers of either merged interdental or voiced (Q) reflexes.

Evidence of an east–west cline, rather than a sedentary-Bedouin distinction is found elsewhere in [Aguadé \(2018\)](#). The affrication of */*t/ > /tʃ/* is only found in Morocco and Algeria, not further east (p. 44) The indefinite articles *wa:hid* and *ʃi:* are used only in Morocco and Western Algeria, while *fard* is common further east in Tunisia and Libya (p. 50). The supposedly ancient genitive particles derived from **mata:ʃ* are most widespread from North-Eastern Morocco into Libya, while the genitive particles similar to *diyya:l* are concentrated further west.

The Hilalian “little bang” narrative does not provide a strong model here for understanding the data. Instead the data shows a remarkable unity between the North African dialects from both pre- and post-Hilalian dialects, while the many contradictions in the traditional isoglosses found for North African dialects strongly undermine the narrative. If the Hilalian narrative was correct, we would expect a clear bifurcation on an essentially north–south axis, between sedentary Hilalian dialects on the coasts and rural/Bedouin dialects further south. The fact that many of the features work on an east–west cline strongly suggests that North Africa is little different from any other region of the Arabic-speaking world, with a gradual process of migration into the region. Dialects coming from further east, which presumably lacked many of the key North African features (but not all, given the eastern dialects shown with yellow in Figure 1) would acquire those features. Distance to the west should also act as a reasonable proxy for time spent in the region, as it takes time to migrate and settle, which would explain the cline going from fewer features to greater the further west the dialects are found.¹⁶

This section provides another piece of the explanation for Holes’ observation about the similarity of Gulf and Algerian Bedouin dialects. The existing big-bang and little-bang models cloud our view of these dialects, forcing us to assume that they are separated by nearly a thousand years of history, with the Algerian Bedouin dialects being the direct ancestors of the Hilalian dialects of the 11th to 14th centuries. Instead, it seems quite likely that the movement of those dialects into that area were much later, which would explain the easy mutual intelligibility he experienced. Moving away from that historical model allows us to instead apply Occam’s razor and propose that many migrations proceeded into (and out of) North Africa even after the supposed Hilalian period.

4. Conservative Dialects and Early Origins

There is another common idea based on apparent linguistic conservatism that combines both the “conservativeness as archaic” and the “big-bang” model. The idea is that

dialects which are more conservative are seen as likely candidates for genealogical ancestors of less conservative dialects. This was, of course, the basic premise behind the long-held view that Classical Arabic is the direct ancestor of the modern dialects, which has largely fallen out of fashion as evidence has begun to more clearly show that Old Arabic deviated in significant ways from Classical Arabic (Al-Jallad 2017).

However, there still tends to be a belief that the modern Yemeni dialects reflect, due to their archaism, a kind of ancestor to most modern dialects. This cannot be entirely separated from the big-bang hypothesis, nor can it be separated from the idea that conservative dialects have genetic posteriority. The Yemeni dialects are highly archaic in a number of ways, from phonology to vocabulary, and many of the features in Yemeni dialects (as a whole perhaps more than individually) appear quite similar to canonical Classical Arabic. For example, the interrogatives are primarily of the *mā* variety, rather than the forms derived from **ayy šay* which are common in both later Classical/Middle Arabic and most dialects.

This apparent linguistic conservatism is coupled with a strong historical tradition that holds that Yemen, and Yemeni tribes, were the origins of many of the Islamic armies. In combination with the big-bang approach, this means that these Yemeni dialects are often held to be the immediate proximate ancestors of modern Arabic dialects. For example, Map 3.2 in Behnstedt and Woidich (2018) depicts “spring pastures of Yemeni tribes in immediate post-conquest Egypt,” while references to that Yemeni origin and influence abound in their article. Elsewhere we see for example Yemeni origins posited as the ultimate ancestors of the Ma’qil tribes that are seen as ancestors of the Hassaniya dialect (Taine-Cheikh 2006, p. 301), a narrative repeated by Watson (Watson 2018, n. 9). Outside of this volume, one can witness the attempt to link Yemeni to Andalusī Arabic (Corriente 2014), though that argument has been criticized (van Putten 2017b).

Beyond the historical reasons for seeing this link, there is also a deeper misinterpretation, again about the significance of “old” features. This idea revolves around the idea that archaic features, or dialects with many retentions versus innovations, are older, and that older dialects must necessarily be ancestors to newer dialects. However, the reality is that the Yemeni dialects (and conservative dialects more generally) are striking precisely in how much they *differ* from other dialects, however conservative that might be. Though this means they could be an ultimate ancestor (or related to an ultimate ancestor) of other modern dialects, there almost certainly exist more proximate ancestors.

A wide-scale comparison of existing Arabic dialects largely confirms this. Magidow, (in preparation) identified 55 isoglosses that represent innovations with regard to Proto-Arabic, and then identified isoglosses present in at least 60% of all sampled dialects. These are shown in Table 2. The geographic distribution of these groups is shown in Figure 2.

Table 2. Core innovations found in the majority of modern Arabic dialects (*n* = 80).

Isogloss	Dialects with Innovation 2
Demonstratives from <i>ḏa</i> : <i>ḏi</i> : not <i>ḏa</i> : <i>ta</i> :	95%
Interrogatives from <i>*ʔayy f ayʔ</i>	92.5%
2fs suffix pronoun <i>*-ki > -ik</i>	72.5%
2nd person pronoun stem <i>*ʔan- > ʔin-</i>	70.0%
3 ms suffix pronoun <i>*-hu > *-u</i>	70%
1 cp pronoun <i>niḥmu > ʔVḥn-</i>	67.5%
“where” <i>ʔayna > wayn</i>	63.5%

What is notable about the distribution of these features is that, as would be expected in the literature, the Yemeni dialects are the only dialects to fall outside of the 90% core. That is to say, the vast majority of Arabic dialects have participated in those two major innovations, while it is primarily the Northern Tihama dialects (Behnstedt 2016, p. 191) that have failed to partake in these innovations. Though not necessarily in the same group of dialects per Behnstedt’s classification, there are further archaisms found only in Yemen that

are absent from most modern dialects. Virtually every modern Arabic dialect has, through analogy with the 1cp suffix *-na*: (both pronominal and verbal) innovated a form of the 1cp independent pronoun from **niħnu* to *nihna* or similar. The primary exceptions are in Yemen (points 153 and 154, just outside of Ta'izz, not far from the point 145 included in Figure 2), while Reinhardt (1894, pp. 21–22) transcribes *honu*: and *nahnu* for some Omani dialects.

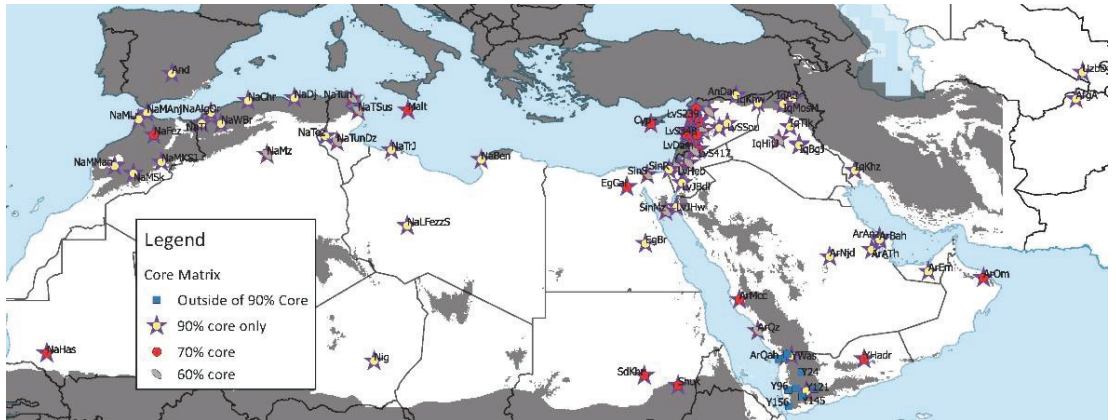


Figure 2. Map of “core” dialects, with dialects not sharing the 90% core features shown as squares.

The implication of this data, when freed from the “archaic as ancestor” narrative is that it is quite unlikely that these Yemeni dialects are the most recent node on a genealogical tree from which all other dialects developed. It is much more likely that most modern dialects derive from a dialect which innovated these core features, while the Yemeni dialects simply did not participate in those changes. The conservatism of these Yemeni dialects and their linguistic features provide a convenient window into the linguistic past of Arabic, however, they do not imply a linguistic ancestry in a historical sense.

Given how widely distributed across space (and time, given their presence in Andalusian Arabic) the dialects with 90% core features are, it would seem reasonable to assume that a single ancestor developed those features before becoming more widely distributed. This dialect, of course, would be unlikely to have been in Yemen, given that its innovations have not fully suffused the area, in contrast to the rest of the Arabic-speaking world. In the “big-bang” viewpoint, this would be the “pre-diasporic” Arabic, the variety that was spoken by the conquering Arabic armies in the early Islamic conquests.

However, as detailed previously, we simply cannot be entirely sure of the how ancient or new any dialectal features are with respect to their distribution in space. There has clearly been a huge amount of dialect movement and change, and many waves of diffusion which have brought specific linguistic features across the Arabic-speaking world. The example of the verb *ʃa:f* ‘he saw’ are illustrative. The verb *ʃa:f* has incredibly wide distribution across the Arab world, found in virtually every Arabic-speaking region today. While this diffusion is often believed to be quite early (Ferguson 1959), many dialects still have other verbs or biforms which appear to be diminishing only in recent times (Behnstedt and Woidich 2011, pp. 330–37). It is hard to determine the earliest this verb came into use—we find an example of the causative *ywʃwfwk* ‘they will show you’ in a papyrus datable as as early as ca. 1000–1100 CE.¹⁷ Though attested relatively early, the modern wide-spread distribution of this form often appears to be quite recent. Cowan (1966) argues that the diffusion of *ʃa:f* must have taken place between the twelfth and sixteenth centuries, based on the places where it is not found (Malta, 15th century Andalusia, and Cyprus). However, in much of North Africa this lexeme’s uptake is clearly quite recent. It is generally not found in Judeo-Arabic varieties in North Africa, while Aguadé (2018, p. 57) notes that

in Djidjelli as of the 1950s, *ʃa:f* was a recent loanword, while in Anjra in the 2000s there was a generational divide between users of *ra* and *ʃa:f*. Blau (1977, p. 200) notes a similar situation in the Tunisian dialect of Marazig, where *ʃa:f* was only widespread among men. Behnstedt and Woidich (2011, p. 333) note many places, in both eastern and western dialects where there are bi-forms, often with *ʃa:f* clearly encroaching on older local forms (mostly **ra*: reflexes, but see their maps for many other previously common verbs with this meaning).

The implication here is that the diffusion of the “core” isoglosses could be either genetic or areal—that is, the diffusion of these features could be an early phenomenon, in a common ancestor of modern dialects, or like *ʃa:f*, they could owe their current distribution to later waves of diffusion. Particularly in the latter case, it would not be surprising that this diffusion would not have made it to Yemen, a peripheral region with difficult terrain, often not included in a meaningful way in the major Arabic-speaking empires. However, in either perspective, while Yemen may have preserved earlier states of the language, Yemeni dialects are at best a great-grandparent or great-aunt to the bulk of modern dialects, but hardly a direct parent.

5. From History to Heuristic: An Alternative Approach

There are several dangers when it comes to using a historical approach for analyzing the macro-history of Arabic dialects. The first is simply that any attempts to reconstruct the history of the Arabic-speaking world are wildly ambitious. The Arabic-speaking world is vast, both in its present reach and historical extent. With Arabic spoken from Mauritania to Afghanistan to Zanzibar, and Arabic inscriptions already attested early in the first millennium, centuries before Islam, any attempt to even scratch the surface of this history will necessarily be extremely superficial. Most historians would be hesitant to even attempt a complete history of the population movements in a region, let alone the entire Arabic-speaking world, especially in the current environment where *longue durée* and macrohistory has become somewhat rare, or is treated at a more popular than academic level. By way of example, Hugh Kennedy’s *The Great Arab Conquests* runs to nearly 500 pages, but covers less than two centuries. Expecting a dialectologist not trained in history to be able to make an original contribution and synthesis of the historical literature is demanding a great deal.

The second is that established histories can become quite difficult to rethink and question once they have become integrated into the dialectological literature. We have seen this already in the big-bang approaches to Arabic, but the little-bang approach of North African history bears additional scrutiny. The historical model established by Marçais (1938) has dominated the dialectology of North Africa, and researchers remain strongly committed to it. Benkato (2019) has documented this extensively (pp. 16–18 especially), but one is still amazed to find statements in Aguadé (2018) such as “it is this role as a junction that, according to William Marçais, made Qayrawan the origin of the spread of all pre-Hilalian Maghrebi dialects, and *there is no reason not to accept this assumption* [emphasis added] (p. 39)” or “the question of whether this description [of Banu Hilal causing destruction] (which was widely embraced by the French colonial ideology in the nineteenth and twentieth centuries) actually reflects historical facts *need not be discussed here* [emphasis added] (p. 42).” The dominant narrative is unquestioned, even as it makes it difficult to account for the many contradictions in the data covered in that chapter, as discussed previously.

Finally, it is not always clear that linguists are always able to do high quality historical work, again because we are largely not trained to do so. A common concept in Arabic dialectology is that of *sprachinseln*, dialect areas which, by virtue of being cut-off from the mainstream of the Arabic language, can allow us to date subsequent changes. However, our historical models for these *sprachinseln* are highly simplistic. One of the most commonly discussed situations is that of Malta, for which Holes (2018a, p. 18) gives a typical interpretation: “a good example [of a *sprachinsel*] is Malta, where a variant of Siculo-Arabic was spoken until the end of the eleventh century, after which all contact with Arabic-speaking communities ceased.” This is echoed by Aguadé (2018, p. 34) where he

states that “Maltese is an important source as *terminus comparationis* since the island of Malta was conquered by the Normans [. . .] in the year 1090,” the value of which is that “Maltese represents an archaic pre-Hilali dialect which evolved uncontaminated by later Hilali interferences.”

However, even a slightly deeper look into the history of Malta suggests that reports of this “cutting off” are greatly exaggerated. The conquest of Malta did indeed bring it under the rule of the Kingdom of Sicily in 1090—but that Kingdom also included Northern Tunisia, and further conquest included Tripoli from 1146 to 1158, and coastal Tunisia from 1134–1148. The Normans did not appear to meaningfully occupy Malta at that time. Rather, Malta was a tributary until 1127 when it was conquered “to use as a transit point for trade,” clearly including trade with Tunisia (Joffé 1990, p. 68). Luttrell (1975, pp. 31–32) reports a Pisan captain, apparently to combat piracy, seizing a Tunisian ship at Malta with the goods aboard it, and throwing the crew into the sea. The island of Pantelleria, 200 km NW of Malta, was even included in an arrangement by which half their tribute went to Tunis (Luttrell 1975). This trade with North Africa clearly continued, where under the Aragonese crown starting from 1283, “Malta’s real usefulness was not as a market or a source of raw materials but as an entrepot and safe harbor on the routes to Beirut and Alexandria, and above all to Tunis and other African ports (Luttrell 1965, p. 6)” where slaves, wood, and cotton all appear to have been traded through Malta. Trade contacts almost certainly would have also resulted in continuing linguistic contact.

Cyprus is held to similarly have been cut off, with Borg (2006, pp. 536–37) stating “the sociocultural parallels between Cypriot Arabic and Maltese is particularly close . . . since in both cases, we are dealing with an Arabic vernacular surviving in complete isolation” referring presumably to Crusader conquest of the island beginning in the twelfth century.¹⁸ This vision of a “cut-off” in Cypriot Arabic leads Tsiapera (1969, p. 11) to speak of a dialect “isolated for some six centuries from other Arabic speakers.” However, here again, there is no clear cut-off, with both migration and trade clearly continuing through both the crusader and Ottoman eras. Movements of Christian refugees from the Levant continued at least through the 13th centuries (Borg 2004, p. 8), with some authors suggesting migrations even in into the Ottoman era in the 16th century (Hourani 1998). As in Malta, trade with the Levantine coast must have continued in both the crusader and Ottoman eras (Borg 2004, p. 10).

The problem here is that linguists, not trained as historians, are making two basic mistakes in their historical research. The first is simply not diving deep enough in their research on these dialects—there is notably almost never a reference cited for the “cutting off” of Malta (none is cited in Holes or Aguadé’s assertions above), and little further research appears to have been carried out in most such accounts. This is somewhat reasonable, as these authors are dialectologists, not historians, and cannot be expected to be deeply immersed in the historical literature. The second is the overly facile equation of a change in political rule or religious affiliation with a wholesale change in a group’s personal and trade relationships. This again seems to be an issue of not having been trained in the surprising mobility and importance of trade that was common in the pre-modern world. Thus, while dialectologists are not ‘to blame’ per se for perpetuating historically inaccurate models, these models can significantly impact how we interpret the data from our dialectological research.

One solution to this problem is to work on more focused, microhistorical work of particular regions. Palva (2009) is a paradigmatic example of this, weaving highly detailed dialectology work with a more sophisticated view of history than we sometimes encounter. He rightly questions the tradition that 1258 marks the decline of Baghdad (see for an example of that idea Fischer (2006) which uses 1258 as the dividing line for “post-Classical Arabic”), and instead notes that the decline of the city has been documented as beginning much earlier. He also correctly identifies that the city may have had cycles of depopulation and repopulation. This kind of highly focused work—in this case, on a single city—allows for a much deeper level of research into the history of the area. Another suggestion would

be for linguists to reach across departmental and disciplinary lines and work with historians specializing in the area to bring a new perspective to their work, and certainly linguistics and dialectologists have their own contributions to make to historical research.

The microhistorical approach, however, does not work as well when dealing with large areas or finding general trends. Here, the danger of missing large chunks of history will always rear its head, and it can be challenging to dismantle the existing historical narratives dominant in the field. For this reason, I suggest instead a general heuristic that draws on general principles, rather than specific historical narratives, to suggest the kinds of linguistic movements that we should expect to see over time and which can be applied as a basic test of whether a historical narrative is plausible.

This heuristic is based on the observation that the movement of linguistic features across the landscape divides into primarily two types. The first is the diffusion of features without a major change in the distribution of populations within a space—that is, the speakers remain in situ, while the linguistic feature itself moves across the landscape. I refer to this simply as “diffusion”. The second is the movement of populations, such that a group moves into an area, bringing the linguistic features of that group with it. I refer to this as “migration.” The speakers are not changing their linguistic behavior, but the linguistic geography of the area changes by virtue of their movement.

Diffusion is by far the most common way a linguistic form moves across the landscape in areas with high population density. This diffusion is generally not simply geographical, diffusing outward to the nearest geographical point, but instead is often hierarchical, with features moving between areas with similar population densities even if they are physically distant, and only later diffusing to areas that are geographically adjacent but with lower population density (Britain 2008). Major movements by sedentary groups are rare, the result of major economic or political changes (e.g., industrialization, urbanization, warfare) and as discussed earlier require quite large changes in the total population for a group to supplant the linguistic behavior of a high population group via migration (Miller 2004; Palva 2009). Less populated areas certainly experience diffusion as well, though being at the “bottom” of hierarchical diffusion they tend to diffuse features from the “bottom up,” that is features tend to diffuse between areas of similar population density (Wikle and Bailey 1996).

In contrast, for areas less conducive to intensive settlement, migration is much more common as a driver of change in dialect geography. The lower carrying capacity of the land means that population densities are lower to begin with, often requiring nomadism to maximize resources, ensuring frequent movement. Therefore, on a given point of land in such areas, the inhabitants are both more likely to move of their own accord, and if another group moves into that area, are more likely to be overwhelmed by the power of numbers. The basic strategy of nomadism means that it is easy for a group speaking a particular linguistic variety to move out of a given territory or into new territory, perhaps due to a single or several seasons of bad weather. Nomadic groups are well prepared for mobility, physically (in terms of their possessions) and culturally (in terms of having the necessary knowledge to survive). Sedentary populations are less likely to move. Even when they do, they have larger populations and thus are more likely to leave behind enough members of the group to maintain the presence of their linguistic variety in their original location. Depopulating a city requires a catastrophe, while for a nomadic group, leaving a given area is a routine seasonal event. For a nomadic group to seize the territory of another nomadic group requires a relatively small migration of people and could radically change the linguistic behavior of an area from a dialectologist’s perspective. On the other hand, nomadic groups cannot generally move into densely populated sedentary areas and impose their linguistic behavior on that area. The nomadic group would need to first win a military campaign against a numerically superior group, and then to maintain and impose their language on a still numerically superior group.¹⁹

Therefore, in this model, densely populated areas are treated as a barrier to migration as a means of language change. For a migrating group to have a linguistic impact, absent a

depopulating catastrophe, it must migrate within marginal lands that support similarly small groups. Of course, there is a limit to the kind of terrain that can support meaningful numbers of people. Life in the Sahara and Empty Quarter require such specialized skills for survival that it might be hard to supplant the small number of speakers in those regions easily. This means that there are narrow corridors between the areas where population density is too high for a migrating group to have an impact linguistically (and where there would be potential resource competition with settled people who were actively using more fertile areas), and where it simply is too difficult to inhabit for a newly arrived group. This model is especially important for the MENA region since those corridors are a more common component of the landscape as vast areas of the Arab world are poorly suited for intensive settlement. The amount of agricultural land in 1961 in the Arab world was approximately 25%, compared to 56% in the Euro area.²⁰ Nomadic groups have played a much more significant role in the development of Arabic than in the modern development of Europe languages, for example, and so our heuristics for the development of Arabic must take this into account.

This model therefore suggests that there are two main corridors of movement of linguistic features over space and time in the Arabic-speaking world. For sedentary populations, movement of linguistic material will likely happen between densely populated areas by way of diffusion of linguistic features without permanent population movement. For nomadic populations, movement would occur through marginally inhabited spaces, and it would be the physical movement of populations that result in a particular distribution of linguistic features in space. Densely populated areas would act as a barrier to the diffusion of linguistic features associated with nomadic speakers of the language. These corridors allow us to anticipate particular kinds of linguistic change over space and time without a need for reference to historical events.

Population density is fundamental to this model, but since we have no access to historical population data, we need a proxy variable. Rainfall provides a good proxy, since agricultural production is roughly proportional to total rainfall, while nomadic pastoralism in the Middle East exploits low-rainfall areas through seasonal migration and use of hearty livestock such as camels. To establish that this proxy variable works well, Magidow (forthcoming), correlated data about the realization of the Q variable in a sample of 88 geo-located dialects against average (modern) rainfall totals. Comparing whether dialects have the voiced realization of (Q) to the average amount of rainfall produces a correlation of 0.50 ($p < 0.001$), explaining nearly 25% of the variation between dialects, a reasonable result for a very rough and simplistic comparison.²¹

Another variable to consider is elevation. Different types of nomadism require different subsistence strategies. One major division is between low-land subsistence (based on movement across large distances) versus elevation subsistence (based on movement up and down in elevation) (Barfield 1993; Barth 1961; Donner 1989). It is notable that in many areas where the Arabs conquered, the 1000 m elevation line represents the limits of Arabicization, from Andalusia to North Africa to the Iraq-Iran border, though Yemen constitutes something of an exception (Donner 1989; Kennedy 2007, pp. 435, 438; Magidow 2013). Figure 3 therefore combines both rainfall and elevation in an attempt to illustrate the migration corridors for linguistic features. The black areas represent either those areas with too much rain (or a river systems likely to support agriculture) or which are at too high an elevation (1000+ m).²² The light grey areas on the map have extremely low rainfall of 50mm/year or less, which are very difficult to inhabit even for nomadic settlement. The white areas are therefore the “Bedouin corridors.” These are the areas where we would expect movement of Bedouin groups, while we expect less movement, and slower movement, elsewhere.²³

Overlaid on the map in Figure 3 are realizations of the proxy variable most strongly associated with Bedouin dialects, use of a voiced reflex of the Q variable. As expected, voiced Q reflexes are primarily found within the Bedouin corridors. The most common exception is in Yemen, which is a geographical and linguistic area noted for being dialect-

tologically unusual. This suggests that in general, the notion of the Bedouin corridor is a good proxy for how nomads help diffuse linguistic features, without a need to reference a particular model of the history of Arabic.

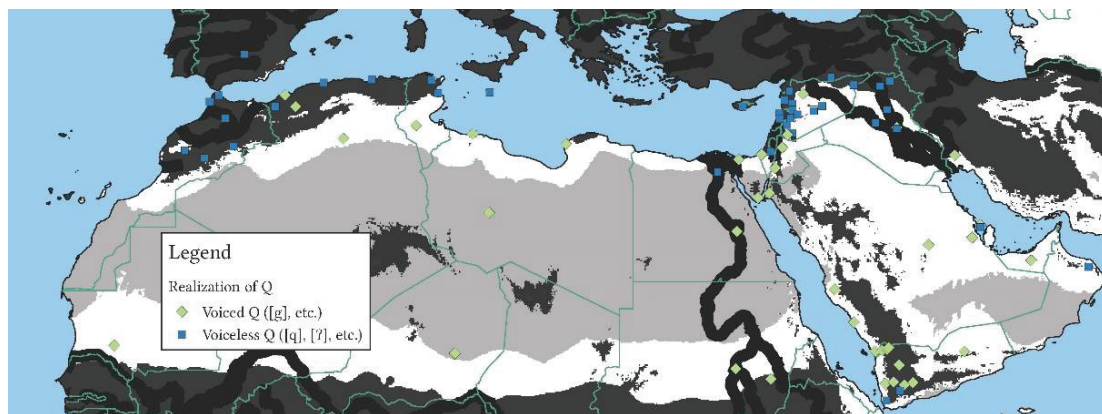


Figure 3. Map of Bedouin corridors (white) contrasted with barriers to Bedouin movement due to rainfall or elevation (black), or extremely low rainfall (grey), overlaid with proxy variables for sedentary vs. Bedouin dialects.

This model also makes meaningful predictions. Among other things, it removes the need to treat the Bedouin-sedentary difference as an essentialized difference with little explanation. Instead, this distinction falls out naturally from the basic idea that there are two different manners in which linguistic features move. Should a dialect spoken by nomads settle into a populated area, it would stop being subject to migration as a driver of linguistic change, and become more influenced by diffusion from more densely populated areas. The similarities between sedentary dialects are therefore due primarily to linguistic diffusion between them, rather than due to a single process of koineization (or even perhaps sharing an early ancestor.) The similarities between Bedouin dialects are similarly a result of their contact with one another, and likely due to rapid movements which erase diversity rather quickly and present an illusion of homogeneity across broad spaces.

This model also informs our historical understanding. In North Africa, we expect to find significant nomadic movement along the coast as far as southern Tunisia. After that, we would expect most of the areas of the coast to be relatively more difficult for nomadic groups to penetrate, and that this would be an area of sedentary in situ linguistic diffusion.²⁴ This replicates relatively well the “pre-Hilalian” vs. “Hilalian” distinction, but without a need to rely on the flawed historical model. We expect the Sinai to be a major land-bridge, with frequent population movements, and that the dialects in that area probably do not represent extremely long-term settlement.²⁵ We expect the movement along the western coast of the Arabian Peninsula to be relatively easy, while the Hijaz mountains would likely divide the dialects to the west and east of that mountain range. We expect to find the desert between Syria and Iraq to act as an extension of the Arabian Peninsula in terms of dialect, but the crescent of sedentary areas from the Syrian coast through Anatolia and into Mesopotamia to form a barrier for further nomadic movement, and for linguistic features to diffuse through that sedentary corridor (Behnstedt 1990; Ingham 1982; Jastrow 1978, p. 78).

Another key prediction of the model is that the linguistic behavior of sparsely populated areas is unlikely to have significant time-depth in terms of features in that space over long periods of time. The chances that the linguistic behavior in the Bedouin corridors has remained the same over millennia is quite low. Indeed, as discussed in Section 3, even the sedentary areas may show less time-depth for their linguistic features than is

typically believed. This has serious implications for how dialectologists analyze the dialects of the Arabian Peninsula. In keeping with the ‘conservative features imply antiquity in place’ fallacy, we often find linguists assuming that Najdi Arabic represents an unbroken linguistic tradition in the region. Owens (2018, p. 211) claims that “Gulf and Najdi Arabic are spoken in the Arabian Peninsula and have been spoken in these areas since pre-Islamic times.” Al-Jallad (2009) similarly states that “the Bedouin dialects of the southern Najd of course never left the Arabian Peninsula.”²⁶ This of course seems highly unlikely. This is a region of very low population density, with well-documented migration out of the Arabian Peninsula into other areas. The vacuum caused by these migrations would almost certainly have changed the dialect landscape of the Peninsula itself. Ingham (1982, map 5) illustrates a very significant reshaping of the linguistic isoglosses in the Peninsula since the 17th century. It is simple to imagine that the thousand years between the Islamic conquests and the start of his map witnessed many similar disruptions to that dialectal map. Indeed, the main argument of Holes (2018b) is that the dialects of the Gulf reflect at least three major layers, one of which may date only as recently as the late eighteenth century. One wonders too about the Baharna layer that he uses as a key piece of evidence. It is in relatively rapid decline, and if traces of that dialect can virtually disappear within two or three centuries, how many layers of Arabic dialects may have been lost in the past millennia? This is not to strictly argue that there cannot be continuity here, but it is highly unlikely and we would need the kind of microhistory mentioned previously to prove continuity or a lack thereof.

This idea that the dialects of the Arabian Peninsula are somehow original to the area based on their archaism is not entirely dissimilar from the argument about Yemeni dialects discussed previously. However, the heuristic here would predict very different histories for the two regions. The highlands of Yemen are relatively inaccessible, and have historically had higher populations than central Arabia, receiving greater rainfall including some from the monsoon. Our model would expect that linguistic change in Yemen would be relatively more difficult than in the Najd. This is not to say that we expect Yemen dialects to be highly durable—it is notable for example that the least “non-core” Yemeni dialects discussed above are located on the Tihama, an area that should be a Bedouin corridor and should not have long-term durability. The geography of the region means that we should consider the possibility that these archaic dialects originated somewhere else, and only arrived relatively recently in Yemen. The model proposed here allows us to still derive observed differences between dialects while being able to explain them in a parsimonious manner based on general principles and using a more accurate analysis of conservative and innovative features.

6. Conclusions

This article began with an observation about dialect similarity that the dialects of the Arabian Gulf and those of southern Algeria, two areas thousands of kilometers apart, were virtually mutually intelligible, while the dialects only a few hundred kilometers to the north were not. This paper has argued that this fact is surprising primarily because of the long-standing narratives in the field of Arabic dialectology, and if one steps away from those narratives, it becomes evident that this situation is both explicable and expected. This article argued that there is a strong tendency to essentialize the idea of linguistic conservatism and apply it more broadly to the linguistic groups which have conservative features. In a similar vein, the field of Arabic linguistics tends to assume the primacy of the earliest layer of settlement during the Islamic conquest, and that modern dialect geography is directly derived from these early settlements. In contrast, this article argued that it is highly unlikely that the earliest settlements successfully Arabized the areas settled and even if they had, it is unlikely that they were able to resist 1400 years of historical changes. The article similarly argued that the long-held idea that Yemeni dialects are ultimate ancestors of most modern Arabic dialects ignores the evidence for the existence of more proximate ancestors from which the dialects descend. Finally, the paper argued that given the limitations of historical models, it may be helpful to apply a heuristic approach

based on sociolinguistics and geography to derive the existing observations found in Arabic dialectology.

This paper is meant to provide some suggestions to strengthen the work being carried out in this field and the following attempts to summarize these into concrete suggestions:

- Refrain from essentializing the concept of ‘archaism’—archaic features can be useful reflections of earlier stages of the language, but they do not necessarily imply that the people who use archaic features, or the totality of the dialect they speak, are necessarily resistant to change or free of innovations
- Be cautious of using social history within linguistic arguments—it can be difficult to be sure that the changes that occurred in one era were not simply erased in following eras
- Consider the possibility that the distribution of features in Arabic dialects may be due to relatively recent changes, rather than the earliest population movements
- Use a heuristic model to check the plausibility of a historical narrative, and to compare historical processes in areas with different geographical landscapes

Finally, there are some larger implications for the field in this paper which will need to be explored elsewhere. If the Arabic dialects are, by and large, new to the areas they currently inhabit, and the linguistic features characteristic of each dialect susceptible to adoption by different dialects as people, or linguistic features, move across the landscape, can Arabic dialects actually be analyzed from a traditional genealogical perspective? Is it meaningful to talk about a long-term persistence of a cluster of dialect features within the same social group? Or are the Arabic dialects reflective of *Wellentheorie* at its most fundamental—a set of human movements bringing features across a vast space, sharing them at times, and then bringing them elsewhere, or disappearing entirely? If a wave on a shoreline leaves behind a seashell, then the following wave moves that shell before depositing a new one, does it make sense to speak of any of those waves as the “origin” of the shore? Will our field be able to reconstruct the dialects that moved into North Africa, or Central Asia, or Andalusia, or the earliest dialects immediately following the Islamic conquests, or do we simply lack the fidelity and density of information needed to do so?

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

Notes

¹ The author wishes to thank Thomas Leddy-Cecere, Luca D’Anna, Phillip Stokes and the three anonymous reviewers for their valuable feedback on earlier drafts of this article. All remaining errors are the author’s responsibility.

² A note on conventions followed in this article: linguistic data original to the article is conveyed in IPA, with Arabic emphasis indicated using pharyngealization (e.g., /sʕ/). Proper names from Arabic unconventionalized in English Arabic are transcribed using standard Orientalist transliteration, which is also used in quotations that use such transcriptions. I consider the term “Hilalian” sufficiently conventionalized that it does not need to be rendered in transliteration.

³ I am deliberately ignoring the social here for the moment, though that is an important aspect of any linguistic analysis, see Magidow (2017) for social factors in relationship to historical dialectology.

⁴ It is important to be clear that features are simply linguistic features—they can only change and move through sociolinguistic or migratory processes, and they change as a result of normal linguistic processes. Features are at times anthropomorphized, so for example we find Procházka (2018, p. 273) stating that gender distinction in the plural is a “feature vulnerable to population movement.” This fails to explain the mechanism behind the loss of feature, or how movement rather than other factors cause it to change.

⁵ Here, and elsewhere, it may be more helpful to remember that almost all “dialects” mentioned in the Arabic dialectology literature are “doculects” in the sense of Cysouw and Good (2013), that is to say that our knowledge of these dialects is largely restricted to published sources, some of which are quite old, based on single speakers, or similar limitations.

- 6 Another aspect of this issue is that our focus on differentiating features, rather than the bulk of the linguistic features in a dialect, means it is often difficult to track the continuity of language that remains even when a small number of features change. A dialect might change its behavior with respect to several sociolinguistically relevant features, such as the realization of (Q) and the interdentals in order to match nearby dialects, but remain quite different in lexicon and grammar. When operating at a very high level of abstraction, we may miss some of these facts and misclassify the dialects involved.
- 7 It is beyond the scope of this paper to probe why this idea of purity or conservatism is so widespread. It features quite heavily in early grammarians' works and the larger narrative of urban life destroying the 'original' dialects. In that case, however, it's important to recognize that the grammarians were undertaking a very different kind of research, much closer to the modern practice of documenting endangered varieties than to performing historical dialectology. It could also be linked to practices in modern dialectology that have focused on non-urban rural males, again in an attempt to document possibly older varieties but which also contain a notion of purity and originalism.
- 8 It should be noted, however, that "innovative" or "retentive" vis-à-vis the proto-language does necessarily mean that a feature is new or old in a given dialect. "Retentive" features can be sociolinguistically important and be passed between dialects, which can be evidence of shared contact in the past, but it is in most cases too difficult to prove that this happened without a doubt. See further Magidow (2017), note 18.
- 9 Since many Bedouin dialects are or were spoken by speakers that are nomadic and tribal, the assumption seems to be that these groups would have had primarily dense connections within the relatively small camping groups in which they travel. However, there are several issues here. First, nomads are hardly divorced from the world around them, being dependent on settled groups for provisions and trade, and often groups that follow transhumanist lifestyles will alternate between nomadism and more settled lifestyles over the year, certainly bringing them into contact with other people and groups in a way that would form weaker network ties (Magidow 2013, chap. 2). Second, the concept of tribe so frequently used as an object of analysis in Arabic linguistics has rarely been submitted to rigorous investigation—what level of analysis does 'tribe' denote? A small group? A larger confederation? One finds scattered critiques of the use of 'tribe' as equivalent to a 'dialect' in, e.g., Stewart (2017) who critiques the idea that Bedouin dialects correspond easily with a territory or space on a map and that they form uncomplicated groups, and who also shows that a single tribal group can show significant linguistic variation across its members.
- 10 This is not always the seventh century expansions—in cases where expansion was later (e.g., Sudan), the earliest layer is still often taken as the most fundamentally important one.
- 11 Holes continues on to contrast the idea of variation within the language of the conquerors against Ferguson's koine idea—however, the framework here recognizes that they are both "big-bang" theories at their heart. Perhaps the idea most in-line with the argument in this paper is Cohen (1962), which Holes summarizes as saying, "it looks more likely that Ferguson's features are the recent *end point* of a process of partial convergence than an ancient single *start point* (p. 8)"—this point tends to be overshadowed by the ongoing, almost subconscious adherence to the big-bang framework.
- 12 For American English, Stanford et al. (2012) explicitly investigated the first-effective founder principle, and found that while in the Northeastern United States a founder effect was operative for nearly two-hundred years, recent changes in modern patterns of contact and dialect use have begun to unravel the earlier situation and that this change is quite rapid, heading towards completion within three generations.
- 13 Certainly it is interesting to compare Holes' Table 5.2, which shows Omani "B" dialects, and as expected, to find these features in Zanzibar Arabic (Reinhardt 1894)—clearly the B dialects were ascendent in some areas even in the relatively recent past.
- 14 A brief explanation of the isoglosses: The 3mp *humma* forms contrast with *hum* forms, and appears to have spread first to the 3mp from the 3fp, before moving to the 2mp *ʔintu:ma*, with most dialects subsequently losing the feminine plural forms. The 1cs pronoun has the first syllable lengthened, the origin of which is unclear, but it well attested in early sources (Lane 1968, vol. 1, pp. 103–4, who cites Ibn Jinni [d. 392/1002]). The *δu*: plural demonstrative forms are clearly formed by analogy with the verbal system, and are virtually exclusive to North Africa (Magidow 2016). The merged gender in suffix pronouns appear to be a major impetus for merging of gender in other pronouns, and may be phonological in nature (loss of a/i short vowel contrasts). **ʔaʃku:n* is derived from **ʔa(yy) ʃayʔ yaku:n* 'which entity is this', but note that *ʃayʔ* should be understood as neutral to animacy (i.e., "entity" not "thing"). The *-ayya* pronoun suffix (e.g., Algiers *ʃkun anti-yya* 'who are you' (Boucherit 2006, p. 63)) appears to derive from emphatic suffixes found on demonstratives. *waqta:ʃ* derives from **waqt aʃ* 'what time?' Dialects across the Arab world have the unusual form 2cs *ʔinti:na*, but those in the Arabian peninsula are feminine forms derived from the verb conjugation suffix, while the North African dialects appear to be originally masculine forms (for example, in Tangiers according to Aguadé (2016) the 2 ms *nti:na* contrasted with 2fs *nti*, which then merged with the 2cs) perhaps derived again from emphatic demonstratives like Djidjelli *ltəmmi:na* 'there' (against a basic *ltəm̄m* 'there')—only the demonstrative related emphatic is considered here, since the verbally derived form in eastern dialects is too likely to be a result of parallel development.
- 15 Anjra, Morocco based on Vincente (2000); Larache, Morocco based on Guerrero (2016); Fez, Morocco based on Caubet (1993); Tlemcen, Algeria based on Marçais (1902); Djidjelli, Algeria based on Marçais (1956).
- 16 Certainly some of the features appear to be quite local, such as *ʔanti:na* 2cs, so dialects further east simply lack access to those forms and so cannot adopt them.

- 17 'an taḥḍura tuḥāsiba-hum fa-qad inḥadarū 'alā 'an yūšawwifū-ka kitābayni, 'wenn Du herkommst, um mit ihnen abzurechnen, denn sie sind in der Absicht heruntergekommen, Dir zwei Schriftstücke', P.Vind.Arab. III 70 https://www.apd.gwi.uni-muenchen.de/apd/show2.jsp?papname=Diem_Vind_III_700&line=4 (accessed on 1 July 2021), from the The Arabic Papyrology Database (www.naher-osten.lmu.de/apd (accessed on 30 September 2021)).
- 18 In writing this section, it proved surprisingly difficult to find a specific, clear explanation of when this supposed cut-off occurred. Arabic dialectologists clearly seem to agree on the parallel between Cyprus and Malta, but do not actually make explicit how the historical situation developed in Cyprus.
- 19 This is of course very nearly the traditional narrative of the Arab conquests, that a group of nomads imposed their language on much more populous groups. Magidow (2013) devotes significant effort to trying to determine how this could have occurred, and while his argument that it is likely due to depopulating events that took place at the time of the Arab conquests, that claim must be re-evaluated in light of the arguments here.
- 20 The year 1961 is the earliest year with data in the World Bank dataset, but it does show land use prior to the more intensive industrialization of the latter 20th century. Data accessed 17 September 2021 via <https://databank.worldbank.org/reports.aspx?source=2&series=AG.LND.AGRI.ZS> (accessed on 30 September 2021). Observe also the shading in Figure 3 in Europe and Turkey, which shows sufficiently high rainfall that we would expect desert-adapted nomadism to be rare in those areas.
- 21 One reviewer notes that climate change and other factors may have changed the shape of these corridors over time, so today's high density areas may not have been yesterday's. This is absolutely an important concern, but one which would probably only be able to be integrated into a more focused historical study of a single area.
- 22 There is not a clear consensus on the required rainfall for dry-land agriculture, with 200 mm a low minimum, so 300 mm was selected as a likely requirement for intensive settlement.
- 23 This approach was partly inspired by the work of Nichols (1992, 2003) who proposes the idea that easily traversed terrain tends to have lower linguistic diversity, while more difficult terrain tends to have greater dialect diversity, at least in part due to the difficulty of supplanting existing speakers in difficult terrain. Her analysis operates at much broader time-scales than considered here, and the heuristic here takes into account specific characteristics of the SWANA region and its peoples, but one is strongly tempted to recognize in Yemen a "residual zone" in her terminology.
- 24 This also correlates well with Arabicization in general in North Africa, where Tamazight languages are well maintained in high elevation areas, while in flat countries like Tunisia they have largely shifted to Arabic.
- 25 This contrasts strongly with the claim in de Jong (2000, p. 14), repeated in Owens (2006, pp. 239–40) that many of the tribes in the Sinai have continuously inhabited the area since pre-Islamic times. Given the model here, this claim should be treated with extreme skepticism. See also Stewart (2017) for a similar point about the history of the Sinai.
- 26 Al-Jallad cites here Ingham 1994, which actually takes quite a nuanced view towards the Najdi dialects, noting their apparent linguistic conservatism (and the fetishization of that conservatism by some observers), but also noting that there are migrations within this area, and that the richest area, between Qašim and Mecca was "much prized and competition over it was fierce," implying continuous changes of area over time. Note that Al-Jallad in his 2012 dissertation makes a similar claim regarding Najdi Arabic (p. 219), and while he has largely disavowed his dissertation (personal communication), it does illustrate that this idea has been current within Arabic dialectology and linguistics.

References

- Aboud-Haggar, Soha. 2006. Dialects: Genesis. In *Encyclopedia of Arabic Language and Linguistics*. Edited by Kees Versteegh. Leiden: Brill, vol. 1, pp. 613–22.
- Aguadé, Jordi. 2016. The Arabic Dialect of Tangier Across a Century. In *Arabic Varieties: Far and Wide. Proceedings of the 11th International Conference of AIDA—Bucharest, 2015*. Edited by George Grigore and Gabriel Biṭună. Bucharest: Editura Universității din București, pp. 21–27.
- Aguadé, Jordi. 2018. The Maghrebi Dialects of Arabic. In *Arabic Historical Dialectology: Linguistic and Sociolinguistic Approaches*. Edited by Clive Holes. New York: Oxford University Press, pp. 29–63.
- Al-Jallad, Ahma. 2017. The Earliest Stages of Arabic and Its Linguistic Classification. In *The Routledge Handbook of Arabic Linguistics*, 1st ed. Edited by Elabbas Benmamoun and Reem Bassiouney. Abingdon and New York: Routledge, pp. 315–31. [CrossRef]
- Al-Jallad, Ahmad. 2009. The Etymology of the Indicative Augment B- in Some Neo-Arabic Dialects. In *Afroasiatic Studies in Memory of Robert Hetzron: Proceedings of the 35th Annual Meeting of the North American Conference on Afroasiatic Linguistics*. Edited by Charles Häberl. Cambridge: Cambridge Scholars Press.
- Al-Wer, Enam. 2003. New Dialect Formation: The Focusing of -Kum in Amman. In *Social Dialectology: In Honor of Peter Trudgill*. Edited by David Britain and Jenny Cheshire. Amsterdam: Benjamins.
- Al-Wer, Enam. 2007. The Formation of the Dialect of Amman: From Chaos to Order. In *Arabic in the City*. Edited by Catherine Miller, Enam Al-Wer, Dominique Caubet and Janet C. E. Watson. London and New York: Routledge.
- Barfield, Thomas J. 1993. *The Nomadic Alternative*. Englewood Cliffs: Prentice Hall.
- Barth, Frederik. 1961. *Nomads of South Persia*. Boston: Little, Brown.

- Behnstedt, P., and Manfred Woidich. 2011. *Wortatlas der arabischen Dialekte: Verben, Adjektive, Zeit und Zahlen*. Leiden and Boston: Brill, vol. III.
- Behnstedt, Peter, and Manfred Woidich. 2013. *Arabic Dialectology*, online ed. Edited by Jonathan Owens. Revised 6 September 2017. Oxford: Oxford University Press, vol. 1. [CrossRef]
- Behnstedt, Peter, and Manfred Woidich. 2018. The Formation of the Egyptian Arabic Dialect Area. In *Arabic Historical Dialectology: Linguistic and Sociolinguistic Approaches*. Edited by Clive Holes. New York: Oxford University Press, pp. 64–95.
- Behnstedt, Peter. 1990. M'addamiyeiye: Ein Neuer Qeltu-Dialekt Aus Syrien. *Zeitschrift Für Arabische Linguistik* 22: 44–66.
- Behnstedt, Peter. 1997. *Sprachatlas von Syrien*. Weisbaden: Harrassowitz.
- Behnstedt, Peter. 2016. *Dialect Atlas of North Yemen and Adjacent Areas*. Edited by Maribel Fierro, M. Şükrü Hanioglu and Renata Holod. Translated by Gwendolin Goldbloom. Leiden and Boston: Brill, vol. 114.
- Benkato, Adam. 2019. From Medieval Tribes to Modern Dialects: On the Afterlives of Colonial Knowledge in Arabic Dialectology. *Philological Encounters* 4: 2–25. [CrossRef]
- Blanc, Haim. 1953. *Studies in North Palestinian Arabic; Linguistic Inquiries among the Druzes of Western Galilee and Mt. Carmel*. Jerusalem: Israel Oriental Society.
- Blanc, Haim. 1964. *Communal Dialects in Baghdad*. Cambridge: Harvard University Press.
- Blau, Joshua. 1977. The Beginnings of the Arabic Diglossia: A Study of the Origins of Neoarabic. *Afroasiatic Linguistics* 4: 175–201.
- Borg, Alexander. 2004. *A Comparative Glossary of Cypriot Maronite Arabic: With an Introductory Essay (Handbook of Oriental Studies)*. Leiden: Brill Academic Pub.
- Borg, Alexander. 2006. Cypriot Maronite Arabic. In *Encyclopedia of Arabic Language and Linguistics*, online ed. Edited by Kees Versteegh. Leiden: Brill.
- Boucherit, Aziza. 2006. Algiers Arabic. In *Encyclopedia of Arabic Language and Linguistics*. Edited by Kees Versteegh. Leiden: Brill, vol. 1, pp. 58–66.
- Bowern, Claire. 2006. Punctuated Equilibrium and Language Change. In *Encyclopedia of Language and Linguistics*. Edited by Keith Brown. Amsterdam: Elsevier, pp. 286–89.
- Britain, David. 2008. Space and Spatial Diffusion. In *The Handbook of Language Variation and Change*. Edited by J. K. Chambers, Peter Trudgill and Natalie Schilling-Estes. Malden: Blackwell Publishing Ltd., pp. 603–37. [CrossRef]
- Caubet, Dominique. 1993. *L'Arabe Marocain: Phonologie et Morphosyntaxe*. Paris: Éditions Peeters.
- Cohen, David. 1962. Koinè, langues communes et dialectes arabes. *Arabica* 9: 119–44. [CrossRef]
- Conrad, Lawrence I. 1981. The Plague in the Early Medieval Near East. Ph.D. dissertation, Princeton University, Princeton, New Jersey.
- Corriente, Federico. 2014. Again on the 'Yemenite Connection' of Andalusí Arabic and Other Western Arabic Dialects. *Folia Orientalia* 51: 9–26.
- Cowan, William. 1966. Two Notes on Arabic Dialectology. *Journal of the American Oriental Society* 86: 416–18. [CrossRef]
- Cysouw, Michael, and Jeff Good. 2013. Languoid, Doculect and Glossonym: Formalizing the Notion 'Language'. *Language Documentation & Conservation* 7: 331–59.
- de Jong, Rudolf E. 2000. *A Grammar of the Bedouin Dialects of the Northern Sinai Littoral: Bridging the Linguistic Gap between the Eastern and Western Arab World*. Leiden: Brill.
- Dixon, Robert M. W. 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
- Dols, Michael W. 1974. Plague in Early Islamic History. *Journal of the American Oriental Society* 94: 371–83. [CrossRef]
- Donner, Fred McGraw. 1989. The Role of Nomads in the Near East in Late Antiquity (400–800 C.E.). In *Tradition and Innovation in Late Antiquity*. Edited by Frank M. Clover and R. Stephen Humphreys. Madison: University of Wisconsin Press, pp. 73–87.
- Ferguson, Charles A. 1959. The Arabic Koine. *Language* 35: 616–30. [CrossRef]
- Fischer, Wolfdietrich. 2006. Classical Arabic. In *Encyclopedia of Arabic Language and Linguistics*. Edited by Kees Versteegh. Leiden: Brill, vol. 1, pp. 397–405.
- Fück, Johann. 1950. *Arabiya, Untersuchungen Zur Arabischen Sprach- Und Stilgeschichte*. Berlin: Akademie-Verlag.
- Garbell, Irene. 1958. Remarks on the Historical Phonology of An Eastern Mediterranean Arabic Dialect. *Word* 14: 303–37. [CrossRef]
- Guerrero, Jairo. 2016. Notes on the Arabic Dialect of Larache (North-Western Morocco). In *Lisan Al-Arab: Studies in Contemporary Arabic Dialects. Proceedings of the 10th International Conference of AIDA. Qatar University 2013*. Edited by Muntasir F. Al-Hamad, Rizwan Ahmed and Hafid I. Aloui. Vienna: LIT Verlag, pp. 195–208.
- Hetzron, Robert. 1976. Two Principles of Genetic Reconstruction. *Lingua* 38: 89–108. [CrossRef]
- Holes, Clive. 1995. Community, Dialect and Urbanization in the Arabic-Speaking Middle East. *Bulletin of the School of Oriental and African Studies, University of London* 58: 270–87. [CrossRef]
- Holes, Clive. 2015. *Dialect, Culture, and Society in Eastern Arabia: Phonology, Morphology, Syntax, Style*. Leiden: Brill, vol. 3.
- Holes, Clive, ed. 2018a. *Arabic Historical Dialectology: Linguistic and Sociolinguistic Approaches*. Oxford: Oxford University Press.
- Holes, Clive. 2018b. The Arabic Dialects of the Gulf: Aspects of Their Historical and Sociolinguistic Development. In *Arabic Historical Dialectology: Linguistic and Sociolinguistic Approaches*. Edited by Clive Holes. New York: Oxford University Press, pp. 112–47.
- Hourani, Guita G. 1998. A Reading in the History of the Maronites of Cyprus From the Eighth Century to the Beginning of British Rule. *Journal of Maronite Studies* 2: 1–16.

- Hoyland, Robert G. 2009. Arab Kings, Arab Tribes and the Beginnings of Arab Historical Memory in Late Roman Epigraphy. In *From Hellenism to Islam: Cultural and Linguistic Change in the Roman Near East*. Edited by Hannah M. Cotton, Robert G. Hoyland, Jonathan J. Price and David J. Wasserstein. Cambridge: Cambridge University Press, pp. 374–400.
- Ingham, Bruce. 1982. *North East Arabian Dialects*. Boston: Kegan Paul International.
- Jastrow, Otto. 1978. *Die Mesopotamisch-Arabischen Qəltu-Dialekte*. Wiesbaden: Kommissionsverlag Franz Steiner GMBH.
- Jastrow, Otto. 2002. Arabic Dialectology: The State of the Art. *Israel Oriental Studies* 20: 347–64.
- Joffé, E. G. H. 1990. Relations between Libya, Tunisia and Malta up to the British Occupation of Malta. *Libyan Studies* 21: 65–73. [\[CrossRef\]](#)
- Kennedy, Hugh. 2007. *The Great Arab Conquests: How the Spread of Islam Changed the World We Live in*. London: Weidenfeld and Nicolson.
- Kossmann, Maarten G. 2013. *The Arabic Influence on Northern Berber*. Leiden: Brill.
- Labov, William. 2001. *Principles of Linguistic Change, Volume 2: Social Factors*. Malden: Blackwell.
- Lane, Edward. 1968. *An Arabic-English Lexicon*. Beirut: Librairie Du Liban Publishers.
- Lentin, Jérôme. 2018. The Levant. In *Arabic Historical Dialectology: Linguistic and Sociolinguistic Approaches*. Edited by Clive Holes. New York: Oxford University Press, pp. 170–205.
- Luttrell, Anthony T. 1975. Approaches to Medieval Malta. In *Approaches to Medieval Malta: Studies on Malta before the Knights*. Edited by Anthony T. Luttrell. London: British School at Rome, pp. 1–70.
- Luttrell, Anthony T. 1965. Malta and the Aragonese Crown: 1282–1530. *Journal of the Faculty of Arts, University of Malta* 3: 1–9.
- Magidow, Alexander. 2013. Towards a Sociohistorical Reconstruction of Pre-Islamic Arabic Dialect Diversity. Ph.D. dissertation, The University of Texas at Austin, Austin, TX, USA.
- Magidow, Alexander. 2016. Diachronic Dialect Classification with Demonstratives. *Al-Arabiyya* 49: 91–115.
- Magidow, Alexander. 2017. Linguistic History and the History of Arabic: A Speech Communities Approach. In *Arabic in Context: Celebrating 400 Years of Arabic at Leiden University*. Edited by Ahmad Al-Jallad. Leiden: Brill, pp. 405–40.
- Magidow, Alexander. Forthcoming. *Dialect Groups and Dialect History: Reconstructing the History of Arabic Dialects via Closed-Class Morphemes*. Oxford: Oxford University Press.
- Marçais, Philippe. 1956. *Le Parler Arabe de Djidjelli, Nord Constantinois, Algérie*. Paris: Librairie d'Amérique et d'Orient.
- Marçais, William. 1902. *Le Dialecte Arabe Parlé a Tlemcen: Grammaire, Textes et Glossaire*. Paris: Publications de L'École des Lettres D'Alger.
- Marçais, William. 1938. Comment L'Afrique Du Nord A Été Arabisée. *Annales de l'Institut d'études Orientales d'Alger* 4: 171–92.
- Miller, Ann M. 1986. The Origin of the Modern Arabic Sedentary Dialects: An Evaluation of Several Theories. *Al-Arabiyya* 19: 47–74.
- Miller, Catherine. 2004. Variation and Change in Arabic Urban Vernaculars. In *Approaches to Arabic Dialects: Collection of Articles Presented to Manfred Woidich on the Occasion of His Sixtieth Birthday*. Edited by Martine Haak, Rudolf de Jong and Kees Versteegh. Leiden: Brill, pp. 117–205.
- Miller, Catherine. 2005. Between Accommodation and Resistance: Upper Egyptian Migrants in Cairo. *Linguistics* 43: 903–56. [\[CrossRef\]](#)
- Miller, Catherine. 2007. Arabic Urban Vernaculars: Development and Change. In *Arabic in the City: Issues in Dialect Contact and Language Variation*. London and New York: Routledge.
- Milroy, Lesley. 2008. Social Networks. In *Handbook of Language Variation and Change*. Edited by John K. Chambers, Peter Trudgill and Natalie Schilling-Estes. Malden: Wiley-Blackwell.
- Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press.
- Nichols, Johanna. 2003. Diversity and Stability in Language. In *The Handbook of Historical Linguistics*. Edited by Brian Joseph and Richard Jones. London: Blackwell, pp. 283–310.
- Ostler, Nicholas. 2005. *Empires of the Word: A Language History of the World*. London: HarperCollins.
- Owens, Jonathan. 2003. Arabic Dialect History and Historical Linguistic Mythology. *Journal of the American Oriental Society* 123: 715–40. [\[CrossRef\]](#)
- Owens, Jonathan. 2006. *A Linguistic History of Arabic*. Oxford: Oxford University Press.
- Owens, Jonathan. 2018. Dialects (Speech Communities), the Apparent Past, and Grammaticalization: Towards an Understanding of the History of Arabic. In *Arabic Historical Dialectology: Linguistic and Sociolinguistic Approaches*. Edited by Clive Holes. New York: Oxford University Press, pp. 206–53.
- Palva, Heikki. 2006. Dialects: Classification. In *Encyclopedia of Arabic Language and Linguistics*. Edited by Kees Versteegh. Leiden: Brill, vol. 1, pp. 604–11.
- Palva, Heikki. 2009. From Qəltu to Qələt: Diachronic Notes on Linguistic Adaptation in Muslim Baghdad Arabic. In *Arabic Dialectology: In Honor of Clive Holes on the Occasion of His Sixtieth Birthday*. Edited by Enam al-Wer and Rudolf de Jong. Leiden: Brill, pp. 18–40.
- Papaconstantinou, Arietta. 2012. Why Did Coptic Fail Where Aramaic Succeeded? In *Multilingualism in the Graeco-Roman Worlds*. Edited by Alex Mullen and Patrick James. Cambridge: Cambridge University Press, pp. 58–76. [\[CrossRef\]](#)
- Procházka, Stephan. 2018. The Northern Fertile Crescent. In *Arabic Historical Dialectology: Linguistic and Sociolinguistic Approaches*. Edited by Clive Holes. New York: Oxford University Press, pp. 257–92.
- Reinhardt, Carl. 1894. *Ein Arabische Dialekt Gesprochen in 'Omān Und Zanzibar*. Stuttgart and Berlin: W. Spemann.
- Rosenhouse, Judith. 2006. Bedouin Dialects. In *Encyclopedia of Arabic Language and Linguistics*. Leiden: Brill, vol. 1, pp. 259–69.
- Stanford, James N., Thomas A. Leddy-Cecere, and Kenneth P. Baclawski Jr. 2012. Farewell to the Founders: Major Dialect Changes Along the East-West New England Border. *American Speech* 87: 126–69. [\[CrossRef\]](#)
- Stewart, Frank H. 2017. Bedouin Dialects of the Sinai: A Review Article. *Jerusalem Studies in Arabic and Islam* 44: 169–220.

- Taine-Cheikh, Catherine. 2006. Ḥassāniyya Arabic. In *Encyclopedia of Arabic Language and Linguistics*, online ed. Edited by Kees Versteegh. Leiden: Brill, vol. 4, pp. 687–99.
- Trudgill, Peter. 1986. *Dialects in Contact*. Oxford: Basil Blackwell.
- Tsiapera, M. 1969. *A Descriptive Analysis of Cypriot Maronite Arabic*. Pdf. Den Haag: Mouton.
- van Putten, Marijn, and Adam Benkato. 2017. The Arabic Strata in Awjila Berber. In *Arabic in Context: Celebrating 400 Years of Arabic at Leiden University*. Edited by Ahmad Al-Jallad. Leiden: Brill, pp. 476–502.
- van Putten, Marijn. 2017a. The Development of the Triphthongs in Quranic and Classical Arabic. *Arabian Epigraphic Notes* 3: 47–74.
- van Putten, Marijn. 2017b. The Illusory Yemenite Connection of Andalusī Arabic. *Zeitschrift Für Arabische Linguistik* 66: 5. [[CrossRef](#)]
- Versteegh, Kees. 1984. *Pidginization and Creolization: The Case of Arabic*. Amsterdam: Benjamins.
- Versteegh, Kees. 2004. Pidginization and Creolization Revisited: The Case of Arabic. In *Approaches to Arabic Dialects: A Collection of Articles Presented to Manfred Woidich on the Occasion of His Sixtieth Birthday*. Edited by Martine Haak, Rudolf Erik De Jong and Kees Versteegh. Leiden: Brill, pp. 343–57.
- Versteegh, Kees. 2014. *The Arabic Language*, 2nd ed. Edinburgh: Edinburgh University Press.
- Vicente, Ángeles. 2012. Sur la piste de l'arabe marocain dans quelques sources écrites anciennes (du XIIe au XVIe siècle). In *De Los Manuscritos Medievales a Internet: La Presencia del Árabe Vernáculo en las Fuentes Escritas*. Edited by Mohamed Meouak, Pablo Sánchez and Ángeles Vicente. Zaragoza: Universidad de Zaragoza, Área de Estudios Arabes e Islámicos, pp. 103–20.
- Vicente, Ángeles. 2000. *El Dialecto Árabe de Anjra (Norte de Marruecos)*. Zaragoza: Universidad de Zaragoza.
- Walters, Keith. 2006. Communal Dialects. In *Encyclopedia of Arabic Language and Linguistics*. Leiden: Brill, vol. 1, pp. 442–48.
- Watson, Janet. 2018. South Arabian and Arabic Dialects. In *Arabic Historical Dialectology: Linguistic and Sociolinguistic Approaches*. Edited by Clive Holes. New York: Oxford University Press, pp. 316–34.
- Wikle, Thomas, and Guy Bailey. 1996. The Spatial Diffusion of Linguistic Features in Oklahoma. *Proceedings of the Oklahoma Academy of Science* 77: 1–15.
- Zelinsky, Wilbur. 1992. *The Cultural Geography of the United States: A Revised Edition*. Englewood Cliffs: Prentice Hall.

Article

A Historical Reconstruction of Some Pronominal Suffixes in Modern Dialectal Arabic

Phillip W. Stokes

Department of Modern Foreign Languages and Literatures, The University of Tennessee,
Knoxville, TN 37996, USA; pstokes2@utk.edu

Abstract: The morphology of the pronominal suffixes in dialectal Arabic are of particular interest for scholars of the history of Arabic for two main reasons. First, multiple dialects attest suffixes that, from a comparative perspective, apparently retain final short vowels. The second and more complicated issue concerns the vowels which precede the suffixes in the dialects, which are thought to either have been case inflecting or epenthetic. In this paper, I take up Jean Cantineau's "embarrassing question" of how to account for the development of the vowels of the pronominal suffixes. Based on data from dialectal *tanwīn* in modern dialects, and attestations from pre-modern texts as well, I will argue that the pre-suffix vowels did originate in case inflecting vowels, but that no historical model heretofore proposed can satisfactorily account for how the various dialectal forms might have arisen. I identify two major historical developments and propose models for each. First, I suggest that dialects in which the pre-suffixal vowels harmonized with the suffix vowels developed via a process of harmonization across morpheme boundaries before the loss of final short vowels. For dialects in which one vowel is generalized, I argue that a post-stress neutralization took place, which led to a single vowel both before suffixes and *tanwīn* as well. Finally, I rely on evidence from the behavior of the suffixes to argue that the final vowel of the 3fs suffix was originally long, but that those of the 3ms, 2ms, and 2fs were most likely short.

Citation: Stokes, Phillip W. 2021. A Historical Reconstruction of Some Pronominal Suffixes in Modern Dialectal Arabic. *Languages* 6: 147. <https://doi.org/10.3390/languages6030147>

Keywords: Arabic dialects; historical dialectology; historical linguistics

Academic Editors: Simone Bettega and Roberta Morano

Received: 24 June 2021

Accepted: 23 August 2021

Published: 31 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The morphology of the pronominal suffixes in dialectal Arabic are of particular interest for scholars of the history of Arabic for two main reasons. First, multiple dialects attest suffixes that, from a comparative perspective, apparently retain final short vowels: dialectal *daras* "he studied", < **darasa* but *abū-ki* < **abVV-ki*, "your (fsg) father." This is despite the fact that most assume a complete loss of final short vowels in the pre-modern ancestors of all Arabic varieties. The second, and more complicated issue concerns the vowels which precede the suffixes in the dialects. In Classical Arabic (henceforth CIAR), this position was occupied by a short case vowel: *bāb+u/i/a +ka* "your door (nom/gen/acc)." In the modern dialects, which lack morphosyntactic case marking, these vowels vary: *bāb-Vk* "idem." Several attempts have been made to account for the development of these vowels from a functional case system (Diem 1991; Jastrow 1991; Blau 2006), though none has yet proved entirely satisfactory. The thorniest issue from a historical perspective is how to account for the functional shift of the pre-suffixal vowels, from morpho-syntactic inflection to their current forms, while also assuming the loss of final short vowels, which should have eliminated several of the vowels of the pronominal suffixes. Alternatively, several scholars have recently argued that the ancestors of the dialects came from varieties of Arabic that lacked case completely (Retsö 1994; Owens 2006, 2018). In this alternative scenario, the pre-suffixal vowels were epenthetic vowels inserted to resolve increasingly un-tolerated consonant clusters.

In this paper I argue in favor of the case vowel origin and propose reconstructions that address both issues: final vowel length and the nature and development of the pre-

suffixal vowel. Following a discussion of the previous proposals and their deficiencies, I discuss both of these questions, with special focus on the problems posed by the third and second person singular forms. I will argue that data from remnants of case vowels before *tanwīn* in the dialects (dialectal *tanwīn*), as well as historical data in which case was retained longer before suffixes than elsewhere, can help understand the development of these vowels. The implications of these observations for the question of the historical length of the vowels of the suffixes will be discussed at length in Section 3. The ultimate goal of these reconstructions is to gain deeper historical insight into the most common surface forms attested across the spectrum of Arabic dialects, as well connect them together in the broader picture of Arabic linguistic history.

2. Pre-Suffix Vowel: Case Vowel or Epenthesis?

The pronominal suffix paradigms for the modern Arabic dialects do not inflect for case but nevertheless attest vowels before the pronominal suffixes. The following data (Table 1) serve as examples of the regular realizations of these suffixes, regardless of syntax (Fischer and Jastrow 1980, p. 81; Jordanian from Al-Wer (2007, p. 510)):

Table 1. Levantine and N. African Singular Pronominal Suffixes.

	Jordanian	Cairo	Damascus	Tunis
3ms	-o, -V:	-uh, -h	-o, -(h)	-u, -h
3fs	-ha, -a	-ha	-a, -ha	-ha
2ms	-ak, -k	-ak, -k	-ak, -k	-ik, -k
2fs	-ik, -ki	-ik, -ki	-ik, -ki	-ik, -k

The origin and nature of the vowels preceding the pronominal suffixes in the modern dialects has become rather controversial. They have traditionally been interpreted as frozen reflexes of case vowels,¹ which eventually harmonized with the short vowels of the suffixes, e.g., nominative *u* was frozen based on harmony with 3ms *hu, etc. (Birkeland 1952, pp. 12, 19; Cantineau 1936/1937, p. 180, vol. 2; Fischer and Jastrow 1980, p. 42; Diem 1991). To account for the vowel harmony, which would ostensibly be possible only after the breakdown of case, and thus presumably the loss of final short vowels, Blau (2006, pp. 87–88) appeals to the intentionality of speakers, who decided to avoid a homophony between masculine and feminine forms, which after the breakdown of case and loss of final short vowels, should have both become ***bāb-k* “your (msg and fsg) door.” Instead, they inserted an *a* vowel before the suffix for the masculine and an *i* to indicate the feminine. Blau cites a phenomenon in CIAr called *naql*, “transfer”, wherein speakers would insert a vowel homophonous with the suffixed case vowel on noun of the pattern of CVCC: *al-bakru* (nom sing) > *al-bakuru*, etc. (ibid., p. 88). However, Blau does not address the biggest issue with his proposal. Given that he believes that a *naql* would have occurred only once a final cluster had been created, it is unclear how speakers would have retrieved the etymological *a* vowel to mark masculine or the *i* vowel for the feminine if they had been lost word-finally.

Diem (1991) offers a different explanation. He recognizes that if loss of case was caused, or at least accompanied by loss of final short vowels, then the vowels of at least some of the pronominal suffixes, like 3ms **-hu* and 2ms **-ka*, should have also been lost. If such was the case, however, then of course the final vowel, with which the case vowel would harmonize, would have been eliminated. Diem, still believing the initial vowels to be remnants of case vowels, argues that the breakdown of the case system was caused by syntactic factors, especially the redundant nature of the Semitic case system, and not the result of a phonetic loss of short final vowels. In Diem’s scenario, case breakdown resulted initially in a state where the short case vowels were in free variation (ibid., pp. 301–3). Eventually, before pronominal suffixes they were harmonized with the suffix vowel and frozen; in word-final position (including, for Diem, before *tanwīn*), one case was levelled.²

This is the state of affairs in what Diem calls “nomadic dialects” (i.e., Bedouin dialects) (ibid., pp. 303–4). Most non-nomadic dialects eventually witnessed the total loss of these remnants due to their lack of any meaningful syntactic function.

This is an attractive explanation³ because it solves the main objection to the traditional assumptions (see above regarding Blau’s 2006 proposal), namely that case was lost due to loss of final short vowels. If case marking was reanalyzed as marked solely by word order etc., then these variants might have eventually been in free variation.⁴ It should be noted here that Diem recognizes the loss of final short vowels at some point, but argues that the breakdown of case was not related to this loss (or, probably, multiple losses), and presumably preceded it.

Alternatively, Owens (2006, chp. 8) has argued that these initial vowels originated as epenthetic vowels, completely unrelated to case vowels, the quality of which varied depending on the following consonant/vowel: 3ms *hu > u-hu (addition of epenthetic vowel) > uh (loss of final short vowel).⁵ Owens’s arguments against reconstructing case in the ancestor of all Arabic varieties is unconvincing. There is thus no reason a priori to suggest these vowels cannot be remnants of case vowels. Still, the complete loss of case vowels could have led to their loss before pronominal suffixes as well, presumably in this case, per Blau (2006), via the levelling of pausal forms to non-pausal position, which in turn could have led speakers to insert epenthetic vowels to resolve these newly-created consonant clusters. In other words, since Ø-marked forms were levelled to other contexts where they would be expected to be preserved if the loss of case were the result of a regular phonetic change (*v > Ø/C_#), such as the construct, then it is also possible that these Ø-marked forms were levelled to the position before pronominal suffixes.

Deciding between these options is difficult due to the uneven knowledge of the contemporary Arabic dialects, and the virtual absence of historical data for the ancestors to the modern dialects. Nevertheless, a careful examination of the implications of these proposals yields meaningful insights. If we imagine that the vowels in question originated as case vowels, then we must account for speakers presumably sacrificing the morphosyntactic role of case vowels before pronominal suffixes in favor of a non-phonemic epenthetic function. Scholars have thus mostly assumed that this happened following the breakdown of case (cf. Diem 1991). Most likely, then, the loss of case marking on the un-suffixed nouns motivated the levelling of one form, as in the external masculine plural oblique form *in*. The presence of very similar suffix forms in Aramaic, and probably for Hebrew as well, provide attractive parallels (Table 2; Aramaic and Hebrew data taken from Hasselbach 2014, pp. 204–5):

Table 2. Reconstructed Aramaic (Syriac) and Hebrew Pronominal Suffixes.

	Aramaic (Syriac)	Hebrew
3ms	<i>malk-eh</i> < * <i>malki-hi</i> (Gen)	<i>malkō</i> < * <i>malku-hu(:)</i> (Acc)
2fs	<i>malk-ek(y)</i> < * <i>malki-kt</i> ¹ (Gen)	<i>malkēk</i> < * <i>malki-ki</i> (Gen)
2ms	<i>malk-aḵ</i> < * <i>malka-ka</i> (Acc)	(non-Tiberian) <i>malkaḵ</i> < * <i>malka-ka</i> (Acc) ²

¹ As indicated by the mater lectionis <Y> of the Aramaic form, the *i* vowel here was apparently long. Benjamin Suchard has suggested to me (p.c.) that this long *i* was the result of **ki* > *kī* based on contamination with the imperfect ending. ² The Tiberian forms are complex. In non-pausal position, Tiberian *malka-kā*, but pausally, *malkēkā*. The latter form suggests < **malki-kah* (Suchard, p.c.), or possibly analogy with the feminine singular form (Suchard 2020, pp. 194–96).

There are several reasons to favor positing etymological case vowels as the origin of the pre-suffixal vowels instead of originally vowelless consonant clusters that were only later broken up via epenthesis. First and foremost is a methodological argument. We can confidently reconstruct case for both Proto-Semitic and Proto-Arabic based on comparative and internal data. A case system in which singular (and broken plural) nominal forms are inflected for three cases (nominative, genitive, and accusative), while duals and sound masculine plurals are inflected for two (nominative and oblique) is attested in Akkadian, Ugaritic, and Canaanite, as well as ClAr (Huehnergard 2019, pp. 60–61). Classical Ethiopic

(Goʻoz) also attests a remnant of the same system, with an accusative/non-accusative system in which accusative is marked with short-*a*, and non-accusative is unmarked, likely as the result of the merger and loss of final **u* and **i* (>ə > ø/C_#) (Butts 2019, pp. 128–29).

Within Arabic, it is not only CIAR that attests this nominal case marking; it is attested, to varying degrees, in pre-Islamic Arabic corpora in the Safaitic (Ancient North Arabian), Nabataean, and Greek scripts as well (Al-Jallad 2018). As argued in Van Putten and Stokes (2018), the Quranic consonantal text (*rasm*) attests a partial case system, distinct from that attested in CIAR. Finally, from a methodological perspective, we should greatly prefer an explanation for the lack of a feature that is most parsimonious and consistent with standard historical linguistic methodology. In attempting to answer these questions we are faced with the scenario either that the Semitic languages in general, and Arabic specifically, should be sub-grouped historically based on the presence or absence of a single feature, or that nominal case marking, which is known to have been lost in a number of languages, was retained completely in some, only partially in others, and lost completely in the rest. The latter is more parsimonious than the former.

Further, Owens' argument that these vowels originated as epenthetic vowels is not as strong as his claims make it seem. Owens himself reconstructs these pronominal suffixes with final short vowels, which were only lost after vowel harmony with the newly inserted epenthetic vowels (Owens 2006, p. 248 ff.). He further argues that when a pronominal suffix is added to a word like *qalb* "heart", the combination would result in a chain of three consonants, CCC, which is often (but not always; see Watson (2007, pp. 340–41)) disallowed in Arabic dialects (ibid., pp. 107–11). This resulted in the insertion of an epenthetic vowel between the stem and the suffix: *qalb* + *hā* > *qalbhā* > *qalb-a-hā* "her heart." The same process was responsible for creating 2ms *ak*: *qalb* + *ka* > *qalbka* > *qalb-a-ka* "your (ms) heart", and 2fs *qalb* + *ki* > *qalbki* > *qalb-i-ki*, and so on. After this stage of epenthetic vowel insertion, final short vowels of these suffixes were lost: *qalb-a-ka* > *qalb-ak*.

While on the face of it this explanation might seem to account for the dialect data without resorting to the notion of case vowels, it is mitigated by some further considerations. First, for Owens's explanation of these forms to obtain, we would need to assume that all dialects resolved these CCC clusters in the same way, namely by inserting the epenthetic vowel after the second of the three consonants, i.e., CCCv(:) > CCvCv(:). In fact, however, as Owens implicitly acknowledges (ibid., pp. 108–9), many dialects resolve these clusters by inserting the epenthetic vowel after the initial consonant, so CCCv(:) > CvCCv(:) (on which, see Watson (2007, pp. 340–48)). In such cases, we would expect two different allomorphs of the suffixes to exist, distributed according to the phonotactic rules of the dialect, but this is not the case. That is, if CVCC nouns were the origin of the epenthetic vowels, as Owens maintains, then we would expect a distribution which matches the two patterns of epenthesis insertion: *qalb-a-k* for CCC > CvCC, and ***qalabk* for CCC > CCvC. Instead, we only find forms that would need to have originated in CCvC-inserting dialects.

Another question that Owens does not meaningfully answer is why epenthesis would have been necessary in most nominal patterns. Owens's examples, cited above, are typically CVCC (including CVW/YC), i.e., *qalb-ak* "you (msg) heart" and *bēt-ha* "her house", etc. Elsewhere epenthesis is not typical. For example, most dialects do not resolve CCV clusters: Jordanian (a CVCC dialect) *katabti* "you (fsg) wrote", and not ***katabiti* "idem" (Al-Wer 2007); Cairene (a CCVC dialect) *ḡasalṭi* "you (fsg) washed", and not ***ḡasaliti* (Woidich 2006, p. 330).⁶ As Owens admits, the pronominal suffixes, with the exception of 3fs (**hā*) and 1cp (**nā*), are to be reconstructed with final short vowels (reviewed above) (Owens 2006, pp. 239–58). Based on the dialectal data on which Owens relies, epenthesis at this initial stage, where the final short vowels of the suffixes were still present, would only be required for nouns of the pattern CVCC. It seems rather implausible that allomorphs necessary only for nouns of one (admittedly frequent) noun pattern were leveled everywhere. However, it is not impossible. If there are any dialects where Owens' reconstruction should be expected to accurately match the distribution, it is thus CCVC-repatterning dialects, like Cairene (Watson 2007, p. 341).

The decisive blow to Owens's reconstruction, however, comes upon further examination of the Cairene distribution. The Egyptian realization of **qalb-a-hā* "her heart" is 'alb-a-ha, **qalb-a-ka* "your (msg) hearth" is 'alb-ak, and **qalb-i-ki* "your (fsg.) heart" is 'alb-ik, each as Owens's reconstruction would predict. However, if the forms with epenthesis were generalized, as could be suggested by the 2ms and 2fs forms, we would expect 3fsg -aha to occur when suffixed to any noun, regardless of morphological pattern. Instead, however, we find that on nouns ending in a single consonant, the form is -ha, i.e., *ba'arit-ha* "her cow", and not ***ba'arita-ha*. The same noun, however, when suffixed with a 2ms or 2fs suffix, would have the supposedly epenthetic allomorph, i.e., *ba'arita-k* (msg.)/*ba'arita-k* (fsg.). In other words, in a dialect that fits Owens' pattern for the creation of epenthetic allomorphs, we would have to suppose that the epenthetic allomorphs of the 2ms and 2fs suffixes were generalized everywhere, but the 3fs suffix retained its original distribution (i.e., occurring only after CC clusters). Thus, Owens's theory does not successfully explain the particulars of the distribution of any dialect type.

The question remains how to model this development from etymological case vowel to the various surface forms. As noted above, both Blau and Diem believe, as I do, that nouns originally ended in a short vowel that marked nominal case on singular, feminine plural, and most broken plural patterns. The question of their development is tied up with that of the simplification and eventual loss of Proto-Arabic morphosyntactic case inflection. For Blau, the breakdown and loss of case is related to the loss of final short vowels, and the subsequent analogical extension of these caseless forms to non-final position. For Diem, on the other hand, case breakdown was only partially, if at all, related to short final vowel loss; rather, for him it was primarily due to the low functional yield of case in Semitic and Arabic.

While both Blau and Diem are undoubtedly correct that both analogy and sound change played roles, neither provided explanations which, in the end, were able to provide satisfactory reconstructions of the process. Blau is ultimately unable to compellingly answer how short final case vowels were lost, thus leading to case loss, without the simultaneous loss of short pronominal suffix vowels. Diem, on the other hand, does not manage to explain what led to the free variation of case vowels, and ultimately the loss of case. As Blau argued successfully elsewhere (Blau 1972), the low functional yield of case in Semitic is not, in and of itself, an explanation for its loss.

Another problem that plagues both accounts detailed above is that neither fully integrates the only other feature which provides direct evidence of case vowels that remain in the dialects, namely dialectal *tanwīn*. Both Blau (2002, p. 44–46; 2006) and Diem (1991, 303 ff.) assume that the attested surface forms of dialectal *tanwīn* reflect unaltered the case vowel which was ultimately generalized. That is, in their reconstructions, any examples of dialectal *tanwīn* -in are automatically assumed to reflect a frozen genitive *-in, any dialectal -an reflects accusative *-an, etc. As argued in Stokes (2020), however, there are reasons based on the dialectal data to doubt this straightforward identification. The clearest example is found in the southern Saudi dialects of Bal-Qarn and Banī 'Abādīl, where an interesting pausal/non-pausal distribution of dialectal *tanwīn* is attested: non-pausal *bēt-in* "a house"/pausal *bēt-u* "idem." The most likely interpretation of this is a vowel merger to a high vowel in final position, wherein before *tanwīn* the surface vowel is short front -i, whereas when word-final the vowel is back high -u (itself apparently lengthened rather than lost; see Stokes (2020, pp. 655–56) for details).

Other evidence suggests a similar merger occurred in most, if not all, ancestors of the modern dialects. For example, in both Bahraini and rural (*fellaḥī*) Iraqi dialects, there is an adverbial *tanwīn* -an that contrasts with the standard dialectal *tanwīn* form -in; crucially, these occur in contexts which could not have been CIAr borrowings (Stokes 2020, p. 650). Rather, the reflexes of dialectal *tanwīn* attest a stage in which phonemic distinction was lost between short case vowels before *tanwīn*: *u,*i,*a > V/C_#(n) (Stokes 2020, pp. 655–60). It is likely that such a merger is responsible for the surface forms of dialectal *tanwīn* in those dialects that retain it.

Finally, the traditional accounts offer only one set of reconstructions for dialects that attest a great deal of variation. In what follows, I will bring together data from dialects across the spectrum in order to provide historical explanations for each of the major patterns attested therein. Where possible and relevant, I incorporate data from dialectal *tanwīn* to inform the reconstruction. I will attempt to show that while the patterns attested in the dialects can effectively be derived from the same Proto-Arabic case-bearing situation, they require different scenarios—each of which differs from those proposed heretofore—to explain the surface manifestations. It is to the data, then, that we now turn.

3. Historical Development of Pronominal Suffixes

Obviously, there is not sufficient space here to treat every attested form. Further, the reconstructions offered here are not final; rather, it is hoped that they will constitute just one more step in the process of understanding the history of the dialects and their interrelationships. There are undoubtedly dialect-specific factors necessary to understand the particular histories of these suffixes.

Table 3 (below) illustrates most of the major strands of morphological variation in the dialectal suffix forms and will constitute the bulk of the data upon which the subsequent discussions will focus. The following should be considered a representative sampling, and are not exhaustive (Najdi from Ingham (1982, pp. 96–98; 2008, p. 328); Šan’āni from Rossi (1939, p. 4), Watson (2009, p. 110), Isaksson (1991, p. 127)); Mardin Arabic from Grigore (2007, p. 228); Levantine from Brustad and Zuniga (2019, p. 411).

Table 3. Pronominal Suffixes from Sample Arabic Dialects.

	Najdi	Dafār	Šan’āni	Mesopotamian Qəltu (Mardin)	Levantine
3ms	C-ih, C/V-h	C-eh, V-h	(Rossi) C-ah, V-h (Watson) -ih, V-h	-u, -hu	-o, -u
3fs	(a)ha	-ha	C-aha/V-hā	-a, -wa, -ya	-a, -ha
3mp	-(i)hum, -ham	-hum	(Rossi) -[o]hum (Watson) -uhum	-ən, -wən	-(h)on, -(h)in, -hum
3fp	-(i)hin, -(a)hin, -han	-hin	(Rossi) -[e]hin (Watson) -ahin	N/A	-hin
2ms	C-ik, C/V-k	C-uk, V-k	C-ak/V-k	-ək, -k	-ak, -ek
2fs	C/V-(i)ġ ~ š	C-iš, V-š	C-iš/V-š	-ki	-ik, -itš, -tši
2mp	-(i)kum, -kam	-kum	(Rossi) -[o]kum (Watson) -ukum	-kən	-kon, -kun, --kin, -kum, -kam, -kim
2fp	-(i)ġin, -ġan	-kin	(Rossi) -[e]kin (Watson) -akin	N/A	-kin, -tšin, -tšan
1cs	C-i/V-ya	-i	C-i/V-ya	-i, -ya	-i
1cp	-(i)na, -a(na)	-na	-[a]na	-na	-na

3.1. Third Person Masculine and Feminine Forms

Numerous and geographically widespread modern dialects attest 3ms C-u/C-o, and -V-. Regarding forms with -o(h), Cantineau (1939) argued that it is the result of *dār* + frozen accusative *a* + *hu*, with subsequent elision of the intervocalic laryngeal:

$$dār-a-hu > dāra-u > dār-o.$$

Jastrow (1991) assumes that *dār-u(h)*/*dār-o*, as well as *dār-a(h)* and *dār-ih* “idem”, were the result of the “freezing” of different case vowels in different dialects. In his scenario, *dār-u(h)*/*dār-o* attested the retention of nominative **-u*, whereas *dāra(h)* attested a frozen accusative **-a* and *dāri(h)* a frozen genitive **-i*. However, like Blau and Diem, Jastrow fails

to provide a scenario which accounts for the development of each. We are not told why the nominative case would be frozen only when followed by the 3ms suffix.

I agree that the 3ms suffix forms *-u(h)* ultimately derives from **u-hu*. As we saw above, explaining these forms involves a model that can explain the freezing of the case vowel before PN suffixes before the loss of word-final short vowels. I would suggest the following steps. From the reconstructed Proto-Arabic distribution, vowel harmony across morpheme boundaries developed, wherein $V^1\text{-}CV^2 > V^2\text{-}CV^2$.⁷

1. Proto-Arabic: **bayt-V^{case}n*; **al-bayt-V^{case}*; **bayt-V^{case}-hu*
2. Harmony Rule: **bayt-V^{case}n*; **al-bayt-V^{case}*; **bayt-u-hu*/**bayt-a-ka*/**bayt-i-ki*

Subsequently, final short vowels were lost.

3. Loss of final short vowels: **bayt-V^{case}n*; **al-bayt*; **bayt-uh*

And finally, the neutralization of phonemic contrast before *tanwīn* occurred:

4. Neutralization of contrast before *tanwīn*: *bayt-Vn*; *al-bayt*; *bayt-uh*

Steps 3 and 4 are not necessarily linear; they might have been simultaneous, or 4 might have preceded 3. Unlike Jastrow et al., however, I also suggest that 3ms *-o(h)* forms originate in **u-hu* as well, and not **a-hu*. The raising of **u* to *o* is widespread,⁸ and is likely behind the realization of the 3ms-*o(h)* forms as well.

Similar types of harmonization are attested in Arabic. For example, Sibawayh (1988, p. 173) mentions a type of assimilation that takes place in pause in which CVCC nouns insert a vowel to break up the final cluster where the second vowel harmonizes with the first:

'iblu "a camel", > *'ibil* "idem",
ḥulmu "a dream", > *ḥulum* "idem."

Blau's discussion of *naql* "transfer" represents a similar process of harmony, wherein a noun in pause inserts an epenthetic vowel that is the quality of the vowel which, when in non-pausal position, would mark its syntactic case:

bakrun > *bakur*

bakrin > *bakir*

bakran > *bakar*

This process⁹ is the same kind as is proposed here ($CV^1CV^2 > CV^2CV^2$). It is important, however, to emphasize the difference between the present proposal and Blau's and Diem's. Blau's appeal to *naql* to explain the pronominal suffix is predicated on the loss of final short vowels as the driver of the breakdown of the case system, which would eliminate the source of the transfer—the final vowels of the pronominal suffixes—he proposes. In the present proposal, a cross-morpheme harmony rule develops before the loss of final short vowels. Case on word-final nouns, as well as before *tanwīn*, likely continued for some time. A round (or several rounds) of word-final short vowel loss subsequently contributed to the loss of case, although it is likely that this process was gradual. The present proposal also differs with Diem's insofar as it proposes phonological rules to account for each stage, whereas Diem posits a period of randomness in the realization of etymological case vowels that is ad hoc.

Returning to the discussion of *dār-u(h)* and *dār-o(h)*, the *-u* and *-o* in *dār-u* and *dār-o* need not therefore be frozen nominatives as such; rather, they should be considered etymological case vowels that have harmonized with the vowel of the suffix. The point is, to my view, significant and thus bears repeating. While the argument here is that the vowels that precede the pronominal suffixes were etymologically case vowels, their development in harmonizing dialects (i.e., where the vowel matches the quality of the PN suffix) represents a stage at which the vowel harmonized regularly with the vowel quality of the following morpheme. It is therefore rather meaningless to speak of "frozen nominative" when referring to the *u*. Not only does this argument account for how the

vowels began to harmonize before the final short vowels of the suffixes were lost, but it also avoids the linguistically ad hoc free variation proposed by Diem.

In addition to forms with *-uh* or *-o(h)*, a minority of modern dialects, however, attest a form *-hu*, which quite possibly derives from **-hū*. In the Arabic of Mardin, for example, post-vocalic 3ms suffix is *-hu* rather than a lengthened vowel *-V:*, e.g., *abūhu* “his father” (Grigore 2007, p. 228; Talay 2011, p. 913). In nearby Sirt, as in most Qəltu dialects, short high vowels have completely merged to /ə/ in every phonetic environment (Talay 2011, p. 913). Thus Sirt *məfləs* (<**muflis*) “broke”, and pronominal suffixes *kən/ən* (<**kun/*kin*) and **hum/*hin* “your (cpl)/their (cpl)” (Grigore and Biṭuna 2012, p. 551). We would thus expect the 3ms suffix, derived from **u-hu* or **uh*, to be realized *-əh* as well. In some SW Yemeni dialects, *-hu* is ubiquitous whether the noun ends in a consonant or a vowel.

Comparative Semitic data suggest reconstructing Proto-Semitic 3ms suffix with a short vowel, **-su*. Additionally, both Hebrew and Aramaic forms suggest a Proto-Northwest Semitic **-hu* (Suchard 2020, p. 43). The Arabic dialectal forms ending in *-u(h)* and *-o(h)*, if the present reconstruction is correct, also suggest a short vowel **-hu*. If the Arabic dialectal *-hu* forms do indeed reflect etymological **-hū*, it could suggest that two by-forms existed at the Proto-Arabic node: **-hu* and **-hū*. There is evidence of long and short vowel variants of pronominal suffixes throughout Semitic. In CIAR, for example, pronominal suffixes following short case vowels are realized long, while those after long vowels were realized short: *bi-hī* “with it”, but *fi-hi* “in it” (Fischer 2002, p. 126). The opposite of this system is attested elsewhere, in, e.g., Aramaic and Gəʿəz, where suffixes with short vowels occur after short vowels, and suffixes with long vowels typically occur after long vowels, cf. Biblical Aramaic (Ezra 5:11) *ʿavdōhi* “his servants.” This distribution led Cantineau (Cantineau 1936/1937, vol. 2) to propose a sort of quantitative harmony, where the length of the suffix vowel harmonizes with the vowel preceding the suffix. A further explanation, offered in Hasselbach (2004), is that the suffixes with short forms are original, with the long forms the result of contamination with the independent form. It is quite possible that various analogies and developments have led to the distribution attested in the various languages (Suchard 2020, pp. 198–214). Whatever the case may be, the presence of long and short forms of the suffixes, originally tied to preceding syllable structure, is very possibly behind the different contemporary realizations of these suffixes across the dialects. This is undoubtedly the explanation for the common distribution of the 2fs suffix, cf. Damascene *C-ik*, but *V-ki*, where the vowel is always historically long (Lentin 2006, p. 548). The *u* forms in Qəltu dialects such as Mardin could represent a levelling of the long form to all contexts, with a subsequent loss of the laryngeal *h* after consonants, thus *C-u* but *V(:)-hu*.¹⁰

Another possibility is that *-hu* forms in some dialects reflect analogical protection of 3ms suffixes, based on analogy with the independent *hū*-based forms. Paradigm pressure can deter an otherwise regular phonetic shift from occurring. For example, the shift in West Semitic from Proto-Semitic **s¹ > h* when word-initial was blocked in some roots where other forms of the paradigm would have remained **s¹*. The verb *samiʿa*, “he heard”, should have become everywhere *hamiʿa* but, likely due to the imperfect forms, which would have been word-internal and thus not shifted (*yismaʿ* not **yihmaʿ*), the initial *s¹* was retained (Al-Jallad 2015).

Finally, in some instances an etymological **-hu* was possibly lengthened by analogical pressure from the long vowel in the 3fs suffix, **-hā*. Elsewhere in the pronominal suffix paradigm feminine forms form the basis for analogical adjustment to the masculine forms. In some cases, for example, the originally feminine *-i* vowel is levelled to masculine forms as well: e.g., Ġabal Fayfāʿ (Yemen) **-hum/*-hin* > *-him/-hin*. Elsewhere, the 3fp are extended via clipping of the feminine plural imperfective suffix **-nah*: e.g., Bani Abādīl (Yemen) 3mp *-him* but 3fp *-hinna* (Behnstedt 1987, p. 67). Masculine forms with a doubled nasal and final *a* were created in many dialects based on the feminine forms: e.g., Aṣ-Ṣalt (Jordan) *-hummu/-hinne* ((Palva 1994, p. 461); apud (Procházka 2014, p. 142)). It is very possible then that the long feminine vowel led to the lengthening of the masculine form in some cases as well.

Other than *-u(h)/-o(h)*, the 3ms suffix also frequently takes the form *-i(h)*, *-eh*, or *-ah*. In a number of Najdi dialects, the vowel preceding all suffixes, singular and plural, is the high vowel *i*, with the exception of the 3fs suffix, which is *a* (Ingham 2008, p. 328). In, e.g., the southern Hijaz and Tihama, (Procházka 1988, p. 192) reports *-ah*. Elsewhere, a fronted *-eh* is reported, for example in the dialect of Yašī' in Yemen (Isaksson 1991, p. 126).

Dialects that attest a 3ms suffix realized as *-a(h)*, including, e.g., some Najdi dialects (Ingham 1982, p. 98) and the southern Ḥ ijāz (Procházka 1988, p. 126), as well as some Sudanese dialects (Owens 1984), have traditionally been interpreted as reflecting the generalization of the accusative case followed by the loss of final short *u*: **a-hu > ah* (Cantineau 1939). Owens (2006, pp. 254–55) suggested instead that there were originally two 3ms suffix variants, one with a high vowel (usually *u*, *i* in Najdi), and one with a low vowel, the quality of which was specifically conditioned by the presence of emphatic consonants. As an example, he cites Eastern Libyan Arabic, where the distribution is supposedly attested. Owens then suggests that some dialects levelled the emphatic variant *ah*, while others the non-emphatic *ih*. As mentioned, Jastrow (1991, p. 170 et passim) argues that *-ah* and *-ih* reflect a frozen accusative and genitive vowel respectively.

There are reasons to doubt each of these explanations, however. To begin with, it is not clear whether dialects with *-eh* reflect either **-i* or **-a*. Relatedly, it is not always clear that contemporary *a* or *i* should be connected with historical **a* or **i*. This is especially true of, e.g., the Najdi dialects that constitute the classic examples of dialects with 3ms suffix *-ih* and *-ah*. In most Najdi dialects, historical high vowels **i* and **u* merged, with phonetic context determining whether the surface form was *i* or *u* (Ingham 2008, p. 327). It is also important to consider issues of transliteration and phonetic reality when discussing these forms. For example, as illustrated in Table 3 above, San'āni Arabic is sometimes transliterated with *-a*, as with Rossi, and other times with *-i*, as with Watson. It is likely that presumed etymological correspondence with historical vowels influences which is used.

I believe the most obvious explanation is that the forms *-ah*, *-ih*, and *-əh* are ultimately the result of a process wherein the distinction between *i* and *a* before *h* word-finally is neutralized. In non-rounded contexts, though transliterated as *ah*, the feminine ending is actually realized as basically a schwa; e.g., Najdi **-ah*, realized in neutral contexts as *ə*, i.e., *xirzəh* “bead”, (Ingham 2008, p. 327). Ingham elsewhere (Ingham 1986, p. 283) notes that 3ms suffix transliterated *-ih*, “his”, is often realized similarly to the *tā' marbūṭa*, namely as *-əh*. Such an explanation would also make sense out of the difference variation in transliteration found in a number of dialects, such as Ṣan'āni, where Rossi (1939, p. 4) gives *-ah*, but Watson (2009, p. 110) gives *-ih*. There is also evidence for such a situation in the Negev (Palestine) dialect of the 'Azazmih, where the reflex of the 3ms suffix is *-ih* when suffixed to most nouns, but *-ah* when suffixed to verbs whose stem vowel is low: *rās-ih* “his head”, but *ḡāb-ah* “he brought it” (Shawarba 2012, p. 193). This is paralleled in some cases by the realization of the feminine nominal marker **-ah*, which is *-it* in construct on most nouns but *-at* on verbs whose stem vowel is low: *nāg-it* “she-camel of”, but *nām-at* “she slept.” The same is true for Eastern Libyan itself, where Owens notes that Mitchell's descriptions (Mitchell 1952, 1960, apud Owens 1984, p. 93) give 3ms suffix *-ih*, whereas Owens transliterates it *-ah*. Owens' examples from Eastern Libyan Arabic are probably similar to other dialects, where the exact quality depends on the phonetic environment. Other dialects have been recorded with a form transliterated as *-eh* (e.g., Yašī'; (Isaksson 1991, p. 126)).

If correct, the traditional historical explanation for these forms, namely that they represent frozen genitive or accusative cases, becomes much less convincing. These variants cannot be explained through the same set of developments as the harmonizing group. Note that the variation attested in the vowel before *tanwīn* (Stokes 2020, pp. 654–58) mirrors that of the 3ms suffix. I suggest that the same neutralization attested before *tanwīn* occurred before PN suffixes in some dialects as well:

$$*bayt-V^{case}-hu/*bayt-V^{case}n > *bayt-V-hu/*bayt-Vn$$

Following the loss of final short vowels, the 3ms suffix would follow this non-phonemic vowel:

bēt-ih/bēt-in.

As we will see, many dialects with 3ms *-ih* or *-ah* attest 2ms *-ik* and 2fs *-iċ*. Thus, it is likely that vowels in this position merged, with a high vowel surface manifestation in every instance.¹¹ It is unclear what might have caused such a neutralization in this position. One possibility is that the change began in the construct, where the case vowel on the initial noun of the construct chain would be in a *waṣl* position, and thus potentially prone to reduction:

**baytu r-rağuli* “the man’s house” > **bayt Vr-rağuli*

Another possibility is that post-stress but non-word final vowels were reduced and merged:

**bayt-u-hu/bayt-un* > **bayt-V-hu/*bayt-Vn* but *al-bayt-u* and *baytu r-rağuli*

Subsequently, the non-phonemic vowel manifested as a high front vowel in most contexts, and final short vowels were lost. The construct perhaps continued inflecting for case, but the otherwise-ubiquitous absolute form eventually replaced it.

This can be summarized by the following steps:

1. Post-stress, non-word final neutralization of vowel contrast: **bayt-ən; bayt-ə-hu; but bayt-V^{case}*
2. Generalization of high vowel realization: **bayt-in; *bayt-i-hu; but *bayt-V^{case}*
3. Loss of final short vowels: *bayt-in; bayt-ih; bayt*

In a few dialects, such as Al-Mahābšeh, however, 3ms is *-eh* but 2ms is *-ak*, suggesting against a generalization to all contexts. In these cases, there might have been a generalization of *-a* before pronominal suffixes, with subsequent raising of *-ah*, as with the *tā’ marbūṭa -ah*, to *-eh* or *-ih*.

The 3fs suffix is in most dialects realized as *-a* or *-(V)ha*. The latter suggests an originally long final vowel **-hā*, but the former is ambiguous. Ahmad Al-Jallad (p.c.) has pointed out that there is some pre-Islamic Arabic epigraphic evidence for a short **-ha* by-form in line one of the Namārah inscription, which reads *mlk ‘l-’rb kl-h/malk ‘al-’arab koll-ah/“king of all ‘Arab”* (M.C.E Macdonald’s translation, from [Fiema et al. \(2016, pp. 405–6\)](#)). For either /kull-ha/ or /kull-hā/ we would expect ***kl-h’* in the Nabataean orthography. [Cantineau \(1936/1937, pp. 78, 182–83\)](#) argued based on data from the Syrian Šammarī and ‘Anaze Bedouin dialects, in which 3fs suffix was realized *C-ah* but post-vocalic *V-ha*, that the vowel was originally anceps (that is, both long and short), and that its length matched the length of the preceding case vowel. While this fits the data on which Cantineau was focusing, this does not fit many other dialects, which show only one form. So, while I agree with Cantineau that there were possibly two by-forms, one long and one short, of the 3fs suffix, based on realizations of the suffix in other Arabic dialects (see below), as well as CIAR data, it seems unlikely that we can posit a length harmonization for the ancestors of all dialects, and almost certainly not for Proto-Arabic.¹²

If we posit short and long by-forms for Proto-Arabic, then examples of 3fs suffix realized as *-ha* seem most likely explained as retentions of **-hā*. More complicated are forms such as *-ah*. If, as is maintained here, final short **-a* vowels were eventually lost in the ancestors of the dialects, then *-ah* forms are most likely explained as reflecting the same harmony across morpheme boundaries described above: **V^{case}-ha > a-ha > -ah*, with harmonization before loss of final short *a*. Alternatively, *-ah* forms could reflect an analogical restructuring of the 3fs suffix based on the 3ms forms: *-uh/-aha > -uh/-ah*. In that case, even these forms could ultimately have descended from an originally long final form **-hā*. From such a situation, whether due to loss of a final short **a*, or analogical restructuring, we can explain the distribution in a minority of dialects in which 3fs is *-eh/-ih*, as in the dialect of Ġiblah, Yemen (Table 4; data from [Isaksson 1991, p. 130](#)):

Table 4. Ġiblah Forms.

3ms	C-uh/C-h
3fs	-ih
2ms	C-ak/V-k
2fs	-ik
1cs	-i

The short *i* vowel of the 3fs suffix in this dialect could reflect one of two paths. One possibility is that it reflects a fronting of short *a* before word-final *-h* (as with *tā' marbūṭa* and 3ms suffixes noted above). Such a distribution is also attested in, e.g., Al-Ḥ ugarīyeh: 3ms C-oh/V-uh and 3fs C-eh/V-ih (Isaksson 1991, p. 132). Alternatively, the 3fs suffix in Ġiblah could be analogical with the 2fs suffix.

Forms like Levantine *-a*, which lack a final laryngeal, could possibly go back to either **-ha* or **-hā*. If they reflect an originally short vowel, then their development would mirror the *-ah* forms discussed above, with elision of the final laryngeal. More likely, in my view, is that they descend from long **-hā* and reflect the loss of the laryngeal *-h* following a vowelless consonant: *bayt-V-hā* > *bayt-hā* > *bayt-ā* > *bayt-a* “her house”. Several pieces of evidence support such a reconstruction. First, the 3fs suffix regularly patterns in terms of pre-suffixal vowel quality and insertion with suffixes that historically were heavy, namely the plural forms, in multiple dialects. That is, these suffixes are typically preceded by a vowel when the phonotactic patterns of the particular dialect dictate the insertion of an epenthetic vowel.

In Meccan Arabic, for example, epenthetic vowels are regularly *a*, which in the paradigm above is inserted before 3fs, 1cp, and 3rd and 2nd plural forms. In the dialect of Al-Maḥall (among others), the anaptyctic vowel harmonizes with the following vowel (see Table 5). The simplest explanation of this distribution is that the vowels before these suffixes originated as anaptyctic insertions, not case vowels, and are thus distinct from the vowels preceding the 3ms, 2ms/2fs, and 1cs suffixes.

Table 5. Patterns of **-CVV* and **CVC* suffixes (Arabian data from Isaksson (1991); Damascene from Brustad and Zuniga (2019, p. 411)).

	Mecca	Al-Maḥall	Damascene
3fs	-[a]ha	-[a]ha	-a
3mp/2mp	-[a]hum/[a]kum	-[o]hum/[o]kum	-on/-kon
3fp/2fp	N/A	-[a]han/[o]kun	N/A
1cp	-[a]na	-[a]na	-na

How can this apparent contradiction be resolved? First, the 3fs and 1cp suffixes, unlike 3ms, 2ms, and 2fs, are re-constructible with long **ā* vowels, 3fs **hā* and 1cp **nā*, which pattern with the heavy CVC syllables of the 3mp/3fp **-hum/*-hin* and 2mp/2fp **-kum/*-kin* suffixes. I argue that as case inflection broke down, the original short (formerly case-inflecting) vowel was syncopated before long (CVV) and heavy (CVC) syllables¹³:

**bayt-V-hā/hum/kum* > **bayt-hā/hum/kum*.

At some point later in their development, a disallowance for CCVV and CCVC syllables arose in many dialects, resulting in insertion of an anaptyctic vowel in the same slot. Elsewhere, in, e.g., the ancestors of Levantine-like dialects, no such disallowance developed, with either the retention of the laryngeal of the 3rd person forms (as in Ḍafār; see Table 3 above) or loss (as in Damascene).

The current proposal can thus be summed up by distinguishing between suffixes that consisted historically of light syllables (i.e., CV) and those that consisted of heavy ones (either CVV or CVC). Before suffixes consisting of light syllables, the final short vowel,

historically a case vowel, in some cases (e.g., Levantine) harmonized with the vowel of the suffixes, and in other cases (e.g., Najdi) lost contrast and manifested as a generalized high vowel.¹⁴ Before suffixes consisting of heavy syllables, there was an initial syncope of the short vowel.¹⁵ In many dialects, a subsequent phonetic rule disallowing strings of heavy syllables resulted in the insertion of an anaptyctic vowel, while in others no such insertion occurred. In dialects such as Damascene, the laryngeal was elided, but elsewhere no elision occurred (see Table 6 for reconstructions).

Table 6. Reconstruction of Meccan and Damascene.

	Meccan-Type	Damascene-Type
3fs/3mp	*bayt-V-hā/*bayt-V-hum > *bayt-hā/*bayt-hum > bayt-a-ha/bayt-a-hum	*bayt-V-hā/bayt-V-hun > *bayt-hā/*bayt-hum > bayt-a/bayt-on
2mp	bayt-V-kum > *bayt-kum > bayt-a-kum	*bayt-V-kun > *bayt-kun > bēt-kon

3.2. Second Person Masculine and Feminine Forms

The 2ms suffix is in many dialects realized *-ak* after consonants and *-k* after vowels. These are derivable from *V^{case}-ka > *a-ka (via the vowel harmony rule mentioned above) > *ak* (after loss of final short *a*), with a post-vocalic variant *-k*. A significant minority of dialects attest another vowel before the final velar stop, usually either high front *-ik*, but in a few instances high back *-ok*/*-uk*. In the case of 2ms *-ik*, the feminine form is almost always an affricated *-itš* or *-iċ*. I have already argued that 3ms *-ih* in many cases reflects a neutralization of vowel contrast in this position. The dialects with 3ms *-ih* in which 2ms and 2fs suffixes are *-ik* and *-iċ* suggest a general loss of contrast, with a high front vowel surface manifestation.

The 2ms suffix form realized as *-ok* or *-uk* has long baffled commentators (cf. Diem 1973, p. 42; Isaksson 1991, p. 127). In Dafār, for example, the paradigm is as follows (Table 7; data from Isaksson 1991, p. 127):

Table 7. Dafār singular forms.

3ms	C-eh/V-h
3fs	-ha
2ms	-uk
2fs	-iš

How can the 2ms *-uk* be explained? One possibility is that, initially, the pre-suffix vowel was universally high: *bayt-V^{case}-PN > *bayt-V^{high}-PN. Subsequently, harmonization of the 2fs to the similarly high suffix vowel occurred: *V^{high}-ki > *i-ki > *i-šī > *-iš*. Before the 2ms, however, the back high *u* was generalized, with no subsequent harmonization. It is curious that the 3ms and 2fs suffixes, both of which were followed etymologically by high vowels, apparently triggered a front high manifestation of the pre-suffix vowel, whereas the 2ms, with a low vowel, did not. Perhaps there was a pattern that led to this:

V-CV^{high} > V^{front}-CV^{high},
but
V-CV^{low} > V^{back}-CV^{low}

The feminine by-forms *-ik*/*-ki* require a bit of comment. Comparative evidence strongly favors a reconstruction of final short *i*, which we would not expect to remain in these dialects. Diem (1991, p. 301) reconstructs 2fs **-kī*. While this is a possible explanation of forms after long vowels with *-ki*, it cannot explain the nearly ubiquitous existence of *-ik* forms. Several scholars have suggested rather that *-ki* forms are due to contamination from the nominative pronoun *inti* and the imperfect verbal suffix, e.g., Jordanian *tuktubi* “you

(fsg) write" (<**tiktubi*) (attested also apparently in Gǝ'əz; (Hasselbach 2004, p. 10, n. 28; Al-Jallad 2014, p. 319)). While it is possible that 2fs patterned originally with 3ms and 3fs in having short and long by-forms, with some length assimilation to explain why *-ki* remains only after long vowels, I believe the lack of 2ms long forms argues in favor of analogy with the independent and imperfective forms.

If indeed the source of the *i* on 2fs-*ki* PN suffix forms is based on analogy with the imperfect, we must still explain its predominant distribution: suffixed to nouns that end in a long vowel but absent following nouns that end in a consonant. In, e.g., Damascene Arabic, the 2fs suffix is *-ik* after a C but *-ki* after a vowel: *kitāb-ik* "your (fs) book", but *abu-ki* "your (fs) father." In other words, if the final vowel of *-ki* was restored from the imperfect, why was it not generalized to all contexts? Its distribution is precisely the same distribution of the 3fs suffix, which is *-a* after a C but *-ha* after a vowel: *kitāb-a* "her book", but *abu-ha* "her father." There is, in other words, a gendered difference in the paradigm. I suggest that, in many dialects, this distinction led to the creation (or retention) of a 2fs suffix with final *-i*, which was available still on the independent and prefix verbal forms:

3ms -VV/2ms -VVK:: 3fs -*ha*: 2fs X > *ki*

The symmetry also holds for other dialects. For example, in Najdi, the 2fs suffix patterns completely with the 3ms and 2ms:

3ms C-*eh*, V-*h*/2ms C-*ik*, V-*k*/2fs C-*ic*, V-*é*

Alternately, in a small minority of dialects, the 2fs suffix patterns completely with the 3fs, against 2ms and 2fs, e.g., the Saudi dialect of Al-Maḥābšeh (Table 8).

Table 8. The 3rd and 2nd singular pronominal suffixes.

	After Consonant	After Vowel
3ms	C- <i>eh</i>	V- <i>h</i>
2ms	C- <i>ak</i>	V- <i>k</i>
3fs	C- <i>ha</i>	V- <i>ha</i>
2fs	C- <i>ki</i>	V- <i>ki</i>

We can therefore explain the 2ms and 2fs suffix forms via developments from originally short final vowels. As we have seen with the 3ms and 3fs (and the plural forms, obliquely), the vowels preceding the suffixes were originally case inflecting short vowels, which lost phonemic value and subsequently harmonized with the retained suffix vowels.

4. Suffix Vowel Length

The length of the final vowels in Proto-Semitic is notoriously difficult to determine because the reflexes of the forms differ across the attested Semitic languages.¹⁶ Various scholars have attempted different solutions to this puzzle. Brockelmann (1908/1913, p. 74 ff.) suggested what has been the majority opinion of the past century, namely that the vowels were anceps, meaning that they could be long or short. Brockelmann's own intuition was that they were originally long, but that some were shortened because they were unstressed (ibid.).¹⁷ Blau (1981, p. 63 ff.) held the opposite scenario, namely that the suffix vowels were originally short, but that paradigm pressure, namely the preservation of gender distinctions on suffixes, led speakers to retain them in languages like Hebrew, where short vowels were otherwise lost.

Despite the difficulties in reconstructing their length in Proto-Semitic, I follow Hasselbach (2004); Al-Jallad (2014) in analyzing most of the forms as originally short in Proto-Arabic. The reasons, discussed above, concern the consistency with which the dialects agree with CIAR. For example, virtually all modern dialects lack the final vowel on 2ms *-*ka*, which seems best interpreted as the loss of final short **a*. It is necessary, and indeed methodologically preferable, to allow for multiple rounds of final-vowel reduction and loss when modeling the development of the dialects from their ancestors. We should

not constrain ourselves to solving the loss of, say, the word-final **-a* of the 2ms suffix in the same stage as the loss of final **-a* on the 3ms prefix verb; rather, we should probably factor in multiple rounds of reduction and loss, depending on dialect. The main difference between my reconstructions and those offered by Hasselbach concern the 3fs suffix, which I reconstruct long for Proto-Arabic, based on the form in CLAr, as well as its almost total retention in the modern dialects. Exceptions to these forms are, I believe, explicable by appeal to analogies (on which, see above). However, the Syrian Bedouin data gathered and analyzed by Cantineau suggest the possibility that, in the ancestors of some dialects, and perhaps even Proto-Arabic, a short by-form for 3fs suffix, and a long by-form for the 3ms suffix, based on length assimilation and/or polarization, was created. Table 9 presents proposed reconstructions of Proto-Arabic PN suffixes.

Table 9. Reconstructed Proto-Arabic Pronominal Suffixes.

3ms	*-hu, *-hū (?)	3mp	*-hum(ū) ¹⁸
3fs	*-ha (?), *-hā	3fp	*-hīn
2ms	*-ka	2mp	*-kum(ū)
2fs	*-ki	2fp	*-kīn
1cs	*C-ī/*V-ya	1cp	*-nā

5. Conclusions

In this paper I have addressed two long-standing issues in the historical study of the modern Arabic dialects. First, I argued in favor of the traditional interpretation of the vowels before the pronominal suffixes of the 3ms, 2ms, and 2fs as originating in etymological case vowels. I provided two different models to account for the modern realizations in the two major attested paradigms. Specifically, I posited two sets of rules. To account for dialects in which pre-suffixal vowels harmonized with the vowels of the pronominal suffixes, I proposed a vowel harmony rule which operated across morpheme boundaries. In order to explain forms in which a single vowel occurs prior to each suffix, I proposed a development by which post-stress vowels were neutralized, leading to a generic vowel (usually front high) surface manifestation. The remaining variants were explained within these models. I contrasted these scenarios with those offered by Blau and Diem, detailing how the arguments here account for variables unaccounted for in these previous proposals.

The second topic examined here is that of the historical length of several pronominal suffixes; specifically, the length of the 3ms, 3fs, 2ms, and 2fs suffixes was examined. I concluded that most 3ms forms can be derived from a historical short **-hu*, although a long by-form possibly existed at the Proto-Arabic node. I further argued that the 2ms and 2fs suffixes are derivable from historically short forms, **-ka* and **-ki*, and that exceptions can be explained via analogies that are readily identifiable. Finally, I argued that the 3fs suffix was in the vast majority of cases derivable from a historically long **-hā*. Furthermore, I argued that the vowels before the 3fs suffixes patterned with those of the 3rd and 2nd plural forms over against the other singular forms, which I argued is due to the fact that the etymological case vowels were syncopated at some point before suffixes of the syllable shape CVV and CVC, with the subsequent insertion of an epenthetic vowel. These vowels behave differently from those of the singular forms because they represent a separate stage of development in the dialects.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

Notes

- 1 Most who posit a nominal case system for Proto-Arabic reconstruct it essentially as it is found in CIAr. For our purposes, the important point is that originally a pronominal suffix would be suffixed to a noun which ended in a case vowel: N + V^{case} + CV(C). So, for example: **kalb-u-hu* “his dog (nom)”/**kalb-a-hu* “his dog (acc)”/**kalb-i-hu* “his dog (gen).”
- 2 Diem (ibid.) argues that the various realizations of DT, i.e., *un*, *in*, and *an*, reflect frozen forms of the three primary Arabic cases. For an alternative proposal, see Stokes (2020).
- 3 That is, other than his distinction between ‘nomadic’ and non-nomadic dialects, for which I can find no good evidence in the realizations of the pronominal suffixes. It seems his distinction is instead based here entirely on the DT. However, DT is also present in sedentary dialects.
- 4 Blau’s main objection to Diem’s analysis, with which he interacts directly, is that redundancy is very common in linguistic systems and cannot therefore be considered a very strong factor on its own (Blau 2006, p. 86). Rather, Blau posits that pausal forms, where *tanwīn* and case vowels were regularly lost, began to intrude into non-pausal positions as well, i.e., he argues for an analogical extension of pausal Ø-marked forms into non-pausal position. Additionally, however, Blau seems to also hold that some sort of final vowel elision took place, perhaps as a result of this pausal intrusion, but it is not clear to me how he envisions the relationships between these two processes.
- 5 It is puzzling that Owens recognizes implicitly in his reconstructions the loss of final short vowels, since this is the typical explanation for the loss of case vowels, which he finds unconvincing (ibid., pp. 2–8 *et passim*). Since he allows for the loss of these final vowels, which is implicit in his suggested proto-forms and necessary to account for the data, it is not clear why that loss could not account for the absence of case in the dialects. Thus, for Owens, final vowel loss seems a very random, haphazard process.
- 6 This point also argues against some kind of metathesis, i.e., *qalb-hu* > *qalb-uh*. First, we would expect that such a metathesis would be motivated by the same impetus as epenthesis, namely phonotactic considerations. My argument in this section is that the epenthetic argument does not account for the distribution of these vowels in any dialect. A further argument against the metathesis argument is the fact that the metathesis would have to have occurred only with pronominal suffixes. That is, if metathesis occurred in the same phonetic environment (i.e., -CCV), we would expect the same to affect suffix conjugation suffixes, such as *darasta* “you (msg) wrote” and *darastu* “I wrote.” That these forms do not have dialectal forms ***darasat* or ***darasut*, but rather just *darasit* (< *darast* <* *darata/darastu*) “I/you wrote,” suggests that this did not occur there.
- 7 This development would have presumably affected nouns in construct when the following noun lacked the definite article: **ibnu malikin* “a king’s son,” > *ibna malikin* “idem.” While the topic is slightly outside the scope of this paper, it seems likely that changes in both the definite and indefinite construct syntagms played a role in the reduction of case inflection.
- 8 It is already attested in the pre-Islamic period in, e.g., the corpus of Graeco-Arabica, transliterations of Arabic into the Greek script from the city of Petra in modern-day Jordan (Al-Jallad 2017, pp. 145–46). The same is true in many modern dialects. For Levantine varieties, see Brustad and Zuniga (2019, pp. 407–8).
- 9 In his chapter on pausal forms, called *bāb al-waqf* (Sibawayh 1988, p. 173 ff.), Sibawayh discusses these forms and the phenomenon to which Blau refers, but without using the term *naql*. Sibawayh typically reserves the term *naql* for the transfer of the vowel on a *hamza* to the preceding consonant with the subsequent loss of the *hamza*: *al-ḥimar* “the red one,” > *laḥmar* “idem.” I thank an anonymous reviewer for pointing this difference out to me.
- 10 Perhaps the CIAr reverse length polarization in the 3ms suffixes, wherein *V-hū* but *VV-hu*, would represent an innovation prodded by the existence of these by-forms.
- 11 Alternatively, we might posit a generalization of a single case, perhaps nominative, before pronominal suffixes. After the merger of **u* and **i*, the high vowel would be generalized.
- 12 In varieties described by the grammarians, which subsequently became standardized for CIAr pronunciation, a reverse length polarity developed for the 3ms suffix (*V-hū*, *VV-hu*), whereas 3fs was ubiquitously long. The dialects cannot, of course, have descended from CIAr since CIAr represents an amalgam of varieties and features. However, the reality of the reverse polarity elsewhere suggests that by-forms existed (or, to use Cantineau’s term, should be reconstructed as anceps), but were distributed differently across dialects. In some, a length distinction became allomorphic; elsewhere, one form was generalized to all contexts.
- 13 This phenomenon is attested and remarked upon by the Quranic reading commentator Al-Farrā’ in his *Luḡāt al-Qur’ān* (Al-Farrā’ 2014, p. 30). He notes specifically that the case vowels -*u* (nom) and -*i* (gen) are syncopeated whenever pronominal suffix with a heavy syllable follows: *bayt-kum* (nom/gen) but *bayta-kum* (Acc). Interestingly, traces of this same phenomenon can be seen in the Quranic reading traditions of Abū ‘ Amr and Ibn Kaṭīr. For example, in Q2:54 the reciters of Abū ‘ Amr read *‘ilā bāri’ ikum* “to your (pl) creator” as *‘ilā bāri’-kum*, without the genitive *i* due to the following heavy syllable. I thank an anonymous reviewer for bringing these examples to my attention.

- 14 Again, it is unclear whether in these dialects either the nominative or genitive case was levelled to all contexts before pronominal suffixes, or rather there was a general loss of phonemicity as case inflection broke down, with the resulting high vowel a result of the general preference in these dialects for open, unstressed syllables to contain high vowels.
- 15 I thank an anonymous reviewer for pointing out that this is similar to a change from Pre-Classical to Classical Latin, in which a light syllable is deleted between two heavy syllable: e.g., *lāridī* > *lardī* “bacon”. Armin Mester (1994) calls this process “prosodic trapping,” in which the light syllable is not metrically footed.
- 16 For example, the 2ms suffix reflects **-ka* in Arabic and Gəʿəz, but is present in some dialects of Hebrew, in which short final **a* should have been lost. Further, 2fs reflects **-ki* in Hebrew and Arabic, but in Gəʿəz, where **u* and **i* merged to ə, the attested form is *-ki* < **kī*.
- 17 This position was also suggested to me by J. Huehnergard (p.c.), who notes the cross-linguistic tendency for long vowels to be realized short when unstressed (as in, e.g., Latin).
- 18 I have not encountered solid evidence for the retention of the so-called long pronominal forms among the modern dialects. Sibawayh discusses the plural forms in (*bābu mātuksaru fī-hi l-hā'u llatī hiya 'alāmatu l-'idmār*) (Sibawayh 1988, p. 195). For an in-depth discussion of these long forms, see Van Putten and Sidky (Forthcoming).

References

- Al-Farrā', 'Abū 'Alī. 2014. *Kitāb Fih Luḡāt Al-Qur'ān*. Edited by Čābir b. 'Abd Allāh al-Sarī'. Unpublished.
- Al-Jallad, Ahmad. 2014. Final short vowels in Ge'ez, Hebrew 'atta, and the Anceps Paradox. *JSS* 59: 315–27.
- Al-Jallad, Ahmad. 2015. Yusap'il or Yuhap'il, that is the question—Two solutions to sound change s¹ > h in West Semitic. *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 165: 27–39.
- Al-Jallad, Ahmad. 2017. Graeco-Arabica I: The Southern Levant. In *Arabic in Context: Celebrating 400 Years of Arabic at Leiden University*. Edited by Ahmad Al-Jallad. Leiden: Brill, pp. 99–186.
- Al-Jallad, Ahmad. 2018. The earliest stages of Arabic and its linguistic classification. In *The Routledge Handbook of Arabic Linguistics*. Edited by Elabbas Benmamoun and Reem Bassiouney. New York: Routledge, pp. 315–31.
- Al-Wer, Enam. 2007. Jordanian Arabic (Amman). In *Encyclopedia of Arabic Language and Linguistics, Vol. II*. Edited by Kees Versteegh. Leiden: Brill, pp. 505–17.
- Behnstedt, Peter. 1987. *Die Dialekte der Gegend von Ṣa' dah (Nord-Jemen)*. Wiesbaden: O. Harrassowitz.
- Birkeland, Harris. 1952. *Growth and Structure of the Egyptian Arabic Dialect*. Oslo: I kommisjon Hos Jacob Dybwad.
- Blau, Joshua. 1972. On the problem of the synthetic character of Classical Arabic as against Judaeo-Arabic (Middle Arabic). *Jewish Quarterly Review* 63: 29–38. [CrossRef]
- Blau, Joshua. 1981. *The Emergence and Linguistic Background of Judaeo-Arabic: A Study of the Origins of Middle Arabic*, 2nd ed. Jerusalem: Ben-Zvi Institute.
- Blau, Joshua. 2002. *A Handbook of Early Middle Arabic*. Jerusalem: Hebrew University.
- Blau, Joshua. 2006. Some reflections on the disappearance of cases in Arabic. In *Loquentes Linguis: Linguistic and Oriental Studies in Honour of Fabrizio A. Pennacchiotti*. Edited by Pier Gorgio Borbone, Alessandro Mengozzi and Mauro Tosco. Wiesbaden: Verlag Harrassowitz, pp. 79–90.
- Brockelmann, Carl. 1908/1913. *Grundriss der Vergleichenden Grammatik der Semitischen Sprachen*. Reprint 1961. Hildesheim: Georg Olms Verlagsbuchhandlung, vol. II.
- Brustad, Kristen, and Emilie Zuniga. 2019. Levantine Arabic. In *The Semitic Languages*, 2nd ed. Edited by John Huehnergard and Na'ama Pat-El. New York: Routledge, pp. 403–22.
- Butts, Aaron Michael. 2019. Gəʿəz (Classical Ethiopic). In *The Semitic Languages*, 2nd ed. Edited by John Huehnergard and Na'ama Pat-El. New York: Routledge, pp. 117–44.
- Cantineau, Jean. 1936/1937. Études sur quelques parlers de nomades arabes d'Orient. In *Annales de l'Institut d'Études Orientales*. Paris: Librairie Larose.
- Cantineau, Jean. 1939. Le pronom suffixe de 3^e personne singulier masculine en arabe Classique et dans les parlers arabes modernes. *BSL XL* 1: 89–97.
- Diem, Werner. 1973. *Skizzen Jemenitischer Dialekte*. Beirut Texte und Studien, 13. Beirut and Baden-Baden: Nomos Verlagsgesellschaft.
- Diem, Werner. 1991. Vom Alterabischen zum Neuarabischen Ein neuer Ansatz. In *Semitic Studies: In Honor of Wolf Leslau on the Occasion of His Eighty-Fifth Birthday, 14 November 1991*. Edited by Wolf Leslau, Alan Kaye and Thomas Leiper Kane. Wiesbaden: Harrassowitz, vol. 1, pp. 297–308.
- Fiema, Zbigniew T., Ahmad Al-Jallad, Michael Christopher Macdonald, and Laïla Nehmé. 2016. Provincia Arabia: Nabataea, the Emergence of Arabic as a Written Language, and Graeco-Arabica. In *Arabs and Empires before Islam*. Edited by Greg Fisher. Oxford: OUP, pp. 373–433.
- Fischer, Wolfdietrich. 2002. *A Grammar of Classical Arabic*, 3rd rev. ed. Translated from the German by Jonathan Rodgers. New Haven: Yale University Press.
- Fischer, Wolfdietrich, and Otto Jastrow, eds. 1980. *Handbuch der Arabischen Dialekte*. Wiesbaden: O. Harrassowitz.
- Grigore, George. 2007. *L'arabe parlé à Mardin: Monographie d'un Parler arabe Périphérique*. Piscataway: Gorgias Press.

- Grigore, George, and Gabriel Bişuna. 2012. Common Features of North Mesopotamian Arabic Dialects Spoken in Turkey (Sırnak, Mardin, Siirt). In *Bilim Düşünce ve Sanatta Cizre: (Uluslararası Bilim Düşünce ve Sanatta Cizre Sempozyumu Bildirileri)*. Edited by Mehmet Nesim Doru. Mardin: Mardin Artuklu Üniversitesi Yayınları, pp. 545–54.
- Hasselbach, Rebecca. 2004. Final Vowels of Pronominal Suffixes and Independent Personal Pronouns in Semitic. *JSS* 49: 1–20. [\[CrossRef\]](#)
- Hasselbach, Rebecca. 2014. Agreement and the Development of Gender in Semitic (Part 1+2). *ZDMG* 164: 33–64.
- Huehnergard, John. 2019. Proto-Semitic. In *The Semitic Languages*, 2nd ed. Edited by John Huehnergard. New York: Routledge, pp. 49–79.
- Ingham, Bruce. 1982. Notes on the Dialect of the Ḍafir of North-Eastern Arabia. *Bulletin of the School of Oriental and African Studies* 45: 245–59. [\[CrossRef\]](#)
- Ingham, Bruce. 1986. Notes on the Āl-Murra of eastern and southern Arabia. *Bulletin of the School of Oriental and African Studies* 49: 271–91. [\[CrossRef\]](#)
- Ingham, Bruce. 2008. Najdi Arabic. In *The Encyclopedia of Arabic Language and Linguistics*, Vol. III. Edited by Kees Versteegh. Leiden: Brill, pp. 326–34.
- Isaksson, Bo. 1991. The personal markers in modern Arabic dialects of the Arabian peninsula. *Orientalia Suecana* 40: 117–45.
- Jastrow, Otto. 1991. ‘Une question embarrassante’—Jean Cantineau über das Pronominalsuffix 3. Sg. m. in den arabischen Dialekten. In *Festgabe für Hans-Rudolf Singer: Zum 65. Geburtstag am 6. April 1990, Überreicht von Seinen Freunden und Kollegen*. Edited by Martin Forstner. Paris: Peter Lang, pp. 167–74.
- Lentin, Jérôme. 2006. Damascus Arabic. In *The Encyclopedia of Arabic Language and Linguistics*, Vol. I. Edited by Kees Versteegh. Leiden: Brill, pp. 546–55.
- Mester, R. Armin. 1994. The Quantitative Trochee in Latin. *Natural Language & Linguistic Theory* 12: 1–61.
- Mitchell, Terence F. 1952. The Active Participle in an Arabic Dialect of Cyrenaica. *Bulletin of the School of Oriental and African Studies* 14: 11–33. [\[CrossRef\]](#)
- Mitchell, Terence F. 1960. Syllabification and Prominence in Arabic. *Bulletin of the School of Oriental and African Studies* 23: 369–89.
- Owens, Jonathan. 1984. *A Short Reference Grammar of Eastern Libyan Arabic*. Wiesbaden: O. Harrassowitz.
- Owens, Jonathan. 2006. *A Linguistic History of Arabic*. Oxford: OUP.
- Owens, Jonathan. 2018. Where multiple pathways lead: A reply to Ahmad Al-Jallad and Marijn van Putten. *Abhandlungen für die Kunde des Morgenlandes* 113: 95–162.
- Palva, Heikki. 1994. Bedouin and sedentary elements in the dialect of es-Salt. In *Actes des Premières Journées Internationales de Dialectologie Arabe de Paris*. Edited by Dominique Caubet and Martine Vanhove. Paris: INALCO, pp. 459–69.
- Procházka, Theodor. 1988. *Saudi Arabian Dialects*. London: Kegan Paul International.
- Procházka, Stephen. 2014. Feminine and Masculine Plural Pronouns in Modern Arabic Dialects. In *From Tur Abdin to Hadramawt: Semitic Studies Festschrift in Honour of Bo Isaksson on the Occasion of His Retirement*. Edited by Tal Davidovich, Ablahad Lahdo and Torkel Lindquist. Wiesbaden: O. Harrassowitz, pp. 129–48.
- Retsö, Jan. 1994. ‘i’ rāb in the Forebears of Modern Arabic Dialects. In *Actes des Premières Journées Internationales de Dialectologie Arabe de Paris*. Edited by Dominique Caubet and Martine Vanhove. Paris: INALCO, pp. 333–42.
- Rossi, Ettore. 1939. *L’arabo Parlato a Ş an ‘ā’*. Rome: Istituto per l’Oriente.
- Shawarba, Musa. 2012. *A Grammar of Negev Arabic: Comparative Studies, Texts and Glossary in the Bedouin Dialect of the ‘Azazmih Tribe*. Wiesbaden: Harrassowitz.
- Sibawayh, ‘Abū Bişr ‘Uṭmān. 1988. *Kitāb Sibawayh, Volume IV*. Edited by ‘Abd al-Salām Muḥammad Hārūn. Cairo: Maktabat al-Ḥāniḡī.
- Stokes, Phillip W. 2020. A Fresh Analysis of the Origin and Diachronic Development of ‘Dialectal *Tanwīn*’ in Arabic. *JAOS* 140: 637–64. [\[CrossRef\]](#)
- Suchard, Benjamin. 2020. *The Development of the Biblical Hebrew Vowels*. Leiden: Brill.
- Talay, Shabo. 2011. Arabic dialects of Mesopotamia. In *The Semitic Languages: An international handbook*. Edited by Stefan Weninger. Berlin and Boston: De Gruyter Mouton, pp. 909–20.
- Van Putten, Marijn, and Phillip W. Stokes. 2018. Case in the Qur’ānic Consonantal Text. *WZKM* 108: 143–79.
- Van Putten, Marijn, and Hythem Sidky. Forthcoming. Pronominal Variation in Arabic among the Grammarians, Quranic Reading Traditions, and Manuscripts. In *Formal Models in the History of Arabic Grammatical and Linguistic Tradition*. special issue of *Language and History* 65: 1. Edited by Raoul Villano.
- Watson, Janet C. E. 2007. Syllabification Patterns in Arabic Dialects: Long Segments and Mora Sharing. *Phonology* 24: 335–56. [\[CrossRef\]](#)
- Watson, Janet C. E. 2009. Ş an ‘āni Arabic. In *The Encyclopedia of Arabic Language and Linguistics*, Vol. IV. Edited by Kees Versteegh. Leiden: Brill, pp. 106–14.
- Woidich, Manfred. 2006. Cairo Arabic. In *The Encyclopedia of Arabic Language and Linguistics*, Vol. I. Edited by Kees Versteegh. Leiden: Brill, pp. 323–33.

Article

Towards a Dialect History of the Baggara Belt

Stefano Manfredi ^{1,*} and Caroline Roset ²

¹ SeDyL (CNRS UMR8202, INALCO, IRD UR135), Campus CNRS de Villejuif, 7 Rue Guy-Môquet, 94801 Villejuif, France

² Department of Arabic Language and Culture, Faculty of Humanities, University of Amsterdam, 1012 WX Amsterdam, The Netherlands; c.j.rosset@uva.nl

* Correspondence: stefano.manfredi@cnrs.fr

Abstract: The Baggara Belt constitutes the southernmost periphery of the Arabic-speaking world. It stretches over 2500 km from Nigeria to Sudan and it is largely inhabited by Arab semi-nomadic cattle herders. Despite its common sociohistorical background, the ethnography of Baggara nomads is complex, being the result of a long series of longitudinal migrations and contacts with different ethnolinguistic groups. Thanks to a number of comparative works, there is broad agreement on the inclusion of Baggara dialects within West Sudanic Arabic. However, little or nothing is known of the internal classification of Baggara Arabic. This paper seeks to provide a comparative overview of Baggara Arabic and to explain dialect convergences and divergences within the Baggara Belt in light of both internally and externally motivated changes. By providing a qualitative analysis of selected phonological, morphosyntactic, and lexical features, this study demonstrates that there is no overlapping between the ethnic and dialect borders of the Baggara Belt. Furthermore, it is argued that contact phenomena affecting Baggara Arabic cannot be reduced to a single substrate language, as these are rather induced by areal diffusion and language attrition. These elements support the hypothesis of a gradual process of Baggarization rather than a sudden ethnolinguistic hybridization between Arab and Fulani agropastoralist groups. Over and above, the paper aims at contributing to the debate on the internal classification of Sudanic Arabic by refining the isoglosses commonly adopted for the identification of a West Sudanic dialect subtype.

Citation: Manfredi, Stefano, and Caroline Roset. 2021. Towards a Dialect History of the Baggara Belt. *Languages* 6: 146. <https://doi.org/10.3390/languages6030146>

Academic Editors: Simone Bettega and Roberta Morano

Received: 26 June 2021

Accepted: 23 August 2021

Published: 30 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: Arabic; Sudanic Arabic; Baggara; comparative dialectology

1. Introduction

In his early account of the Shuwa Dialect of Bornu, Nigeria and of the Region of Lake Chad, the British colonial governor G. J. Lethem (1920, p. xi) stated that Nigerian (i.e., Shuwa) Arabic “is part of the Arabic dialects of the Sudan, of which Shuwa is the westernmost”. By using these words, Lethem was most likely the first to recognize ‘the dialects of the Sudan’ as a homogenous dialect group, distinct from other Arabic varieties (i.e., Maghrebi, Levantine, Gulf, etc.). Later on, Blanc (1971) adopts the label ‘Sudanic Arabic’ for referring to the dialect continuum running across the vast region delimited by Lake Chad (Nigeria) in the west, by the Red Sea coast (Sudan) in the east, by Lake Nasser (Egypt) in the north, and by the Nuba Mountains (Sudan) in the South. According to Blanc (1971, p. 503) Sudanic Arabic “does not fit too neatly into either the East-West or the nomadic-sedentary dichotomy, though on the whole it is more Eastern than Western and more nomadic than sedentary”. In spite of the vagueness of the dialect portrait drawn by Blanc, both historical and linguistic data unmistakably point out that Sudanic Arabic mainly (but not exclusively) emerged following the penetration of Arabic-speaking nomadic groups from Upper Egypt into Sudan in the first half of the 14th century (cf. 2, see also Thomas A. Leddy-Cecere this special issue). During the last decades, several studies (Blanc 1971; Kaye 1976; Owens 1985, 1993b; Roth-Laly 1994a; Manfredi 2012) have identified a number of isoglosses for pinpointing the Sudanic area within the wider Arabic-speaking world. These Pan-Sudanic features include:

- The phonological reflexes *q > g (k), *t > ṭ (t); *ḍ > ḍ̣ (d) (e.g., *baqar > baqar ‘cows’; *ṭawb > ṭōb ‘women dress’; *kaḍḍāb > kaḍḍāb ‘liar’);
- The presence of the phoneme č (e.g., čalbaq ‘stick.3SG.M’);
- The partial preservation of an etymological a in prepausal position (e.g., gul-t(a) ‘say-1SG/2SG.M’, dagga ‘hit.3SG.M’);
- The generalization of the preformative a- in imperative forms (e.g., a-gīf ‘IMP-stop’);
- The presence of postnominal demonstrative determiners lacking the etymological element *ha- (e.g., da ‘PROX.SG.M’);
- The large use of the lexemes abū ‘father’ and umm/aḥḥim ‘mother’ as compound formatives (e.g., abū gurūn ‘rhino’, aḥḥim garn ‘ground hornbill’);
- The presence of ideophones as a lexical category (aḥmar čall red IDPH ‘bright red’).

Concerning the internal classification of Sudanic dialects, [Kaye \(1976\)](#) examines selected phonological and morphological features of Chadian and Sudanese Arabic in light of Ferguson’s monogenetic theory of dialect emergence and stresses the relative homogeneity of these varieties. [Owens \(1993a\)](#), in contrast, proposes a patchwork approach that does not reveal any eco-linguistic difference between ‘Bedouin’ and ‘Sedentary’ dialects, while showing clear affinities between Sudanic and Upper Egypt Arabic. [Roth-Laly \(1994a, 1994b\)](#), on her part, stresses the generalization of traditional ‘Bedouin’ features in Sudanic Arabic and identifies new local isoglosses for opposing ‘Bedouin’ to ‘Sedentary’ dialects across the Sudanic area. Regardless of their different approaches, the previously mentioned scholars agree in identifying two main dialect sub-types within Sudanic Arabic: West Sudanic Arabic (henceforth ‘WSA’) and East Sudanic Arabic (henceforth ‘ESA’). WSA encompasses the dialects spoken in Nigeria ([Lethem 1920](#); [Owens 1993a, 1993b](#)), Cameroon ([Owens 1993a](#)), and Chad ([Carbou 1913](#); [Roth-Laly 1972, 1979](#); [Hagège 1973](#); [Decobert 1985](#); [Zeltner and Tourneux 1986](#); [Abu-Absi 1995](#); [Jullien de Pommerol 1999a, 1999b](#)) as well as in the western Sudanese provinces of Darfur and Kordofan ([Manfredi 2010, 2012, 2013](#); [Roset 2018](#)). ESA covers the remaining parts of the Sudanic dialect area (i.e., the central and eastern part of Sudan) and it includes the koine of the capital Khartoum (i.e., Khartoum or Sudanese Arabic, [Bergman 2002](#); [Dickins 2011](#)) and the rural dialects spoken in the Gezira and Butana regions ([Reichmuth 1983](#)). Even if this geographical split between WSA and ESA is supported by strong linguistic evidence, it hardly reflects the high degree of diatopic and eco-linguistic variation affecting Sudanic Arabic. In this regard, it is worth remarking that [Hillelson \(1925, p. xv\)](#) distinguishes at least four distinct dialect subtypes in Sudan, counting:

“The speech of the Northern Sudan, including Berber Province and the Arabic-speaking parts of Dongola; the speech of the Central Sudan, including Omdurman, the Gezira, and the country to the east of the Blue Nile; the idiom of the Western Sudan, embracing the White Nile, Kordofan and Darfur; and the dialect of the Baggara tribes. It should further be noted that the speech of nomad Arabs everywhere differs from that of the settled population.”

Hillelson’s introduction to his Sudanese Arabic dictionary therefore warns about the specificity of the “speech of nomad Arabs” in Sudan of which the semi-nomadic Baggara are part. In light of the above, this paper seeks to answer the question of to what extent Baggara Arabic constitutes a homogenous dialect sub-type within West Sudanic Arabic. For this aim, we provide a comparative overview of five different Baggara dialects in order to assess their degree of structural proximity and to explore the sociohistorical factors underlying the diffusion of linguistic innovations across the Baggara Belt.

A preliminary version of this paper was presented at the 47th North Atlantic Conference on Afroasiatic Linguistics, INALCO, Paris, 24–25 June 2019. The paper is organized as follows. In Section 2, we offer a sociohistorical and linguistic introduction to the Baggara Belt. Section 3 briefly presents the data and the sources used for our comparative analysis. In Section 4, we explore the diatopic variation affecting a number of phonological, morphosyntactic, and lexical features in Baggara Arabic, while trying to reconstruct both internally and externally motivated diachronic changes. Section 5 finally attempts at re-

constructing the dialect history of the Baggara Belt and provides some new hints on the internal classification of West Sudanic Arabic.

2. The Baggara Belt: Sociohistorical and Linguistic Background

MacMichael (1922, p. 271) argues that “Baggara means no more than cattlemen”. Accordingly, the term Baggara (from the agentive noun PL *baggāra* ‘cattlemen’, SG.M/F *baggāri/baggāriye*) has neither ethnic nor genealogical pertinence, as it rather stresses the specificity of an agro-pastoral system of production based on cattle herding and sorghum cropping (Cunnison 1966, p. 10; Teitelbaum, 1984; Braukämper 1993, p. 14; Manfredi 2010, p. 10). There is a broad agreement on the fact that the center of origin of the Baggara tribes is to be found in present-day Chad. Nevertheless, two contrasting hypotheses have been put forward to trace back the way by which Arab nomads reached Chad. On the one hand, Carbou (1913, p. 4) and Henderson (1939, p. 52) allege that Arab nomadic groups entered Chad via the Fezzan area in Libya. In this perspective, the Baggara should be seen as an offshoot of the Arab groups that pushed southwards from Maghreb to central Africa following the Hilalian invasion in the 11th century. Even if this hypothesis is corroborated by Baggara oral traditions referring to Abu Zayd al-Hilali (Manfredi 2010, p. 12), there are neither historical nor linguistic arguments supporting the suggestion of a Maghrebi origin of Baggara groups. On the other hand, MacMichael (1922, p. 275) affirms that:

“On the migration of these Arabs from the east there cannot be the least doubt. They advanced gradually through the Negroland. [...] Their dialect is quite different from the Maghrebi, while in many respects it still preserves the purity and the eloquence of the language of Hijaz.”

Accordingly, the Baggara would have split apart from the Juhayna groups that penetrated Sudan from Upper Egypt (Cunnison 1971). This latter hypothesis is also supported by Braukämper (1993, p. 19) who links the beginning of the westwards migration of the Baggara ancestors with the famines that affected the Nile valley during the second half of the 15th century. Further to this, there is an unmistakable linguistic evidence pointing to an influx of Upper Egypt Arabic into Baggara dialects (Owens 1993b, cf. 4.1).

Against this backdrop, the question of when and how Arab nomads abandoned camels in favor of cows remains quite controversial. Braukämper (1993, pp. 17–20) suggests that the Baggaraization process started after the overthrow of the Tunjur dynasty in Wadai (eastern Chad) in 1635. This event, which is referred to as *Šagat al-Nāga* ‘the division of the she-camel’ in Chadian and Sudanese oral traditions (MacMichael 1912, p. 151), would have led to a southward movement of the Arab groups that supported the Tunjur dynasty. Following this population displacement, Arab nomads came into contact with Fulani cattle herders settled in the low rainfall savannas and eventually switched from camel to cattle breeding, while maintaining Arabic as their ancestral language. At variance with this hypothesis, Owens (1993b, p. 166; 2003, p. 723), making use of both historical and linguistic data, claims that the Baggaraization process took place as early as the 15th century in the area of Kanem-Bornu and Baguirmi (western Chad), since Arabs and Fulani had already reached this region by that time.

In the present paper, we stress instead that Baggaraization should be seen as a progressive process of socio-economic integration rather than a sudden ethnolinguistic hybridization induced by the adaptation to new ecological conditions. We therefore argue that the need for economic differentiation of both sedentary and nomadic groups is the main factor behind the emergence and the diffusion of the Baggara semi-nomadic production system across eastern Sahel. In this regard, Haaland (1969) convincingly shows that the Baggara tribes of Darfur (i.e., Ta’isha, Rizeygāt, Bānu Ḥalba, cf. Figure 1) are characterized by the incorporation of local sedentary groups, as cattle-owning Fur farmers frequently establish themselves as Baggara nomads. In a similar manner, Manfredi (2010) observes that the Ḥawāzma, the larger Baggara tribe of Sudan (cf. Figure 1), are mixed with both sedentary non-Arab groups of Kordofan (mainly Nuba) and with camel-herders coming from eastern Sudan (the so-called *Abbāla*). The same is true for the Bānu Sulēym tribe of the White

Nile (cf. Figure 1) who integrated eastern Arab groups (mainly Aḥāmda) as well as the sedentary Shilluk of South Sudan. The fact that Baggarrization is a gradual and ongoing process of socio-economic integration is also revealed by recent genetic studies (Čížková et al. 2017; Priehodová et al. 2020; Nováčková et al. 2020), which indicate that, despite a remarkable degree of ethnic admixture between agro-pastoralist groups of the Sahel, biological contacts between Fulani and Arab nomads must have been rather infrequent. These circumstances support the idea that the Baggarrization process took place at different times, across a wide geographical front, and involved different Arab and non-Arab groups. As we will see, this underlying ethnolinguistic heterogeneity is the main reason for the absence of interference from a single substrate language (i.e., Fulani) in Baggara Arabic (cf. 4.1, 5).

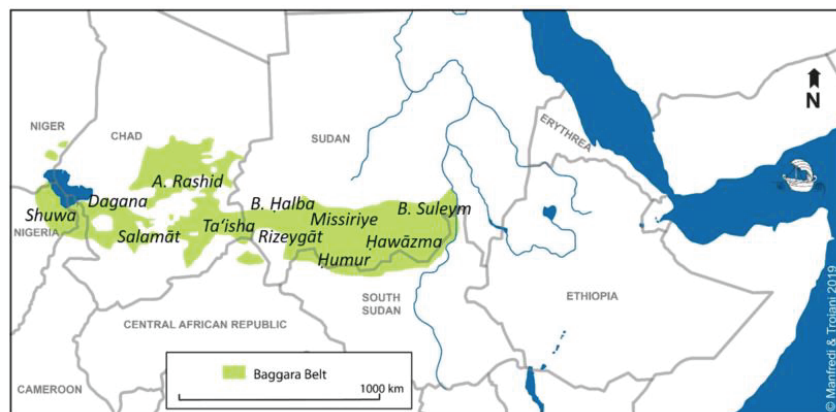


Figure 1. The Baggara Belt and its main tribes.

At the present time, Baggara Arabs are involved by different dynamics of language contact, mainly depending on their degree of sedentarization and their relative demographic weight. On the one hand, the Shuwa of north-eastern Nigeria represent a largely sedentarized linguistic minority. Accordingly, speakers of Nigerian Arabic present a high degree of bilingual proficiency in Kanuri and/or Hausa, while maintaining transmission of their ancestral language to younger generations (Owens 2020, p. 177). On the other hand, Baggara Arabs of Sudan represent an ethnolinguistic majority and they still hold on to their semi-nomadic production system. Accordingly, they have hardly developed any bilingual competence in the different languages of sedentary communities of Darfur and Kordofan. Nevertheless, due to the dominant position of Arabic in Sudan, western Baggara groups are affected at different degrees by dialect mixing and leveling towards Sudanese Arabic (Manfredi 2013, cf. 5).

Finally, it is worth remarking that Baggara Arabic historically represented the target language of non-Arab sedentary bilingual communities of Chad and western Sudan. Most sedentary communities dispersed across the West Sudanic dialect area speak Arabic as a vehicular language (see Roth-Laly 1979 for the variety of Abbeche, eastern Chad). In such a context, an increasing number of town dwellers in western Sudan (Darfur and Kordofan) are shifting from their ancestral languages to Arabic (Manfredi 2012; Roset 2018). It is thus not surprising that Baggara dialects and the Arabic varieties spoken by non-Arab sedentary groups display a high degree of mutual intelligibility. However, we will also see that, due to the stronger influence from local languages, the Arabic varieties of sedentary communities display a number of divergent morpho-phonological features (e.g., depharyngealization, lack of implosive consonants, weakening of F.PL as morphological category, cf. 4.1, 5) that allow us to draw a distinction between Baggara and ‘Sedentary’ West Sudanic Arabic.

3. Sample and Sources

The data used for our comparative overview of Baggara Arabic come from different sources. First, we refer to a heterogeneous literature that provides linguistic information on different Baggara dialects of Chad and Sudan. These bibliographical sources have been completed by new first-hand data gathered during fieldwork in the White Nile region (Sudan) in 2018. The data and the sources can be summarized as follows:

- Nigerian Baggara Arabic (NA) Shuwa tribe (Carbou 1913; Lethem 1920; Owens 1993a, 1993b);
- Baguirmi Baggara Arabic (BgA) Walād ‘Ali tribe (Zeltner and Tourneux 1986; Decobert 1985);
- Batha Baggara Arabic (BaA), Walād Rashid tribe (Decobert 1985);
- Kordofanian Baggara Arabic (KA), Ḥawāzma and Missiriye tribes (Manfredi 2010, 2013); and
- White Nile Baggara Arabic (WA), Bānu Sulēym tribe (own fieldwork data).

For the aims of this study, we will also make reference to Shukriyya Arabic (ShA) (Reichmuth 1983) which provides room for comparison between Baggara Arabic and an Eastern Sudanic dialect. Moreover, we will largely disregard both ESA and WSA Sudanic sedentary dialects in the quest to provide evidence for dialect convergence or divergence within the Baggara Belt. Nonetheless, the comparison between Baggara Arabic and the sedentary dialects of Chad and Sudan offers a number of interesting hints concerning the internal classification of WSA as a whole (cf. 5). The geographical distribution of the dialects included in our sample is shown in Figure 2.

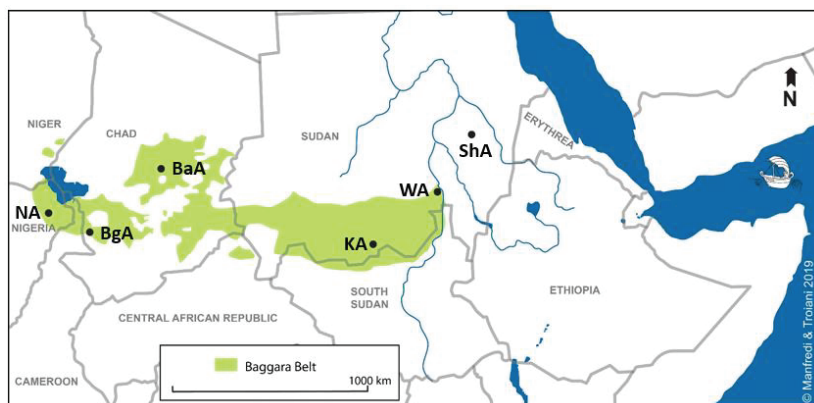


Figure 2. The geographical distribution of the Arabic dialects of the sample.

4. Assessing Diatopic Variation across the Baggara Belt

In this section, we provide a qualitative overview of selected dialect features of Baggara Arabic. The analysis is primarily intended to show the geographical distribution of Figure 2 these features and to explain the dynamics of dialect convergences and divergences across the Baggara Belt. For this purpose, we will analyze phonological (4.1), morphosyntactic (4.2), and lexical (4.3) isoglosses from the perspective of internally motivated, and externally motivated as well as multi-causal changes.

4.1. Phonological Features

If we omit a few phonological features attested all across the Baggara Belt (ex. **-a > -e* in pre-pausal position, e.g., **kabīr-a > kabīr-e* “big (F)”; presence of backness vowel harmony, e.g., **simsim > *sumsim > sumsum* “sesame”), Baggara Arabic is affected by a high degree of phonological variation. This is actually not surprising as phonological features typically have a low stability gradient and they are therefore more likely to undergo both internally

and externally induced changes. If we take a look at the domain of pharyngealized (i.e., emphatic) consonants, Baggara Arabic presents a number of phonological splits producing a rich set of non-etymological pharyngealized consonants.

- | | | | |
|----|---------------------|-----|---|
| 1. | <i>aḡm</i> “mother” | vs. | <i>ʕamm</i> “paternal uncle” |
| | <i>karro</i> “cart” | vs. | <i>karr-o</i> ‘they cooed’ (coo-3PL.M) |
| | <i>gʕlb</i> “heart” | vs. | <i>galab</i> “he came back” (come_back.3SG.M) |
| | <i>bāba</i> “daddy” | vs. | <i>bāb = a</i> ‘its door’ (door = 3SG.M) |

However, the etymological pharyngeal consonants *ħ and *ʕ are diversely affected by depharyngealization. Owens (1993b, 2020) claims that the phonological developments *ħ > h and *ʕ > ʔ, Ø represent a defining feature of WSA and he further argues that “this change could have been due to substratal influence, originally non-native speakers having difficulty in mastering ħ/ʕ.” (1993b, p. 163). This hypothesis is indeed plausible for sedentary varieties of WSA spoken by non-Arab groups, which are characterized by the complete loss of pharyngeal consonants (Jullien de Pommerol 1999b, p. 11; Roth-Laly 1972, p. 68; Manfredi 2013, p. 24; Roset 2018, p. 18). However, the situation is quite different when it comes to the distribution of *ħ and *ʕ across the Baggara Belt.

- | | |
|----|--|
| 2. | NA- <i>*ħilim</i> > <i>hilim</i> ‘he dreamt’ (dream.3SG.M), <i>*gaʕad</i> > <i>gaʔad</i> ‘he sat down’ (sit.3SG.M) |
| | BbA- <i>*ħille</i> > <i>hille</i> ‘village’, <i>*naʕla</i> > <i>naʔala</i> ‘sandal’ |
| | BaA- <i>ʕud</i> ‘stick’, <i>ħille</i> ‘village’ |
| | KA- <i>ħilim</i> ‘he dreamt’ (dream.3SG.M), <i>gaʕad</i> ‘he sat down’ (sit.SG.M) |
| | WA- <i>ħalla</i> ‘he released’ (release.3SG.M), <i>gaʕad</i> ‘he sat down’ (sit.SG.M) |

Example (2) shows that, with the exception of NA and BbA, Baggara dialects retain pharyngeal consonants. Furthermore, KA gives evidence of pharyngealization of the etymological glottal stop in intervocalic position (e.g., **raʔā > riʕa* see.3SG.M, Manfredi 2010, p. 232). Further to this, at the beginning of the 20th century, Carbou (1913) and Lethem (1920) reported the presence of pharyngeal consonants in western Chad and Nigeria, respectively. This state of affairs inevitably weakens the hypothesis that depharyngealization in western Baggara dialects is a product of substrate interference due to second language acquisition. Contrariwise, if we consider that Nigerian Arabs have developed a high bilingual proficiency in Kanuri and/or Hausa following their progressive sedentarization (cf. 3), a more plausible hypothesis is that depharyngealization is a relatively recent phenomenon triggered by language attrition. In this perspective, speakers of NA gradually lost their ability to produce the etymological sounds *ħ and *ʕ and they replaced them with their laryngeal and glottal counterparts (cf. Lucas and Manfredi 2020, p. 6). All things considered, depharyngealization is not a defining feature of WSA, but it is rather an important phonological feature distinguishing Nigerian and Baguirmi Arabic from other Baggara varieties.

The innovative nature of the westernmost Baggara phonologies (i.e., NA and BgA, cf. Figure 3) is confirmed by other features differentiating them from eastern Baggara varieties. This is the case of the insertion of an epenthetic vowel after *x*, *h*, and *q*, whose occurrence is also limited to Nigeria and western Chad.

- | | |
|----|---|
| 3. | NA, BbA <i>*aħmar</i> > <i>ahamar</i> “red”, <i>axdar</i> > <i>axadar</i> “green” |
| | BaA, KA, WA, ShA <i>aħmar</i> “red”, <i>axdar</i> “green” |

Owens (1993b, pp. 96–97, 161) and Owens and Jidda (2006, p. 710) consider guttural epenthesis a generalized feature of WSA. This is because this syllable change is also attested in most sedentary dialects of Chad and western Sudan (Jullien de Pommerol 1999b, pp. 28–29; Roth-Laly 1979, pp. 107–8; Roset 2018, p. 29). In this general context, eastern Baggara dialects (BaA, KA, and WA) as well as ShA are characterized by a higher degree of stability of syllable structures, as they do not display guttural epenthesis.

An important segmental feature subjected to diatopic variation across the Baggara Belt is the reflex of the etymological dental emphatic *ṭ. It has been argued that the common reflex of *ṭ in WSA is an implosive emphatic ḍ (Owens and Jidda 2006, p. 709). Nonetheless, sedentary dialects of Chad and western Sudan have *t* as the most common reflex of the

etymological voiceless dental emphatic (Jullien de Pommerol 1999b, pp. 28–29; Roth-Laly 1972, p. 69; Manfredi 2013, p. 24; Roset 2018, p. 41). Looking at Baggara dialects, the implosive emphatic *ɗ* presents different phonological statuses.

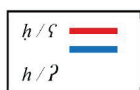
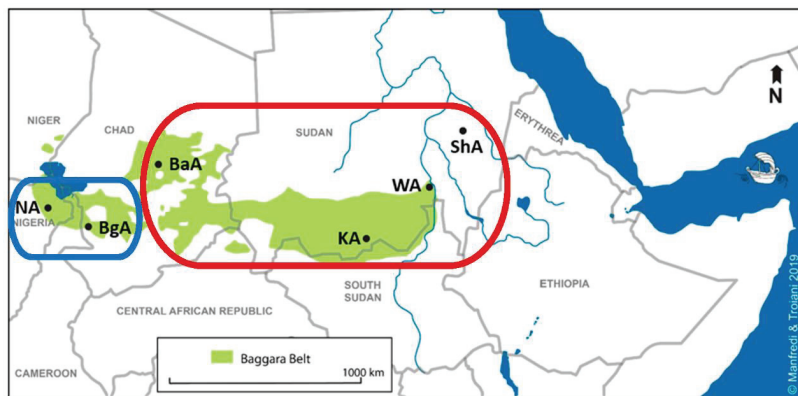


Figure 3. Depharyngealization.

4. NA **rabaɗ* > *ɗabaɗ* ‘he tied’ (tie.3SG.M),
 BgA *ɗawwal* ‘he was late’ (be_late.3SG.M)
 KA *ɗaršān* [ɗarša:n], *ɗīn* ‘mud’
 WA *ɗawwa* ‘he lifted’ (lift.3SG.M)

If *ɗ* is a full-fledged phoneme in NA, in KA the implosive emphatic [ɗ] only occurs as an allophone of *ɗ* before open vowels (Manfredi 2010, p. 44). The other three Baggara varieties included in our sample align to ESA dialects in presenting an etymological *ɗ (see Reichmuth 1983, p. 44 for ShA). Concerning the origin of the implosive realization of *ɗ, Owens (2020, p. 179) argues that it represents a possible candidate for substrate interference in WSA, as Fulani also has a dental implosive consonant. This hypothesis, however, neglects the fact that Upper Egypt dialects also present a glottalized realization for the etymological *ɗ. Khalafallah (1969, p. 29), for example, states that *ɗ* and its glottalized reflex are in partial complementary distribution in Sa’idi Arabic. Behnstedt and Woidich (1985), on their part, claim that a glottalized realization of *ɗ* is attested from Asyut to Aswan. More recently, Schroepfer (2016, p. 152) shows that, similar to what is observed in KA, in Aswan Arabic, *ɗ* is in variation with [ɗ] in pre-vocalic position. Accordingly, it seems plausible to think that the origin of *ɗ* is an inherited feature from Upper Nile dialects rather than a phonological innovation due to substrate interference from Fulani. In this perspective, the phonological status of *ɗ* in NA would have been strengthened only at a later stage due to broader areal diffusion, as most local languages of the Chari-Baguirmi region present implosive consonants (Maddieson 2013). Conversely, the absence of the reflex *ɗ* in the other Baggara varieties could be explained in light of dialect leveling towards regional standards lacking glottalized realizations (i.e., Sudanese and Chadian Arabic, *ɗ*). All things considered, the complex geographic distribution of the implosive emphatic *ɗ* across the Baggara Belts should be interpreted as the result of a multi-causal change involving language inheritance from Upper Nile dialects, areal diffusion from local languages in the Lake Chad region, and dialect leveling.

The phonological reflexes of the etymological voiced velar fricative *g are also variably affected by (im)plosivization.

5. NA *ǧādi > qādi 'there', *šūǧul > suqul 'thing'
 BgA, BaA *ǧayyar > ǧʻayyar 'he changed' (change.3.SG.M), *šūǧul > šuǧʻul 'thing'
 KA *ǧanam > qanam [ǧʻɑnɑm] 'goats', *šūǧul > šoqol 'thing'
 WA *ǧanam > ǧanam 'goats', *šūǧul > šūǧul 'thing'

Example 5 shows that Chadian Baggara dialects (i.e., BbA, BaA) have a uvular implosive ǧʻ as basic reflex of the voiced velar fricative *ǧ (Decobert 1985, pp. 45–47; Zeltner and Tourneux 1986, pp. 16, 23), whereas NA (Owens 1993a, p. 20) and KA (Manfredi 2010, p. 231) present a voiceless uvular plosive q, which is typical of Levantine Bedouin dialects (Rosenhouse 2006, p. 261). Similar to what is observed with t [dʻ] above, in KA, the uvular implosive [ǧʻ] can occur as an allophonic realization before open vowels. Lastly, in line with ESA dialects (i.e., ShA, Reichmuth 1983, p. 46), the Baggara variety of the White Nile does not present any innovative development of the voiced velar fricative *ǧ. In this context, it should be also remembered that WSA sedentary dialects stand apart from Baggara dialects in that they typically present a voiceless reflex x for the etymological *ǧ (Roth-Laly 1994a, p. 77; Roset 2018, p. 36).

According to Owens (1993b, p. 165) the occurrence of a uvular implosive ǧʻ in the Chari-Baguirmi region provides strong evidence for a Fulani substratal input in WSA, as Fulani (Niger-Congo) is among the few languages in the area with implosive ǧʻ. In fact, if we look at the geographic distribution of ǧʻ across the Baggara Belt (cf. Figure 4), it is plausible that ǧʻ represents an innovation emerging in Chad from a former voiced velar plosive q, which is still attested at the fringes of the Baggara dialect continuum (i.e., Nigeria in the west, Kordofan in the east). Furthermore, it is also true that ǧʻ is rarer than dʻ in the local languages spoken across the Baggara Belt. Despite this, there is no particular reason to postulate a Fulani substrate in Baggara Arabic, as the uvular implosive ǧʻ is also found in Afro-Asiatic (e.g., the Chadic languages Tera and Bole) and Nilo-Saharan (e.g., Central Sudanic-Sara-Bongo) languages spoken in the wider Lake Chad region (Maddieson 2013). In view of the above, ǧʻ can be better analyzed as a phonological innovation that emerged in Chad due to areal diffusion and whose geographical dispersion across the Baggara Belt is affected by both the persistence of conservative phonological features (i.e., q in NA and KA) and by the influence of ESA varieties (i.e., ǧ in WA).

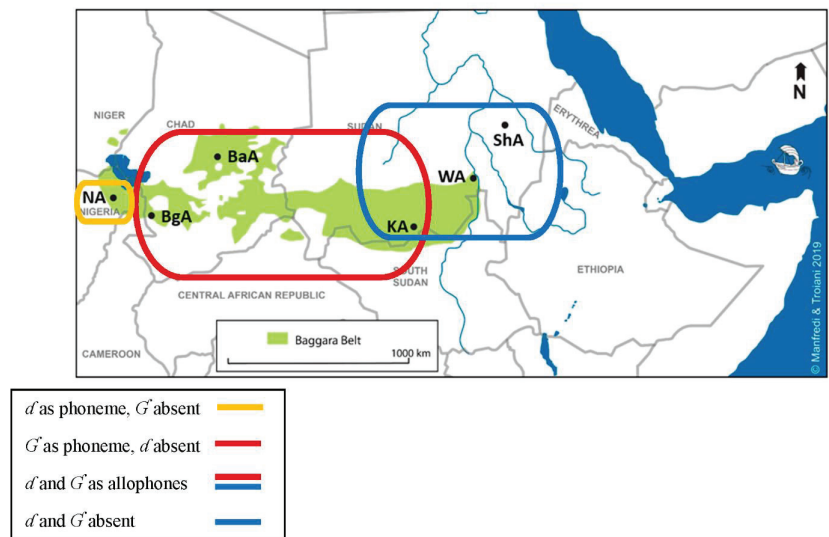


Figure 4. Implosive consonants.

Another segmental feature that draws attention in our phonological comparison of Baggara dialects is the occurrence of the voiceless postalveolar affricate \check{c} [tʃ]. Owens (1993b, p. 161) and Roth-Laly (1994b, p. 77) consider \check{c} as a Pan-Sudanic feature (cf. 1). In fact, if we exclude ESA Bedouin dialects (i.e., ShA, Reichmuth 1983, p. 43), \check{c} seems to be attested all across the Sudanic dialect area. Despite this, the phonological status of \check{c} varies a great deal across the Baggara Belt. In most cases, \check{c} is found either in ideophones (cf. 1) or in loanwords from different local languages.

- 6. NA *ču* IDPH, *čabaq* 'he waded through' (wade_through.3SG.M)
- BgA *čat* IDPH, *čilal* 'milvus'
- BaB *čut* IDPH, *kolči* 'groundnuts'
- KA *čall* IDPH, *čororo* 'topping for sorghum'
- WA *čall* IDPH

These non-etymological occurrences of \check{c} may suggest a marginal phonemic status of this phoneme in Baggara dialects. Nevertheless, the origin of the phoneme \check{c} can also be traced back to internal phonological changes.

- 7. NA, BgA **šakka* > *čakka* 'he pierced' (pierce.3SG.M)
- KA **šakšāka* > *čakčāka* 'drizzle'
- 8. BaA, BgA **wajh* > *wiči* 'face'
- KA **wajh* > *wičč* 'face'

As we can see in the previous examples, in Baggara Arabic \check{c} may represent either a minority reflex of the etymological \check{s} or the output of phonological assimilation between the voiced postalveolar affricate j and a following laryngeal h . Still, none of these internal changes is attested in WA (e.g., *šakka* 'he pierced'; *wašš* 'face'). The weakening of the phonological status of \check{c} in the White Nile region suggests a western (i.e., Chadian) origin for this phoneme.

Lastly, an interesting case of suprasegmental change that variably affects Baggara Arabic is represented by the regressive assimilation $nt > tt$ in 2nd person independent personal pronouns (cf. Figure 5).

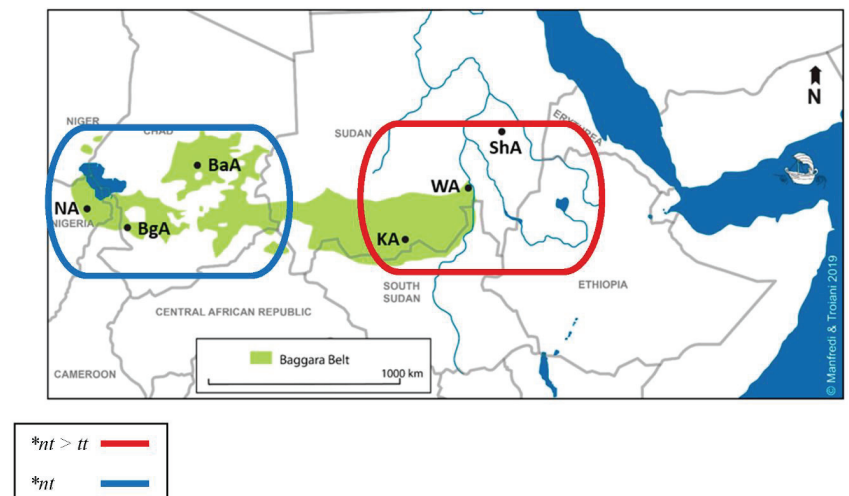


Figure 5. $*nt > tt$ in 2nd person independent pronouns.

As we can see in Table 1, western Baggara dialects (i.e., NA, BbA and BaA) are characterized by conservative pronominal forms retaining the nasal-alveolar cluster $*nt$, whereas eastern Baggara dialects (i.e., KA and WA) give evidence of the regressive assimilation $*nt >$

tt. Given that Baggara Arabic as a whole is characterized by a remarkable stability of bound pronouns,¹ regressive assimilation in independent pronouns suggests that free morphemes are more likely to undergo phonological change than bound morphemes. In comparative terms, the west–east split in the domain of 2nd person independent personal pronouns proves the integration of ESA Bedouin features in eastern Baggara dialects (cf. 4.2, 4.3, 5), as the assimilation *nt* > *tt* is also attested in ShA (Reichmuth 1983, p. 102).

Table 1. 2nd person independent pronouns.

	NA, BgA, BaA	KA	WA, ShA
2SG.M	<i>inta</i>	<i>itte</i>	<i>itta</i>
2SG.F	<i>inti</i>	<i>itti</i>	<i>itti</i>
2PL.M	<i>intu</i>	<i>uttu</i>	<i>ittu</i>
2PL.F	<i>intan</i>	<i>ittan</i>	<i>ittan</i>

4.2. Morphosyntactic Features

Baggara Arabic is bound by a few innovative morphosyntactic changes that distinguish it from other WSA and ESA dialects. For example, all the Baggara dialects included in our sample elide 1st singular and 2nd singular masculine pronominal affixes in the suffixed conjugation of consonant-final verbs lacking of nominal/pronominal objects, as shown by Table 2.

Table 2. Elision of pronominal subjects in the suffixed conjugation.

<i>katáb</i> write\1SG	<i>katáb-t = a</i> write-1SG = 3SG.M
<i>katáb</i> write\2SG.M <i>kátab</i> write.3SG.M	<i>katáb-t = a</i> write-2SG.M = 3SG.M <i>Kátab = a</i> write.3SG.M = 3SG.M

9. KA, elision of 1st SG person in absence of nominal objects

wiṣīl

arrive-1SG

‘I arrived.’

wiṣīl-ta kudūgli

arrive-1SG Kadugli

‘I arrived in Kadugli.’

In these conditions, stress is grammatically distinctive as it distinguishes between 1st singular/2nd singular masculine and 3rd singular masculine pronominal subjects (Zeltner and Tourneux 1986, p. 72; Owens 1993a, p. 111; Manfredi 2010, p. 240). In a different manner, sedentary dialects of Chad and Sudan present the suffix *-ta* for both 1st singular and 2nd singular masculine persons (Roth-Laly 1979, p. 2; Owens 1993b, p. 131; Dickins 2006, p. 563; Manfredi 2013, p. 15; Roset 2018, p. 177), whereas the form *-t* seems to be limited to ESA Bedouin dialects (i.e., ShA, Reichmuth 1983, p. 281).

In contrast to the above, verbal inflection may also be affected by an important degree of diatopic variation across the Baggara Belt. This is the case of 1st singular/1st plural person marking in the prefixed conjugation.

As is well known, Arabic dialects can be broadly classified into three morphological types depending on the 1st singular/1st plural pronominal affixes of the prefixed conjugation (i.e., type-1 (*b-*)*a-*... 1SG vs. *n-*... 1PL; type-2 *a-*... 1SG vs. *n-*... *-u* 1PL; type-3 *n-*... 1SG vs. *n-*... *-u* 1PL). If type-1 is mainly found in eastern (i.e., Levantine) Arabic dialects, type-3 is generally supposed to be a western (i.e., Maghrebi) feature spreading up to eastern Egypt. Type-2, on its part, seems to be limited to a few buffering zones in the Nile Delta and in Upper Egypt (Behnstedt 1998).

Despite important differences in their historical reconstruction, Owens and Jidda (2006) and Behnstedt (2016) agree on the fact that the attestation of type-3 in Chad is proof

of the migration of speakers out of Upper Egypt into the Sudanic region. Nevertheless, if we look at the paradigms in Table 3, it clearly appears that the diffusion of type-3 in Baggara Arabic is affected by both internal developments and dialect contact. On one side, Chadian Baggara dialects (i.e., BgA, BaA) present type-3 forms *n*-... 1SG vs. *n*-... -*u* 1PL.

Table 3. 1SG/1PL marking in prefixed conjugation.

	NA	BgA, BaA	KA	WA, SHa
CvCvC 1SG	<i>ba-ktub</i>	<i>na-ktub</i>	<i>(b-)a-ktub</i>	<i>(b-)a-ktub</i>
CvCvC 1PL	<i>na-ktub</i>	<i>na-ktub-u</i>	<i>na-ktub(-u)</i>	<i>na-ktub</i>
CvCv 1SG	<i>ba-mši</i>	<i>na-mši</i>	<i>(b-)a-mši</i>	<i>(b-)a-mši</i>
CvCv 1PL	<i>na-mši</i>	<i>na-mš-u</i>	<i>na-mš-u</i>	<i>na-mši</i>

On the other side, WA shares with ShA a more conservative type-1 paradigm, *a*-... 1SG vs. *n*-... 1PL. Nigerian Arabic, on its part, presents an innovative type-1-derived paradigm in which the preverbal marker **b*- has been integrated into the 1st singular affix **a*-, i.e., *ba*- 1SG vs. *n*-... 1PL. In light of the above, it seems plausible to think that both type-3 and type-1 dialects played a role in the emergence of Baggara Arabic, the former type still covering the core of Baggara Belt (i.e., Chad), with the latter now being limited to the geographical fringes of the dialect continuum (i.e., Nigeria, White Nile). In this context, KA falls again into a contact zone characterized by a mixed type 2-paradigm, *a*-... 1SG vs. *n*-... -*u* 1PL. It is also worth remembering that most WSA sedentary dialects present a type-3 prefixed conjugation (Roth-Laly 1979, p. 3; Jullien de Pommerol 1999b, p. 131; Roset 2018, p. 178) and that for this reason, the forms *n*-... 1SG vs. *n*-... -*u* 1PL can well be considered as a WSA feature.

However, non-native varieties of Arabic in western Sudan tend to neutralize number distinction of 1st persons. Accordingly, they generalize the use of the prefix *n*- to both 1st singular and plural persons (Manfredi 2013, p. 42). This instance of paradigm simplification by analogy (i.e., *n*-... 1SG vs. *n*-... -*u* 1PL > *n* - 1) proves that Baggara dialects represent the main target varieties of non-native speakers of Arabic in Chad and western Sudan. Figure 6 resumes the distribution of the pronominal prefixes of the prefixed conjugation across the Baggara Belt.

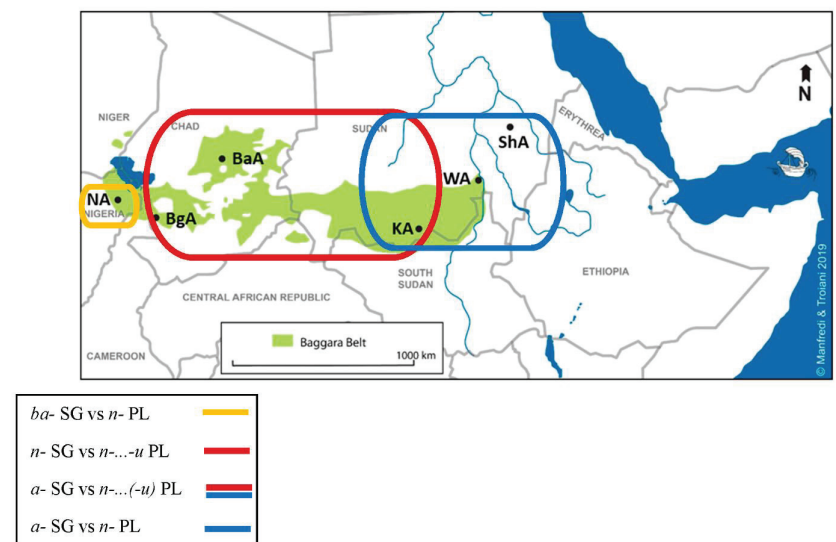


Figure 6. 1st person marking in prefixed conjugation.

Morphosyntactic variation across the Baggara Belt can also be induced by the emergence of isolated features due to internal change. This is the case of the interrogative pronominals “who” and “which one”.

10. NA, BgA *miné* ‘who’ / *atú* ‘which (one)’
 WA *min = ú* I ‘who’ (who = 3SG.M) / *yat = ú* ‘which (one)’ (which = 3SG.M)
 KA *at = ú* ‘who, which (one)’ (which = 3SG.M)

Generally speaking, like most Arabic dialects, Baggara Arabic marks a distinction between the non-selective pronoun ‘who’ and the selective pronoun ‘which one’. In WA, these interrogative pronominals are inflected for number and gender by means of accented clitic pronouns, whereas western Baggara dialects (i.e., NA, BbA) present two invariable pronominal forms. In this context, KA is the only Baggara dialect to express both non-selective and selective meanings by means of a single morphological form inflected for number and gender (i.e., *at=*, Manfredi 2010, p. 218). Given that Niger-Congo and Nilo-Saharan languages in contact with Baggara Arabic in the Nuba Mountain region (Southern Kordofan) formally distinguish ‘who’ and ‘which one’, this isolated feature of KA can only be imputed to an internal change not shared by other Baggara dialects.

Finally, diatopic variation in morphosyntactic structures can also be a product of the diverse impact of areal diffusion on Baggara Arabic. This kind of contact-induced change can be exemplified by two competing comparative constructions across the Baggara Belt.

11. NA, BgA, KA, exceed comparative with *fāt* ‘pass, surpass’
ħajm = i b = u-fūt = ak
 size = 1SG IND = 3SG.M-surpass = 2SG.M
 ‘I’m bigger than you.’
 WA, ShA, elative form with locative marking
ana akbar min = ak
 1SG big from = 2SG.M
 ‘I’m bigger than you.’

Example 10 shows that, in line with most Sub-Saharan languages (Stassen 2013), Baggara Arabic presents exceed comparative constructions in which the standard is constructed as the object (i.e., *=ak* 2SG.M) of the transitive verb *fāt* ‘surpass’. Still, this instance of grammatical calquing induced by areal diffusion does not reach WA which, similarly to ShA and other ESA dialects, presents a more common locational comparative construction with the standard introduced by the preposition *min* ‘from’. The absence of exceed comparative constructions in the eastern fringes of the Baggara Belt is reasonably another output of dialect leveling towards ESA and it is another argument in favor of a west–east migration of Baggara groups.

4.3. Lexical Features

The Baggara dialects included in our sample share a number of interesting lexical innovations. These include the lexemes *ħ/harrāy* ‘sun’ (Behnstedt and Woidich 2011, p. 402) and *elmi* ‘water’ (< **al = mi* < **al = mā?*, Behnstedt and Woidich 2011, p. 420), which differ from the more common *šemis/šemiš* ‘sun’ and *mōya* ‘water’. Furthermore, Baggara Arabic gives evidence of a few conservative lexical features, as in the case of the root **ra?ā* for the verb ‘see’ (Behnstedt and Woidich 2014, p. 330; cf. 4.1). Despite these surface affinities, Baggara Arabic displays a high degree of lexical variation. First of all, the westernmost Baggara dialects (i.e., NA, BgA) stands out from the other varieties included in our sample due to a number of loanwords from Kanuri and other languages spoken in the Lake Chad region (Behnstedt and Woidich 2011, p. 23; 2014, p. 730).

12. NA, BgA *kāka-y* ‘(my) grandmother’ (from Kanuri *kākā*)
 BgA, KA, WA *aṃṃ aṃṃ-í, aṃṃaṃt-í* ‘(my maternal) grandmother’ (lit. ‘the mother of my mother’)
 13. NA, BgA *čokol* ‘fork’ (from Hausa *cóokàlìí*)
 KA, WA *šōka* ‘fork’

Most commonly, lexical variation across the Baggara Belt results from dialect contact. The possessive particles (POSS) in Table 4 are a case in point.

Table 4. Possessive particles.

	NA	BgA	KA	WA	ShA
POSS.SG.M	<i>hana</i>	<i>hana</i>	<i>hān, hūl</i>	<i>hūl</i>	<i>hūl</i>
POSS.SG.F	<i>hil, hinta</i>	<i>hil</i>	<i>hint, hūl</i>	<i>hūl</i>	<i>hūl</i>
POSS.PL	<i>hinen</i>	<i>hiney</i>	<i>hinēn, hilēl</i>	<i>hilēl</i>	<i>hilēl</i>

As we can see, western Baggara dialects (i.e., NA and BgA) are bound by the forms *hana* POSS.SG.M and *hine(n)* POSS.PL, as opposed to *hūl* POSS.SG.M and *hilēl* POSS.PL in WA. Conversely, the form *hil/hūl* POSS.SG.F is attested all across the Baggara Belt, whereas the singular feminine form *hint* is limited to NA and KA. In this overall situation, KA clearly falls into a buffer zone in which western and eastern lexical forms are still in competition (Manfredi 2012). On the one hand, *han(a)/hint(a)/hinēn* forms originate in Upper Egypt (Owens 1993b, p. 111) and they represent a WSA feature within Sudanic Arabic as they are also attested in the sedentary dialects of Chad. On the other hand, *hūl/hūl/hilēl* possessive particles are common to ESA dialects spoken by groups that penetrated Sudan directly from the Arabian Peninsula (i.e., ShA, Reichmuth 1983, pp. 111–12) and they are not attested in the urban dialects of eastern Sudan. The complex geographical distribution of *hān(a)/hint(a)/hinēn* and *hūl/hūl/hilēl* across the Baggara Belt (cf. Figure 7) seems to indicate a longstanding coexistence of these possessive forms and may corroborate the idea that speakers of ESA dialects have also been historically involved in the Baggarization process (cf. 2, 5).

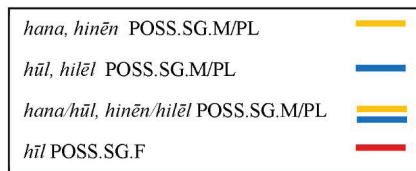
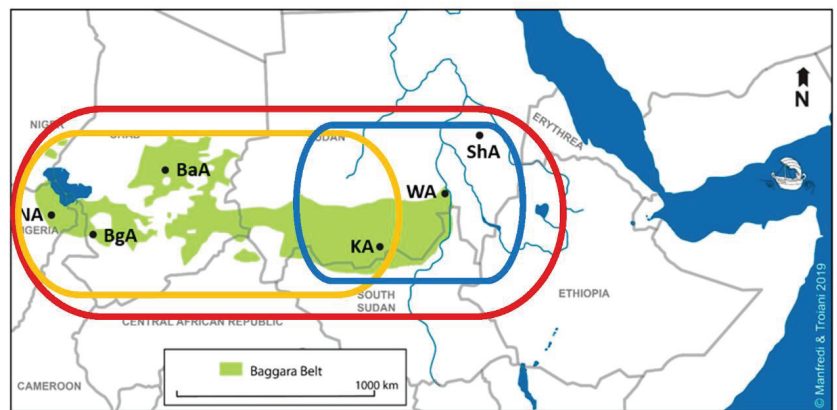


Figure 7. Possessive particles.

The intensifier ‘very’ provides another example of lexical variation due to dialect contact and leveling across the Baggara Belt.

14. NA, BgA *bilhēn* ‘very’; BaA, KA *bilhēn* ‘very’
WA *šedīd* ‘very’

In fact, except WA, all the Baggara dialects included in our sample display the form *bilhēn/bilhēn*, which finds its origin in the prepositional phrase **balḥayl* ‘very’ (< **bi-l-ḥayl* ‘by strength’) attested in a number of Middle Eastern Bedouin dialects (Rosenhouse 2006, p. 267). WA, on its part, aligns with ESA in using the adjective **šadīd* ‘strong’ as intensifier. This lexical isogloss confirms that the Baggara dialect of the White Nile is the most affected by contact with eastern Sudanic dialects.

5. Conclusions

Based on the previous comparative overview of Baggara Arabic, we can now attempt at reconstructing the main dynamics of dialect convergence and divergence across the Baggara Belt. First of all, despite their common ethnolinguistic and sociohistorical background, Baggara dialects display a high degree of diatopic variation. Indeed, if we exclude common Pan-Sudanic features (cf. 1), there are only a few isoglosses that are shared by all five varieties included in our sample. These comprise the vowel change **-a > -e* in pre-pausal position, the presence of backness vowel harmony (cf. 4.1), the forms of bound personal pronouns (cf. note 2), the elision of pronominal subjects in the prefixed conjugation (Table 2, ex. 9), and several lexical isoglosses (cf. 4.3).

Secondly, the lack of a number of WSA innovations (e.g., etymological *č*, ex. 7–8; *bilhēn* ‘very’, ex. 14) in the White Nile region supports the hypothesis of a west (i.e., Chad) > east (i.e., Sudan) migration of Baggara groups. In fact, if KA still gives evidence of competing WSA and ESA features (e.g., type-2 prefixed paradigm, Table 3), WA is clearly more affected by contact with ESA dialects, and therefore it is more similar to ShA (e.g., type-1 prefixed paradigm, Table 3; *hūl/hilēl* SG.M/PL possessive particles, Table 4). This suggests that there is no overlap between the ethnic and the dialect borders of the Baggara Belt, as WA lost most of its WSA features while integrating several ESA innovations. In this context, the attestation of both WSA and ESA features across the Baggara Belt (e.g., type-1/type-3 paradigms, Table 3; *hūl/hint* F.SG possessive particles, Table 4) points to a longstanding coexistence of these dialect sub-types and provides evidence that the Baggarization process did not exclusively involved speakers of WSA varieties.

In terms of contact-induced change, we have shown that there is no linguistic evidence for a Fulani substrate in Baggara Arabic. In fact, the heterogeneity of the languages spoken across the Baggara Belt limits the possibility of a substrate interference via language shift. In such conditions, contact-induced innovations are mainly a product of areal diffusion on a west–east axis, from the Lake Chad region in direction of the White Nile. This is the case of both the implosive consonant *ɠ* (ex. 5) and exceed comparative constructions (ex. 11) whose grammatical productivity tend to fade eastwards. As far as NA and BgA are concerned, contact-induced changes also occurred as a consequence of language attrition. This is the case of depharyngealization (ex. 2) which should be seen as a relatively recent innovation induced by the high degree of Kanuri/Arabic bilingual proficiency of local sedentarized Baggara Arabs. The prominence of the adstrate over a supposed Fulani substrate is also testified by a number of loanwords occurring in the basic vocabulary of the western-most Baggara dialects (ex. 12). All things considered, NA and BgA undoubtedly represent the most innovative Baggara varieties of our sample and they cannot therefore be adopted as a dialect prototype for Baggara Arabic. Furthermore, the previous linguistic arguments potentially support the hypothesis of Baggarization as a gradual process of socioeconomic integration rather than a sudden ethnolinguistic hybridization between Arab and Fulani agro-pastoralist groups.

As a final remark, it is without doubt that Baggara Arabic as a whole represents a WSA dialect sub-type. Nonetheless, we have also shown a number of isoglosses opposing most Baggara dialects to sedentary dialects of Chad and western Sudan. These include the presence of pharyngealized and pharyngeal consonants (ex. 1–2), the presence of implosive consonants (especially *ɠ*, ex. 5), and the elision of pronominal subjects in the

prefixed conjugation (Table 2). This suggests that, despite the generalization of traditional Bedouin features across the Sudanic area (cf. 1), WSA is actually affected by an important eco-linguistic variation. Further to this, the structural divergences between Baggara Arabic and the sedentary dialects of Chad and western Sudan reduce the geographical extent of a number of isoglosses that were formerly thought to represent pan-WSA features (i.e., depharyngealization, implosivization). However, this is only partially true for NA and BgA that, being predominantly spoken by sedentarized Baggara Arabs, are phonologically closer to sedentary WSA dialects (ex. 3). All things considered, WSA features vary significantly according to both diatopic and eco-linguistic factors. Although there is no sharp boundary between Bedouin and Sedentary dialects in the Sudanic area, eco-linguistic factors still matter and should be therefore taken into account in further research of the area.

Author Contributions: Both authors contributed to several aspects of the study, specifically, conceptualization, S.M and C.R.; methodology, investigation, resources, and data curation S.M.; writing—original draft preparation S.M. and C.R.; writing, review and editing, S.M. and C.R.; supervision S.M. All authors have read and agreed to the published version of the manuscript. Both authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

-	affix boundary	BaA	Baṭḥa Baggara Arabic
=	clitic boundary	BgA	Baguirmi Baggara Arabic
\	ablaut	ESA	Eastern Sudanic Arabic
*	reconstructed form	KA	Kordofanian Baggara Arabic
1, 2, 3	1st, 2nd, 3rd person	NA	Nigerian Baggara Arabic
F	feminine	ShA	Shukriyya Arabic
IDPH	ideophone	WA	White Nile Baggara Arabic
M	masculine	WSA	West Sudanic Arabic
OBJ	object		
OBL	oblique		
PL	plural		
PROX	proximal		
SG	singular		

Note

¹ Among these, we can recall the forms = a 3SG.M, = ki 2SG.F (after both consonant- and vowel-final items) = ku 3PL.M which are also variably found in the sedentary dialects of Chad and western Sudan.

References

- Abu-Absi, Samir. 1995. *Chadian Arabic*. München and Newcastle: Lincom Europa.
- Behnstedt, Peter, and Manfred Woidich. 1985. *Die Ägyptisch-Arabischen Dialekte*. Beihefte Zum Tübinger Atlas Des Vorderen Orients. Reihe B, Geisteswissenschaften. Wiesbaden: L. Reichert.
- Behnstedt, Peter, and Manfred Woidich. 2011. *Wortatlas der arabischen Dialekte, Vol. 1*. Leiden: Brill.
- Behnstedt, Peter, and Manfred Woidich. 2014. *Wortatlas der arabischen Dialekte, Vol. 3*. Leiden: Brill.
- Behnstedt, Peter. 1998. La frontière orientale des parlers maghrébins en Egypte. In *Peuplement et Arabisation au Maghreb Occidental. Dialectologie et Histoire*. Edited by Jordi Agudé, Patrice Cressier and Angeles Vicente. Madrid-Zaragoza: Casa de Velázquez, pp. 85–96.
- Behnstedt, Peter. 2016. The *niktib-niktibu* issue revisited. *Wiener Zeitschrift für die Kunde des Morgenlandes* 106: 21–36.

- Bergman, Elisabeth M. 2002. *Spoken Sudanese Arabic: Grammar, Dialogues, and Glossary*. Washington: Dunwoody Press.
- Blanc, Haim. 1971. Arabic. In *Linguistics in Sub-Saharan Africa*. Edited by Jack Berry and Thomas Albert Sebeok. Berlin: De Gruyter Mouton, pp. 501–9.
- Braukämper, Ulrich. 1993. The Origin of Baggara Arab Culture with Special Reference to the Shuwa. *Sprache und Geschichte in Afrika* 14: 13–46.
- Carbou, Henri. 1913. *Méthode Pratique pour L'étude de L'arabe parlé au Ouaday et à l'est du Tchad*. Paris: Geuthner.
- Čížková, Martina, Pavel Munclinger, Mame Yoro Diallo, Iva Kulichová, Mohammed G. Mokhtar, Alioune Dème, Luísa Pereira, and Viktor Černý. 2017. Genetic Structure of the Western and Eastern African Sahel/Savannah Belt and the Role of Nomadic Pastoralists as Inferred from the Variation of D-loop mtDNA sequences. *Human Biology* 89: 281–302. [CrossRef]
- Cunnison, Ian. 1966. *Baggara Arabs. Power and Lineage in a Sudanese Nomad Tribe*. Oxford: Clarendon.
- Cunnison, Ian. 1971. Classification by genealogy: A problem of the Baqqara belt. In *Sudan in Africa*. Edited by Yusuf Fadl Hasan. Khartoum: Khartoum University Press, pp. 186–96.
- Decobert, Christian. 1985. *Phonologies Arabes du Tchad*. Paris: Geuthner.
- Dickins, James. 2011. Dialects of Egypt and Sudan. In *The Semitic Languages. An International Handbook*. Edited by Stefan Weninger, Geoffrey Khan, Micheal Streck and Janet C. E. Watson. Berlin/Göttingen: De Gruyter Mouton, pp. 935–53.
- Dickins, James. 2006. Khartoum Arabic. In *Encyclopedia of Arabic Language and Linguistics, Vol. 2.* Edited by Kees Versteegh, Mushira Eid, Alaa Elgibali, Manfred Woidich and Andrzej Zaborski. Leiden: Brill, pp. 559–71.
- Haaland, Gunnar. 1969. Economic determinants in ethnic processes. In *Ethnic Groups and Boundaries: The Social Organization of Culture Difference*. Edited by Friederik Barth. Long Grove: Waveland Press, pp. 58–73.
- Hagège, Claude. 1973. *Profil d'un Parler arabe du Tchad*. Paris: Geuthner.
- Henderson, Kenneth D.D. 1939. A note on the migration of the Messiria tribe into South West Kordofan. *Sudan Notes and Records* 22: 49–77.
- Hillelson, Sigmar. 1925. *Sudan Arabic: English-Arabic Vocabulary*, 2nd ed. London: The Sudan Government.
- Jullien de Pommerol, Patrice. 1999a. *Dictionnaire Arabe Tchadien-Français*. Paris: Karthala.
- Jullien de Pommerol, Patrice. 1999b. *Grammaire Pratique de l'arabe Tchadien*. Paris: Karthala.
- Kaye, Alan. 1976. *Chadian and Sudanese Arabic in the Light of Comparative Arabic Dialectology*. The Hague: Mouton.
- Khalafallah, Abdelghani. 1969. *A Descriptive Grammar of Sa'i:di Colloquial Egyptian Arabic*. The Hague: Mouton.
- Lethem, Gordon. 1920. *Colloquial Arabic: Shuwa Dialect of Borno, Nigeria and of the Lake Chad Region*. London: Crown Agent for the Colonies.
- Lucas, Christopher, and Stefano Manfredi. 2020. Introduction. In *Arabic and Contact-Induced Change*. Edited by Christopher Lucas and Stefano Manfredi. Berlin: Language Science Press, pp. 1–33.
- MacMichael, Harold Alfred. 1912. *The Tribes of Northern and Central Kordofan*. London: Frank Cass & Co.
- MacMichael, Harold Alfred. 1922. *A History of the Arabs in Sudan and Some Account of the People Who Preceded Them and of the Tribes Inhabiting Dar Fur*. Cambridge: Cambridge University Press, Vol. II.
- Maddieson, Ian. 2013. Glottalized Consonants. In *The World Atlas of Language Structures Online*. Edited by Mathew Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online: <http://wals.info/chapter/7> (accessed on 7 May 2021).
- Manfredi, Stefano. 2010. A Grammatical Description of Kordofanian Baggara Arabic. Ph.D. dissertation, Università degli Studi di Napoli "L'Orientale", Naples, Italy.
- Manfredi, Stefano. 2012. Dialect mixing and dialect levelling in Kordofanian Baggara Arabic. In *Dynamiques Langagières en Arabophonies: Variations, Contacts, Migrations et Créations Artistique. Hommage offert à Dominique Caubet par ses élèves et Collègues*. Edited by Alexandrine Barontini, Christophe Pereira, Angeles Vincente and Karima Ziamari. Zaragoza: Universidad de Zaragoza, Colección Estudios de Dialectología Árabe, pp. 141–62.
- Manfredi, Stefano. 2013. Native and Non-native varieties of Arabic in an emerging urban centre of western Sudan. Evidences from Kadugli. In *African Arabic: Approches to Dialectology*. Edited by Mena Lafkioui. Berlin and New York: De Gruyter, pp. 13–51.
- Nováčková, Jana, Martina Čížková, Mohammed G. Mokhtar, Pavel Duda, Vlastimil Stenzl, Petr Triska, Zuzana Hofmanová, and Viktor Černý. 2020. Subsistence strategy was the main factor driving population differentiation in the bidirectional corridor of the African Sahel. *American Journal of Physical Anthropology* 171: 496–508. [CrossRef]
- Owens, Jonathan. 1985. Arabic dialects of Chad and Nigeria. *Zeitschrift für arabische linguistik* 14: 45–61.
- Owens, Jonathan. 1993a. *A Reference Grammar of Nigerian Arabic*. Wiesbaden: Harrasowitz.
- Owens, Jonathan. 1993b. Nigerian Arabic in Comparative Perspective. *Sprache und Geschichte in Afrika* 14: 85–175.
- Owens, Jonathan. 2003. Arabic dialect history and historical linguistic mythology. *Journal of the American Oriental Society* 123: 715–40. [CrossRef]
- Owens, Jonathan. 2020. Nigerian Arabic. In *Arabic and Contact-Induced Change*. Edited by Christopher Lucas and Stefano Manfredi. Berlin: Language Science Press, pp. 175–96.
- Owens, Jonathan, and Hassan Jidda. 2006. West Sudanic Arabic. In *Encyclopedia of Arabic Language and Linguistics*. Edited by K. Versteegh, Mushira Eid, Alaa Elgibali, Manfred Woidich and Andrzej Zaborski. Leiden and Boston: Brill, Volume IV, pp. 708–19.
- Priehodová, Edita, Frederic Austerlitz, Martina Čížková, Jana Nováčková, François-Xavier Ricaut, Zuzana Hofmanová, Carina M. Schlebusch, and Viktor Černý. 2020. Sahelian pastoralism from the perspective of variants associated with lactase persistence. *American Journal of Physical Anthropology* 173: 423–36. [CrossRef] [PubMed]

- Reichmuth, Stefan. 1983. *Der arabische Dialekt der Šukriyya im Ostsudan*. Zürich: Georg Olms.
- Rosenhouse, Judith. 2006. Beduin Arabic. In *Encyclopedia of Arabic Language and Linguistics*. Edited by K. Versteegh Mushira Eid, Alaa Elgibali, Manfred Woidich and Andrzej Zaborski. Leiden and Boston: Brill, Vol. II, pp. 259–69.
- Roset, Caroline. 2018. A Grammar of Darfur Arabic. Ph.D dissertation, University of Amsterdam, LOT Publications, Amsterdam, The Netherlands.
- Roth-Laly, Arlette. 1972. Esquisse de la phonologie du parler arabe d'Abbéché. *Comptes Rendus du Groupe Linguistique d'études Chamito-Sémitiques (G.L.E.C.S.)* 16: 33–79.
- Roth-Laly, Arlette. 1979. *Esquisse Grammaticale du Parler Arabe d'Abbéché (Tchad)*. Paris: Paul Geuthner.
- Roth-Laly, Arlette. 1994a. Dialectologie comparée au Tchad et au Soudan-De quelques matériaux et discriminants. *Matériaux Arabes et Sud-Arabiens* 6: 69–102.
- Roth-Laly, Arlette. 1994b. A propos de la pondération des discriminants dans la comparaison dialectale: Le cas de l'accent dans l'aire tchado-soudanaise. In *Actes des Premières Journées Internationales de Dialectologie arabe de Paris*. Edited by Dominique Caubet and Martine Vanhove. Paris: Publications Langues'O, pp. 205–16.
- Schroepfer, Jason. 2016. Ethnic Variation of */tʕ/ in Aswan Arabic. In *Selected Papers from New Ways of Analyzing Variation (NWAY 44)*. U. Penn Working Papers in Linguistics. Philadelphia: Penn Graduate Linguistics Society, vol. 22, pp. 151–60.
- Stassen, Leon. 2013. Comparative Constructions. In *The World Atlas of Language Structures Online*. Edited by Mathew Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology, Available online: <http://wals.info/chapter/121> (accessed on 19 May 2021).
- Teitelbaum, Joel M. 1984. The transhumant production system and change among Hawazma nomads of the Kordofan region, Western Sudan. *Nomadic Peoples* 18: 51–64.
- Zeltner, Jean-Claude, and Henry Tourneux. 1986. *L'arabe dans le bassin du Tchad*. Paris: Karthala.

Article

Mahdia Dialect: An Urban Vernacular in the Tunisian Sahel Context

Cristina La Rosa

Department of Humanities (DISUM), University of Catania, 95124 Catania, Italy; cristinalarosa@unict.it

Abstract: This paper aims to present some preliminary results of the linguistic analysis of the dialect of the Wilāya of Mahdia on which few studies exist, focused mainly on phonology. My analysis, here extended to the morpho-syntactic level, is based on a corpus of interviews taken from some social media pages. The sample will be composed of respondents of different geographical origin (from Mahdia and some nearby towns), gender, age and social background. A deeper knowledge of the Arabic of Mahdia region, which is a bundle of urban, Bedouin and “villageois” varieties, would contribute to throw new light on the features of the Saḥli dialects and would add a small piece to the complex mosaic of Tunisian and Maghrebi dialects, whose traditional categories of classification should be reconsidered.

Keywords: Mahdia Arabic; Maghribi Arabic; Tunisia; Sahel; urban dialects; Bedouin dialects; *villageois* dialects; Arabic dialectology; Sociolinguistics

1. Introduction

During the recent years, the need to more precisely examine and describe the varieties of Arabic used in the Tunisian Sahel has been central to the scientific dialectological debate. In fact, there are few systematic studies available on Sahel varieties and the data offered need to be partially reinterpreted. In 1950, William Marçais, with regard to “les parlers villageois” of Tunisia, claimed: “on a surtout en vue ici ceux des bourgs et des petites villes du Sahel [. . .] n’ont fait encore l’objet d’aucune enquête” (Marçais 1950, pp. 210–11).¹ The scholar classified the non-coastal dialects of Sahel as “villageois” and introduced a third category of dialects sharing both sedentary and Bedouin features (see also (Marçais and Guiga 1925, p. XXV)).

The first systematic study on a Sahel variety is *Textes arabes de Takroûna* by William Marçais. According to him (Marçais and Guiga 1925, p. XIX), “Le parler arabe de Takroûna concorde dans l’ensemble avec ceux des centres agricoles, bourgs et villages, qui parsèment la région côtière de la Tunisie centrale, communément appelée Sâhel [. . .] depuis le moyen âge. Séparés les uns les autres par des différences de détail, ces parlers relèvent, quant à la phonétique et à la grammaire, d’un même type general dont le takroûni n’est qu’une variété particulière”.² Marçais and Guiga offered a linguistic study and also a monumental glossary of the variety considered (Marçais and Guiga 1958).

In 1980, Talmoudi classified the Tunisian dialects in four groups: The varieties of Sahel, which are urban and semi-urban, these latter “spoken in small villages as Ksibet Sûsa and Khnis display features of both nomadic and sedentary dialects”. The Northern dialects are also “divided into two types: urban and rural. The villagers on the left side of Oued Medjerda speak so-called ṣba:li dialects (mountain dialects) which have features in common with North East Algerian vernaculars”. The Central Western dialects are rural and nomadic. The rural ones “resemble in several respects the East Arabic dialects”. The Southern dialects are divided in three groups: “urban dialects in Sfax, rural in the oasis and nomadic. The latter is spoken by semi-pastoral people in Sahara” (Talmoudi 1980, pp. 10–11). “The genuine dialect of Sûsa”, according to the scholar, is spoken in the

Citation: La Rosa, Cristina. 2021. Mahdia Dialect: An Urban Vernacular in the Tunisian Sahel Context. *Languages* 6: 145. <https://doi.org/10.3390/languages6030145>

Academic Editors: Simone Bettega and Roberta Morano

Received: 2 July 2021

Accepted: 23 August 2021

Published: 27 August 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Medina by the older generation, but the author also took into account the innovations of the younger generation for the composition of his study on the Arabic of Sousse (Talmoudi 1980, p. 13).

Lajmi, in 2009, Zammit, in 2014, and Sellami, in 2019, conducted some investigation on some features of Sfaxi Arabic, providing new elements for the knowledge of the dialects of Sahel.

Mion (2015, 2018)³ offered some reflections on the origin of this “third category” of dialects, that is the “villageois”, and on the phenomena that characterise Tunisian “villageois” dialects, whose ‘mixed’ features are the product of a long process of interdialectal contact between a sedentary and a Bedouin variety of Arabic. The latter was introduced by the Banū Hilāl who invaded the Maghreb around the 11th century.

Mion and Luca D’Anna during the “Prima giornata di dialettologia maghrebina” (Cagliari May 16 2019), launched the research project “The Tunisian Sahel: Dialectological, Historical and Sociolinguistic Perspectives”, which aims at shedding new light on the features of the Arabic varieties of Tunisian Sahel, which, in fact, should be better described regardless of the existing rigid classification criteria (Bedouin/urban/rural), which do not highlight the richness characterising the varieties spoken in the region (Mion 2015). A century after Marçais’ studies, however, no systematic research on the Bedouin and urban varieties of the Sahel region has been conducted, yet.

My paper intends to be a modest contribution to the knowledge of the varieties of the area, starting from the analysis of the dialect(s) of Wilāyat al-Mahdiyya (henceforth Mahdia).

Marçais and later scholars have included Mahdia Arabic among the varieties used in the coastal towns of Sahel, such as Monastir, Sousse and Sfax. Saada (1984, p. 17) included the urban dialect of Mahdia among “les parlers arabes des capitales”, but she also stated that she did not have elements to classify the varieties used in the neighbouring villages.

At present, there are few studies entirely dedicated to the Arabic of Mahdia, and the dialects of the surrounding towns and villages have never been described. Attia (1969) offered a phonological analysis of the variety used by the fishing community in the 1960s. What emerges from his paper is a quick description of the phonological inventory of the Mahdia dialect, accompanied by few examples. In his paper, the scholar highlighted some well-known features of Mahdia Arabic: The passage of the interdental /t/ and /d/ to /d/ and /t/, the reduction of the diphthongs /ay/ and /aw/ to /ē/ and /ō/ in the middle of a word and the voiced articulation of /q/ in some terms such as *gar’a* “irrigated land” and *gamra* “moon” (Attia 1969, p. 125). He also quickly focused on combinatory phonetics, vowels and vocalic phenomena such as final *imāla*, syllable patterns and prosody. Based on these phenomena, Attia defined Mahdia Arabic an urban variety.⁴ Yoda’s study (2008) added some important information to the knowledge of the vocal system of Mahdia dialect. In fact, the scholar focused on the special status of Mahdia Arabic having /ē/ and /ō/ phonemes, unlike the other sedentary dialects, and attributed the presence of these phonemes to the influence of the nearby village dialects. Besides, in 2019, Yoda published some texts in Mahdia Arabic, accompanied by a simple grammatical sketch concerning mainly phonology. In his paper, he writes that Mahdia Arabic: “is an eastern Maghribi sedentary dialect showing some features of the village dialects of the region, the most conspicuous being the word-final *imāla* ē. Among the sedentary dialects of Tunisia, Mahdāwī dialect [. . .] is characterized by a five-long-vowel system [. . .] and the correspondence of the interdentals of Old Arabic [. . .] to plosives, as attested in most of the Jewish dialects of Tunisia. In this respect the dialect in question is worthy of more detailed descriptions”. (Yoda 2019, p. 55).

It is clear, however, that the few existing studies have focused mainly on the vowel and consonant system. Therefore, the linguistic analysis of Mahdia variety should be extended to all linguistic levels. It would also be fruitful to extend the research to the varieties used in the territory near the town, which remain to be investigated (except for D’Anna’s study about Chebba: see (D’Anna 2020)). In fact, Mahdia Arabic is worthy of study because it

is surrounded by village varieties and, since no language can live isolated and separated from the adjacent varieties, it could have some peculiarities owing to contact phenomena. That is, as Yoda already showed, it is an urban variety containing some village features.

In addition to this, a deeper knowledge of the Arabic of Mahdia region, which is a bundle of urban, Bedouin and “villageois” varieties, would contribute to shed new light on the features of the Sahel dialects and would add a small piece to the complex mosaic of Tunisian and Maghrebi dialects. In fact, a deeper knowledge of the varieties spoken in this area would also be useful to understand through which historical and socio-linguistic dynamics these dialects originated, since we have scant data on the history of Tunisian dialect(s). More generally, this would also help provide a more detailed and specific classification of the dialects of the area (Taine-Cheikh 2017; Guerrero 2018; Benkato 2019), which share many common features, but present many differences that deserve to be highlighted.⁵

Mahdia and the surrounding towns

Mahdia is a town whose importance and splendour are rooted centuries ago. The geographer al-Idrīsī (d. 1175–1176) describes Mahdia as a beautiful town, two days from Sfax and Kairouan. Mahdia, not long before his arrival, had a harbour visited by merchant ships from everywhere: Maghrib, Mashriq, al-Andalus and Christian countries. al-Idrīsī informs us that, during those times, Mahdia was already famous throughout the world because of its goods and renowned clothes that were exported to all the other countries, but he also adds that, since the Norman conquest, its trade has been strongly reduced. Moreover, the geographer defines the wall surrounding the town a wonder worthy of mention (Bresc and Nef 1999, pp. 183–84). Mahdia, during al-Idrīsī’s times, was composed of two towns: Mahdia, the seat of the power and the residence of the sovereign, and al-Zawīla. The latter was beautiful and densely populated with merchants. The geographer writes of Mahdia with a sort of nostalgia because the invasion of the “Arabs”, that is the Banū Hilāl,⁶ and of the later Norman conquest, destroyed many important and emblematic aspects of the town. Mahdia, however, remained the capital of Ifrīqiyyā (Bresc and Nef 1999, pp. 183–86).

Nowadays, the vestiges of a flourishing medieval city can be retraced in its old *madīna*. Mahdia was built on a peninsula (see Figure 1), situated on the Eastern coast in the centre of the Republic of Tunisia. It is 200 km from the capital, Tunis, has a mild climate, which is usually affected by the Mediterranean air currents, and its economy is based on agriculture, especially on oil production, fishing and craft industries specialised in producing silk, leather clothing and mosaics. Thanks to its position on a 75-km-long coast, close to Sousse and El Jem, tourism also plays an important role in its economic activities.⁷ Mahdia is a town of pre-Hilalian foundation that later underwent the Hilali invasion. Founded in 909 A.D. by the Fatimid Caliph ‘Ubayd Allāh al-Mahdī as the new capital of the Reign, replacing Kairouan, the Aġlabid capital, Mahdia was the first urban settlement in the peninsula. The history of the town is quite complex. The ancient nucleus was the already mentioned quarter of al-Zawīla, which was also the commercial core. During the Hilalian invasion, Mahdia regained its role of capital, after a short time in which the capital was al-Manṣūriyya. In 1087, Mahdia was conquered by Pisans and Genoans, then by Normans in 1123, in 1134 by the Hammādids and, in 1140, the Normans of Sicily imposed harsh conditions on the town. Then, Roger II caused the end of the Zīrid Dynasty until the French protectorate of 1884, and the town was attacked and conquered by numerous dynasties and conquerors and was even destroyed and plundered (Talbi 1986, pp. 1246–947). The suburb of al-Zawīla was completely destroyed during the Hilalian conquest and rebuilt in 1200. As a consequence of centuries of riots, pillages and plagues, in the 16th and 17th centuries, the composition of the population changed above all because of two elements: The arrival of the Muslim refugees coming from al-Andalus and the introduction of the Turkish garrisons. According to Talbi, in 1986, 60% of the population was composed of

descendants of the Kouloughlis (Bearman et al. 1986, p. 366),⁸ affecting onomastics and customs (Talbi 1986, p. 1247).

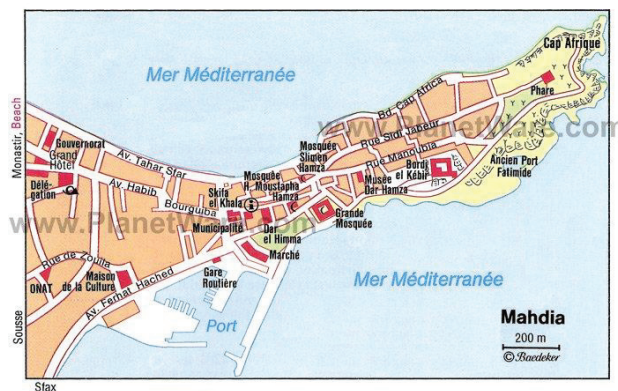


Figure 1. Contemporary map of Mahdia⁹.

Goitein (2010, p. 311) claimed that Mahdia and some coastal towns of Tunisia resisted the perpetual chaos in which the Hilalian invasion threw the country,¹⁰ but their hinterland was lost and was exposed to the subsequent attacks from the Normans and Italians. In the Geniza documents, the Christian and the Almohad conquests are well documented as being a catastrophe and a cause of economic decline of the region. Despite this, the economic exchanges between Tunisia, Italy, Spain and Syria continued, but some merchants with their families were obliged to move to Sicily and Egypt. Moreover, in the 11th century, the repeated pillaging by the Banū Hilāl obliged the inhabitants of the villages to seek safety within the walls of the surrounding towns; besides, some nomadic populations have returned to the steppes (Decret 2003). These population movements, only in part caused by the Hilalian invasion, must have had an important influence on the Tunisian Arabic language, regarding which we currently have scanty or no data. Mion (2015, pp. 275–76) distinguishes the two phases of the Arabization of Sahel and underlines some relevant historical events which had linguistic consequences: In the pre-Hilali period, the region was probably inhabited by some sedentary Arabized people, and during this phase, the main urban features of Sahel Arabic developed. The Hilalian invasions of the 11th century deeply troubled the region to the extent that many urban centers and villages were threatened and disappeared, so that Ibn Ḥaldūn (d. 1406), in the 14th century, writes that only some weak traces of sedentary culture could be found in some families from Kairouan or Mahdia. A relevant piece of information, linked to the population movement in Mahdia region and its consequences on the Arabic of the region, is that “en 945–46 Ismā’il al-Manṣūr [. . .] quitta la ville de Mahdia pour établir sa résidence à Sabra, en provoquant la ruine de l’ancien siège de l’empire fatimide et la perte des habitants et ses faubourgs, ce qui nous incite aujourd’hui à voir en cela un événement qui laisse le champ libre, plus tard, à un repeuplement de la part de gens beaucoup moins urbanisés” (Mion 2015, p. 275).

If we continue our imaginary journey, guided by al-Idrīsī, we find Monastir (al-Munastīr) at 30 miles of navigation from Mahdia. The geographer reports only that the town has some castles in which some fruits are produced and then exported to Mahdia and that the inhabitants of the latter bury their dead in the cemetery of Monastir (Bresc and Nef 1999, pp. 184–85). Actually, after the Arab conquest of the mid 8th century, Monastir became renowned for religious reasons because of its *ribāṭ* and its cemetery in which important personalities, such as the last members of the Zirid dynasty, were buried.

As recounted by the Imām al-Māzarī (d. 1141), Monastir and Mahdia appear to be very connected and seem to have been spared from the Hilalian invasions. Monastir was described as a prosperous town whose religious importance was proved by the pilgrimage

of numerous people from the nearby regions, similarly to what happened in the holy city of Kairouan (Soucek 1993, pp. 227–29).

Monastir is a town located on the southern end of the Gulf of Hammamet, about 160 km south of Tunis, and today is part of the Wilāyat al-Munastir (see Figure 2). Its main commercial activities are tourism, the textile industry for wool processing, the production of salt, which is an ancient activity already described by al-Bakrī (d. 1094) due to the fact that the town was built near a salt pan, soap, olive oil and fishing. Monastir is also a university town (for further details on the history of Monastir, see (Soucek 1993, pp. 227–29)).¹¹

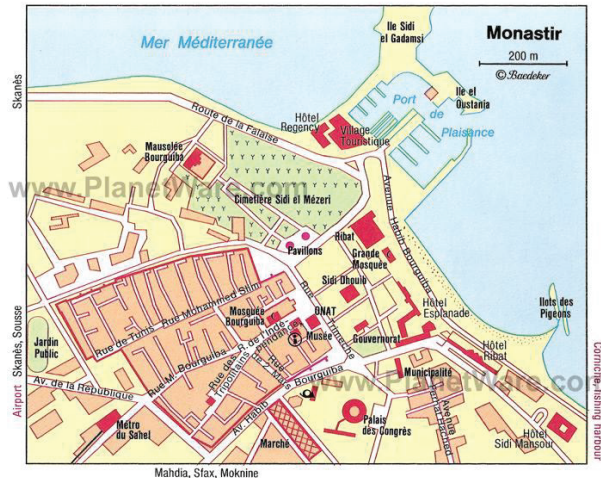


Figure 2. Map of Monastir.

Msaken (Arabic: M’sākin) is a small town of the Tunisian Sahel located at a dozen kilometres to the south of Sousse. Administratively dependent on the governorate of Sousse, it has an increasing number of inhabitants in the summer due to the return of expatriates, most of whom work in France.¹² For this reason, its inhabitants call Msaken “la Petite Paris”. Its economy is based on olive oil production. There are, however, a number of handicraft and industrial enterprises in the surrounding area, set up by old emigrants who returned to Msaken in the 1980s (Ma Mung 1984).

Msaken was founded during the Hafsid dynasty in the 14th century (Bouhlel 2009, p. 125). A brief description of Msaken as a holy place in Tunisia together with Kairouan is contained in *Tunisie et tunisiens* by François Bournand (Bournand 1893, pp. 311–13) who actually quotes the information offered in *Promenades d’une Française dans la régence de Tunis* by Voisins d’Ambre (1884, pp. 171–74). Bournand states that Msaken is a small town of about 9000 inhabitants, 9 km from Sousse, built in a slightly mountainous area planted with olive trees, famous as a religious place because of the *madrasa* of Sīdī ‘Alī b. Ḥalīfa, renowned in North Africa because of the high number of students and the high level of teaching of its university. Msaken is considered by the author to be like Seville, Padua, Oxford or Cologne. The town occupied a large area with houses surrounded by greenery, while the centre of the town developed along the main road and had several schools and mosques. According to M.me de Voisins, entry to the holy city was forbidden to Christians and especially Europeans. For this reason, the author claims to have seen Msaken from a nearby hill and that she found it fascinating to the point that she was very impressed by it. Then, she gives a rather exotic and Eurocentric description of the town (Voisins d’Ambre 1884, pp. 171–74; Bournand 1893, pp. 311–14).

Ma Mung (1984, p. 163), stated that in 1984, Msaken had 41,219 inhabitants and that it was among the largest 15 or 16 towns of Tunisia, together with Mahdia. In his study,

he focused on the importance of the migratory movement that affected a large part of the population from the 1960s onwards, which had important effects on the economy of the country. Today, Msaken is the second town of Sahel in terms of size; its inhabitants constantly move back and forth to the rural and village hinterland as well as to Sousse. The growing integration of this town into the economic space of the regional metropolis, Sousse, means that it has become a basin of intense activity (industrial, construction and services) and, above all, of employment for the populations of the surrounding areas. The main categories of commuters are those employed in industry and government, especially teachers (Boubakri and Lamine 1992).¹³ According to Bouhlel (2009, p. 126), Msaken is close to several small towns that show different linguistic features from each other as well as some common features. The scholar pointed out that Msaken Arabic has some peculiarities that distinguish it from the other Tunisian varieties. Moreover, this variety is in continuous evolution because of its inhabitants' migration movements, and therefore the "original" Msaken variety would be spoken only by aged people, children, emigrated and housewives.

2. Materials and Methods

This paper offers some preliminary results of the linguistic analysis of the dialect(s) of the Wilāya of Mahdia, based on a corpus of interviews taken from some social media pages, which will be presented below. The sample is composed of speakers of different geographical origin, gender, age and social background. The videos analysed have been collected between 2019 and 2021 and the speakers speak varieties of Tunisian Arabic that are more or less dialectal in the sense that they can be more or less influenced by Modern Standard Arabic, and may have mixed features, depending on the the speakers' levels of education and their role (i.e., politicians and teachers tend to use a higher register).¹⁴ Moreover, their speech can be more or less influenced by the medium used, i.e., radio or video. The role of the Arabic of the capital Tunis and of the main city of Sahel, Sousse, will be highlighted too.¹⁵

Starting from the previous studies on Arabic dialectology, and particularly of the Maghribi area and on Tunisian Arabic, I will observe and analyse some of the urban, Bedouin and village isoglosses indicated by Marçais in his studies (Marçais 1950, pp. 207–14) and I will then add some selected morpho-syntactic elements attested in the Arabic of Mahdia.

Even if the focus of this study is the urban vernacular of Mahdia, some linguistic elements of some nearby towns will be underlined in order to follow the suggestions of earlier scholars. For this reason, some video-recordings of speakers from some small villages near Mahdia, such as Tlelsa and Teboulba, have been analysed as well as audios of respondents from Monastir and Msaken, since, according to Marçais, the two towns have village features (Marçais 1950, p. 207). Similarly, Saada's statement "On a recueilli en outre des informations concernant le parler de Bqalta¹⁶ qui ne suffisent pas à le classer (parler de Musulmans)" attracted my attention.¹⁷

Therefore, several dialogues and interviews contained in numerous online radio and tv programs and social media information pages have been analysed. Due to the difficulty of determining the real background and origins of the speakers involved in the videos available online, deriving from the analysis of online material, I have selected mainly the materials in which the origin of the speakers has been specified by the speaker or the interviewer.

Sources for Mahdia and Monastir Arabic

For linguistic data on Mahdia and Monastir, I consulted the following online radios that broadcast on their relative Facebook pages:

- * *Menara Fm* (<http://www.menarafm.net/> accessed on 19 August 2021) is a radio station based in Mahdia since 2019 whose editorial line is based on independence and freedom of expression. It offers many interviews with Tunisian artists and craftsmen

and those from other countries in the Arab world, as well as reports, news and live radio broadcasts focused on debates on topical issues chosen by the radio speakers. It also broadcasts some fixed radio programs, such as *Nahārek zīm*, i.e., “Have a nice day”, in which every morning some presenters give the weather forecasts and the main news of the day.

- * *Radio MFM Tunisie* (<http://www.mfmtunisie.net/> accessed on 19 August 2021) is a generalist radio station that broadcasts in the regions of Mahdia, Sousse and Sfax and offers programs on sport, news, music and society. It also has daily news programs such as *Hadret l-youm*, i.e., “Conversation of the day”, and *Mid Mag*, broadcasted at midday.
- * *Radio Mahdia* (<https://www.radio-mahdia.info/> accessed on 19 August 2021) broadcasts almost exclusively news programs.
- * *Radio Mahdia 1* (<http://www.radiomahdia1.com/> accessed on 19 August 2021) is specialised in tourism and cultural issues. It offers news programs, interviews of the citizens made on the road and interviews of local artists taken in the radio studio, and has a program devoted to children, called *Matinal children*.
- * *Dār al-šabāb Ražīs* (<https://tinyurl.com/2fmb88fx> accessed on 19 August 2021) is a web radio based in Mahdia, and particularly in Rajiche, a small coastal town in the Mahdia Governorate, created under the patronage of the Ministry of Youth and Sports of the Tunisian Republic. Among the programs broadcast, *Mahdia taḥki*, that is “Mahdia speaks”, is interesting since it often brings together people from different generations who discuss a cultural or social topic. This helps analyse the language of speakers of different ages and sex.
- * *Radio Monastir* (<http://www.radiomonastir.tn/> accessed on 19 August 2021) not only provides news from Wilāyat al-Monastir, but also from all Tunisia, and was very useful for analysing the speech of natives of Monastir and Mahdia. It offers news programs, reportages and topical talk shows.
- * I also analysed videos taken from the Facebook pages:
- * *Revolution Mahdia*, which is a page open to all the inhabitants of Mahdia region whose mission is to foster and promote the freedom of expression and circulation of information in the zone. This is often done through the broadcasting of interviews of the citizen taken on the road.
- * *News Mahdia* and *Ville de Mahdia*, both broadcasting news, social and entertainment programs.

Sources for Msaken Arabic

News on Msaken is provided by the online tv channel *Msaken Tv* and by *Radio RM FM*, which is a radio station based in the town broadcasting entertainment programs and news. Both broadcast in their Facebook pages.

Sources for Bekalta Arabic

As regards Bekalta, I analysed some videos from the Facebook pages *100% ba9louti* and *Bekalta Today*, which provide daily news, reportages and interviews about the town.

Except for the programs involving children and young people mentioned earlier, the speakers involved in the radio programs analysed are men and women of all ages.

When analysing an oral corpus on media or on social media, the researcher should be aware of many challenges of such a work. In fact, as Van-Mol (2010, pp. 67–68) states, “people of all layers of society appear in these media to participate, by means of oral expression, which reflects a wide variety of language capabilities and layers”. Van-Mol also distinguishes two main classes of speakers: The first is made of professional workers, such in the case of radio presenters in my corpus, or intellectuals who use a top-down strategy of communication by using a higher register with some dialectal features in order to also be understood by less educated people. The other is composed of less educated people and non-professional workers, i.e., the majority of the speakers in my corpus. However,

even if, as Van-Mol states, oral media Arabic gives us the opportunity to observe how people with different linguistic competences communicate with each other, many elements that can influence the linguistic variety and register used by speakers should be taken into account. First of all, the kind of media and of program in which they are involved influences their linguistic choices as far as regards the register and the spontaneity of their speech. Furthermore, the different linguistic competences of young and aged people and of women and men (i.e., linked to age and sex) have a role in their linguistic production. In addition, the topic dealt with and kind of oral text produced by them, such as dialogues, interviews, monologues and multilogues, are important. Moreover, the audience who is going to listen to the speakers has to be taken into account because it has an influence on the speaker's linguistic choices.¹⁸

In spite of these challenging aspects, conducting a linguistic analysis through social media resources has some advantages. Firstly, it offers the opportunity to observe many examples of informal communication among several individuals. In fact, it is possible to reach many speakers in different locations, such as small towns and villages thanks to their social media pages. Secondly, it allows research to continue despite the sanitary crisis by remotely studying a linguistic variety and formulating hypotheses that will then be verified during the fieldwork.

In fact, even if using social media resources for linguistic analysis has some advantages, it cannot substitute fieldwork. For this reason, the first results presented below will be compared and integrated with those obtained from a period of field research in which I will record some interviews. This was the very first aim of my research, but due to COVID-19, reaching Tunisia has become impossible, at the moment. The data provided by the in-field interviews will show further elements useful in verifying the hypothesis formulated and will provide a more precise description of the varieties used in the Wilāya. During the field work, eventual phenomena linked to diatopic and diastratic variation will be highlighted in order to point out the possible local variants belonging to any social group such as older and younger people, men and women. Moreover, any possible diachronic element that might have contributed to the formation of Mahdia Arabic will be identified and described. This will be done by choosing a sample of speakers that is as varied as possible and by consulting some Arab historical and geographical works.

Some selected phonological, morphological and syntactic features will be presented below.

3. Results

3.1. Phonological Remarks

3.1.1. Interdentals

It is well known (among others, (Marçais 1950, p. 201; Attia 1969, p. 22; Yoda 2008, 2019)) that in Mahdia, Arabic interdentals $\underline{d}/$, $/\underline{t}/$ and $/\underline{d}/$ are substituted by the respective dental consonants $/d/$, $/t/$ and $/d/$. The data collected through the online social media pages analysed are not so unequivocal and show a certain variation.¹⁹

$/\underline{t}/$

There were several cases of dental realisation of $/\underline{t}/$ in $/t/$ in initial and median positions: *istitnā'iyy*, *istitnā'iyya* "exceptional", *aktar* "more", *al-tēni*²⁰ "the second", *ymattəl* "represents/plays", *ātār* "antiquities, ruins" and the plural *ātārīn*, *el-ǧim'a l-tēni wella l-tlēta* "the second or the third week". The interdental phoneme is also attested, such as in *tawra* "revolution", *iṭnīn* "two (m.)" *ṭmāniya* "eight", *al-tēniya* "the second", *ṭlēta* or *tlēta* "three (f.)", *aktar* "more", *miṭṭel* "example", *ṭnāš* "twelve" and *kāriṭiyya* "catastrophic". In some cases, the pronunciation of dental/interdental is present in the same informant: *miṭṭel* "example", *ṭnāš* "twelve", but *ymattel* "represents/plays" and *tlātamiya* "three hundreds", *ṭmāniya* and *tmāniya* "eight (f.)". D'Anna (2020, p. 88) states that in the word *tlāta*, only the second interdental is preserved, but in my corpus, there is a certain oscillation in the realisation of $/\underline{t}/$ in this word.²¹

The general tendency seems to show that initial /t/ is more likely to be preserved. In my opinion, the trend in the preservation of the interdental /t/ in the Arabic of Mahdia may be due to the influence of Tunis Arabic, which has it. In fact, the Arabic of the capital applies a centripetal force to the other “peripheral” varieties of Tunisia because of its increasing prestige, due to the fostering of travel and education during the recent years (Gibson 2002, pp. 29–30). Media and social media, on which several programs are broadcast in *dāriġa*, have a relevant role in this process too. In general, as Gibson showed, Arabic dialects do not follow the general trend affecting many other languages of the world, that is they do not tend towards a levelling to Modern Standard Arabic, but they are inclined to be closer to the standard dialect variety of the region, such as in the case of Tunis Arabic.²²

/ d /

The phenomenon is less evident in the voiced interdental consonant, which, however, sometimes oscillates with /d/: *asātida* “professors”, *mādabiyya* “I want/I would like to”, *hādāka* “that (m.)”, *hāda* (Yoda 2019, p. 59), which alternates with *hāda* “this”, *hēdi* “this m.”, *hēdiyya* “this (f.)” (attested in Tlelsa; see also Yoda 2019, p. 65; Attia 1969, p. 123). The same phenomenon is extended to all the demonstratives containing a /d/. In the same speaker, we find the two realisations of the consonant in the same utterance: *kadē w-kadē* “so and so”, *hadēya* but also *hādiya* “this (f.)” *hāda* “this (m.)”, *nahdāw* “we take”, but *mādabiya* “I want/I would like to”. This feature is attested in a wide audience, different by age and sex and has also been attested by Yoda, for instance, *hdē* “he took” (Yoda 2019, p. 63).

In Sousse, the verb *hdā* loses its final /d/ in the imperfect conjugation (Talmoudi 1980, p. 93), and some examples are also available in Mahdia Arabic: *yāḥū* “he takes”. The phenomenon is also attested by Yoda (2019, p. 63) and in Bekalta too, where we find, for instance, *nāḥū aḥna* “we take it”. The weakening of /d/ is an ancient phenomenon, attested in Sicilian Arabic where it sometimes even passed to /l/ and the consonant has not been pronounced in Maltese Arabic since the 14th century (La Rosa 2019, pp. 116–17; Avram 2012, p. 102).

In Monastir and in Bekalta, interdentals are generally maintained, but in the latter, a certain oscillation, depending on free variants and the speakers’ levels of education, is attested in the following cases: *aktar min talātina* “more than thirty”, *nāḥdu* “we take”, *yāḥduh* “they take it”, *nāḥduh* “we take it”, but in the same speaker we find *hādī* “this (m.)” and *hādāya* “this (f.)”, *hādiya* “this (f.)”, *hādāka* “that (m.)”, which alternates with *hādāka*, and *hdīna* “we took”.

3.1.2. /q/

In Mahdia, Monastir, Msaken and Bekalta, /q/ is generally pronounced as voiceless, but it might be realised as /g/ in the following examples: *gālū lī* “they told me”, *gutlak* “I told you”, *sil’ a mangūla* “transported goods”, *bagra* “cow”, *mungēla* “watch”, *ynaggəz* “he jumps”, *gā’ dīn* “[the ones] staying”. The phenomenon was noticed also by Attia (1969, pp. 125–26) who offered few examples that have been already mentioned above. As regards Msaken, Bouhlel observed that the inhabitants of the town generally pronounce the uvular stop /q/ as voiceless because they consider themselves as “citizens”, such as in the following words: *iqalla’* “eradicate”, *yūq’ud* “he sits down”, and *qalb* “heart”. According to Bouhlel, the consonant is pronounced as /g/ in some words, such as *mungēla* “watch” or *bilgəz* “well”, and in his opinion this would be the “real” pronunciation of the consonant. Bouhlel also shows a certain variation in the pronunciation of some terms, such as *zqūqu* and *zgūgu* “pine nuts”, *baqrī* and *bagrī* “veal” (Bouhlel 2009, p. 127). Actually, none of the Tunisian dialects know the exclusive use of /q/ and/or /g/. If compared to the spread of /g/, the realisation /q/ is in fact limited and restricted only to the cities of Bizerte, Tunis, Sousse, Monastir, Mahdia, Sfax, and Kairouan. Mahdia is, therefore, “entouré de g” (Mion 2015, p. 271); see the maps in (Skik 2000).²³ Moreover, Yoda (2008, p. 484) stated that the appearance of /g/ instead of /q/ in some words in Mahdia Arabic, as well as in

other Tunisian dialects, is very common.²⁴ In addition to this, already in 1984, Saada (1984, pp. 27–28) talking about the Arabic of Tozeur, stated that the existence of a /g/ sound is not a valid criterion for establishing the “non citadinité” of a dialect in Tunisia. She also added that throughout rural Tunisia, the phoneme /g/ was used for /q/, but that it was also present in some coastal zones and towns, in Judeo-Arabic and Muslim varieties, even if the Judeo-Arabic varieties of Tunisia are usually an exception and all have /q/, except Tataouine.²⁵

Free Variants

/k/ as /q/

In some aged inhabitants of Mahdia, among which are some fishermen, natives of the quarter Burż al-rās, /k/ is articulated as /q/ in few words or expressions, such as *kif kif* “the same”, pronounced *qif qif*, and *flūqa* “felucca, rowing boat”, generally attested as *flūka* in Tunisian Arabic.²⁶ The term *flūka* is pronounced with /k/ three times in the same video by two other speakers, that is a Mahdāwī poet, while reciting a poem dedicated to Mahdia, and the speaker. Both used a higher register: The poet because of the nature of the text that he was reciting and the speaker because of his role of presenter of the documentary. Therefore, their roles may have influenced their linguistic choices. Regarding previous studies on Mahdia Arabic, Attia (1969, p. 125) did not highlight this feature, but identified an emphatic variant of /k/ indicated with /ḳ/, such as in *kāni* “I am”.

3.1.3. /ṛ/

The phenomenon is attested in some aged fishermen natives of Mahdia, such as in the words *buṛż*, of the toponym Burż al-rās, *bḥar* “sea”, *raḥma* “mercy”, *aktar* “more”, *barḥa* “out” and *dār* “house”. The same informant may also pronounce some words in which /r/ is not emphasised, such as *ḡārek* “your neighbour”. The phenomenon has also been attested in some other mid-aged speakers, women and men, but not in all the speakers of the video analysed. The realisations of /r/ as /r/ and /ṛ/ may also oscillate in the same informant, a forty-year-old female nurse from Mahdia, such as in the word *arba* and *arba* “four”. The emphatic /ṛ/ is also attested in Mahdia Arabic by Yoda (2008, 2019). The presence of /ṛ/ does not seem to allow fronting or raising of /a/ and /ā/. In the speakers involved in the videos analysed who do not pronounce /ṛ/, /a/ is maintained as well and is not raised nor fronted, such as in *arb'a* “four”, *dinār* “dinar” and *barra* “out”. However, further investigation is required to verify the presence and the distribution of the phenomenon since it has been attested in few speakers, at the moment.²⁷

Regarding Msaken Arabic, Bouhleb (2009, p. 127) stated that, differently from some other regions of Tunisia, the *rā'* is generally not “amplifié”, and is even pronounced *re*, such as in the words *ureq* “leaves”, *krēheb* “cars”. Bouhleb added that some exceptions are available, such as *mrā* “woman” and *ḥrābiš* “pills”, but he does not explain in which conditions the phenomenon occurs. In my corpus, some cases of emphasisation of /ṛ/ in Msaken Arabic have been attested, as well as the raising of /ā/ after /r/, such as in *krēheb* “cars”, as indicated by Bouhleb.

3.1.4. De-Emphasisation

Some cases of loss of emphasis have been attested in Mahdia and Bekalta in speakers of different sex and age, such as in the words *atfāl* “children”, *tbīb* “doctor”, *musāba* “infected” *durūfa* “conditions”. In the latter, we notice the possible shift /ḍ/ > /d̥/ > /d/ because of the loss of the interdental articulation. Some other examples are *tufultī* “my childhood”, *bi-sifa* “in quality of”, *abyad* “white”, *tul tul* “directly, straight”, *muwātin* “citizen”. The phenomenon involves all the emphatics /ḍ/, /ḏ/, /ṭ/ and /ṣ/ and does not seem to affect the vowel quality, which is maintained. In the corpus of Bekalta, we also find some examples of de-emphasisation and sonorisation such as *mazdar* “source” and *mazraḥī* “theatral”, in which there is a shift /ṣ/ > /z/ with a possible intermediate step of /ṣ/ > /s/. The phenomenon of de-emphasisation is also attested in Kairouan, where we find, among

other words, *matālib* “requests”.²⁸ In a woman of Monastir, we find a de-emphasis of /t/ realised as lightly affricated /ts/ in the exclamation *yā latif!* “oh my God!”. A light affrication of /t/, already highlighted by Cantineau (1960, p. 37) as a feature attested in Algerian and Moroccan Arabic, generally also characterises Tunisian Arabic, but here involves emphatic consonants too.²⁹ Saada (1984, p. 24) identified a /ts/ sound in the speakers of some tribes of Tozeur and Dallaji (2017, pp. 153–57) reported that in Nabeul there is a tendency to affricate /t/ that people define through the verb *taštaš* and the terms *taštīš* and *taštaša*. Furthermore, Maamouri (1967) described the variety of Nabeul as being characterised by a strong affrication of /t/.³⁰ As regards the loss of emphasis, it is also typical of Tunisian Judeo-Arabic (Taïeb and Sayah 2003). Cohen (1975, vol. 2, p. 14 and n. 7) states that some phenomena in Tunis Judeo-Arabic lead us to think that there must have been a period in which the articulation of emphatic consonants was stronger and “sans doute plus forte qu’à Tunis musulman, ce qui expliquerait le fait que les musulmans qui veulent imiter le parler des juifs, exagèrent l’emphase en même temps que les modulations expressives de la phrase”.³¹ Cohen’s statement also shows that in Tunisia the weakening of the articulatory strength in pronouncing the emphatics has long been attested. However, the pronunciation of emphatic consonants with more or less articulatory strength from different confessional groups, i.e., Jews/Muslims, is a phenomenon of old, present in different Arab contexts because of community or identity reasons.³² Saada (1984, pp. 83–84) noticed the phenomenon in the Arabic of Tozeur and indicated a series of conditions for its occurrence. Further studies will be necessary to be able to indicate which are the causes in Mahdia Arabic and if any regularity can be found.

3.1.5. Dropping of Final /n/

In Monastir and Mahdia, the weakening or dropping of the final /n/ of numerals in pausal position, that is not followed by the *ism al-ma’ dūd*, should be noted: *‘išrīn milyū* “twenty millions”, *sab’at w-‘išrī* “twenty-seven”, *ṭmānī* “eighty”, *ḥamsī* “fifty”. The dropping of the final /n/ is an old feature, attested in Andalusī Arabic only in dual nouns (see (Ferrando 1995, pp. 50–51; Corriente et al. 2015, pp. 110, 125–26). It also should be noticed that in Yoda’s transcriptions, this feature does not emerge (see, for instance (Yoda 2019, p. 63)). In Saada’s study on the Arabic of Tozeur, some consonantal dropping involving final /n/ have been attested, such as the cases of *ṭnē* “two”, *mnē* “from where”, *lē* “until when”, *ga’ dī* “standing”, *takū* “you will be”. The scholar defines this phenomenon, which in Tozeur is wider spread than in Mahdia and involves some other consonants, *tarḥīm*, i.e., “softening” or the process of shortening a word (for a more complete definition, see (Carter 2007, p. 17), and states that only one informant uses it and therefore judges this feature as in decline/regression (Saada 1984, pp. 39–42).

3.1.6. Vowels

According to Attia (1969, pp. 126–27), Mahdia Arabic has five short vowels /a/, /e/, /i/, /o/, /u/ and five long vowels /ā/, /ē/, /ī/, /ō/ and /ū/.³³ Yoda (2008, 2019) stated that Mahdia Arabic has three short vowels, /a/, /i/, /u/, and five long vowels, and assumed that Mahdia dialect, as an urban variety, acquired /ē/ and /ō/ as phonemes because of the influence of the surrounding village dialects. He also added that “Because of the existence of /ē/ and /ō/, Mahdāwī dialect is unique among the Tunisian sedentary dialects” (Yoda 2008, p. 489).³⁴ Therefore, Mahdia Arabic vowel system would be made of five long vowels and five short ones for Attia, but only of three short vowels for Yoda.³⁵ Starting from this vowel scheme, Mahdia Arabic long vowels can undergo some qualitative changes; i.e., some cases of opening of the close vowels /i/ and /u/ are present in the varieties of the region, such as in Bekalta, where we find both *zīt* “oil” and the variant *zēt*, or *hōnī* “here” in Monastir (Attia 1969, p. 129) and a certain variation is also attested in Msaken, such as in the cases of *rōh* “soul” and *fūl* “broad beans”. Bouhleb (2009, p. 128) also provides some minimal pairs: *qūm* “stand up” / *qōm* “people” pejorative, *kūn* “be” / *kōn* “world”.

In this section, I shall focus exclusively on the treatment of /i/ when it is followed by the third person singular suffix pronouns. What emerges is that, in general, /i/ + hā and /ū/ + hā are maintained, such as in *mā fī-hāš* “there is not” (as well as in Takrouna Arabic see (Marçais and Guiga 1925, p. XII, n. 2) and in Sousse, (Talmoudi 1980, p. 152)). However, some cases of *fē-hā* “in it” have been attested (Yoda 2008, p. 489; 2019, pp. 61, 65, 66). In the videos analysed in my corpus, I found only rare cases of *fē-h* “in it (m.)” and of *fē-ha* “in it (f.)” in a speaker from Knaies, a small village 11 km near Msaken, and in a small group of people allegedly from Mahdia. The majority of the people speaking in the videos analysed, coming from Mahdia or Monastir, uttered *fī-h* or *fī-ha*. The reduced presence of the phenomenon may be due, once again, to the linguistic levelling towards the Arabic of Tunis or of the other coastal towns of Sahel such as Sousse (see (Mion 2015, p. 272)) or to the conditioning of the medium used, which may push the speaker to use a higher register.³⁶

Therefore, the vowel system of Mahdia seems similar to that of Monastir, which, according to Gibson (1998, p. 276), has *villageois* features since it has five long vowels as in Figure 3.

Villageois (Jemmel, Monastir)

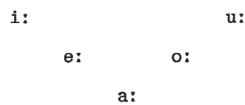


Figure 3. Jemmel and Monastir Vowel system, taken from (Gibson 1998, p. 276).

According to Marçais, Takrouna Arabic has “une particularité de détail propre aux dialectes sâhli: l’ouverture en ē et en ō de ī et de ū” that we find in Mahdia too (Marçais and Guiga 1925, pp. XXI–XXII).

3.1.7. Diphthongs

According to Yoda (2008) and Mion (2015, p. 271), the vowel system of Sahel Arabic is characterised by five long vowels, ā, ē, ī, ō, ū, obtained from the monophthongisation of the diphthongs /ay/ and /aw/ in ē and ō. According to the traditional categories of classification, which contrast some allegedly Hilali and pre-Hilali features, in the Hilali system, /ay/ and /aw/ would be reduced to ē and ō, while the passage of /ay/ > /ī/ and /aw/ > /ū/ would be typical of the pre-Hilali dialects. Actually, the situation is more complex than this, because, as Mion (2015, p. 271) showed, the reduction of the diphthongs to ē and ō “is shared by both /q/ and /g/ varieties”. As already dealt above, the pronunciation of /q/ has been the main criterion for establishing whether a dialect had Bedouin or sedentary features, but Mion and some other scholars have raised doubts about the validity of this principle.³⁷ Based on what emerges from my corpus, the realisation of etymological diphthongs is usually sedentary in Mahdia, in Monastir and in Bekalta, such as in the following cases: *al-yūm* “today” (Tielsa), *ḥīr* “better”, *ḍīf* “guest”. Even if Yoda (2019, p. 59, n. 9) states that the diphthongs in Mahdia are usually reduced to /ē/ and /ō/, the cases identified in my corpus are similar to those of Sousse, where the diphthongs /ay/ and /aw/ are reduced to /ī/ and /ū/ (Talmoudi 1980, p. 55). Some other existing studies on Sahel varieties all go in the same direction; for instance, to recount it with Marçais’ words regarding the Arabic of Takrouna, “Comme les dialectes congénères du Sâhèl, le takrouni y suit sur quelques points des voies propres qui, l’éloignant des parlers citadins de la Régence, le rapprochent de certains dialectes bédouins du Maghréb oriental [. . .]: les diphtongues anciennes ai et au accentuées et non en finale absolue s’y réduisent généralement à ē et ō” (Marçais and Guiga 1925, pp. XXI–XXII).³⁸ As D’Anna (2020, pp. 88–89) showed, in Chebba, a small village 35 km south of Mahdia, etymological diphthongs

are reduced to /ē/ and /ō/ and are occasionally partially diphthongised. According to Bouhlel (2009, p. 128), in Msaken Arabic, the diphthongs /aw/ can be reduced to /o/ in many terms, i.e., *lōn* “color”, *mōz* “bananas”, *dōra* “tour”. The situation seems to be quite complex since, once more, it is clear that Mahdia Arabic is an urban dialect surrounded by mixed or contact varieties, which can influence it in different ways and at different levels.³⁹

3.1.8. Imāla

Several cases of final *imāla*, etymological or not, have been found in all positions in Mahdia and the nearby towns: *hmē* “here”, *‘alē* “on/upon”, *mšē* “he went”, *ḥdē* “he took” ((Yoda 2019, pp. 59, 63); see also (Attia 1969, p. 130)), At the moment, the pronoun *ēni* “I” is the only case of *imāla šadīda* found in Mahdia and Monastir. (Bouhlel 2009, p. 127) added some other examples attested in Msaken Arabic, such as *mī* “water” and *smī* “sky”.

Imāla is a phenomenon shared by many Tunisian vernaculars. Gibson (1998, p. 279), for instance, analysed the pronunciation of the verbal forms *mšā* “he went” and *mšāt* “she went” and stated that in Sousse the average vowel of the feminine form is /ε/ and those of the masculine form are /ε/ and /e/. He added that among his Sousse informants, nobody uttered a long /e/, while in Monastir, there was only one speaker with /ē/ as the tense vowel. Marçais affirmed that in Takrouna Arabic “ā accentué en finale absolue y passe à peu près constamment a ē” (Marçais and Guiga 1925, pp. XXI–XXII). The feature of *imāla* is also shared by Tunisian Judeo-Arabic, the /ā/ is affected by *imāla* in all positions (Cohen 1975, vol. 2, p. 56).⁴⁰

3.1.9. Raising of Final—a

Mahdia Arabic is also characterised by the raising of the final feminine singular ending –a such as in the cases of *maḥdūde* “limited”, *ḥamse* “five”, *insaniyye* “human”, *bnayye* “child (f.)”. There is a role played by syntactic elements since we also find *ḥamsa ṣḡār* “five children”, with the numeral in *iḏāfa*. The presence of emphatic and pharyngeal consonants usually blocks this raising.

3.2. Nominal Morpho-Syntax

3.2.1. Personal Pronouns

The existing studies on the personal pronouns of Tunisian Arabic show a complex situation, which can be summarised in Figure 4:

1 ^{ère} pers.sg.:	
<i>āna</i> :	Tunis et tous le Nord-Est; Sfax.
<i>āni</i> :	Sousse et le Sahel.
<i>yāna</i> :	Kairouan.
<i>nāy(a)</i> :	Le Kef (Nord-Ouest extrême) ; zones rurales et bédouines du Sud.
2 ^{ème} pers.sg. :	
<i>anti</i> (m/f) :	Tunis, Sousse, Kairouan, Sfax.
<i>anta</i> (m) / <i>anti</i> (f) :	toute variété rurale ou bédouine du pays.
1 ^{ère} pers.pl. :	
<i>aḥna</i> :	Tunis, Sousse, Sfax.
<i>naḥna</i> :	Le Kef, Mateur, Kairouan.
<i>ḥnā</i> :	La Chebba (Sahel), Gabès.

Figure 4. Distribution of the first p. sing., second p. sing., first p. plu. of the personal pronouns in Sahel Arabic (Mion 2015, p. 274).

Mion, in his paper *Réflexions sur la catégorie des «parlers villageois» en arabe tunisien*, offered a brief sketch of the first singular and plural personal pronouns “I” and “we” and of the second singular person “you” and focused on gender opposition in the pronominal system. If the village and Bedouin dialects share the presence of gender opposition in the second person singular, according to Marçais’ classification, the urban varieties do not (Marçais 1950, p. 208). With respect to this element, in Mahdia Arabic, there is no gender

distinction in pronouns and verbs, as well as in Monastir, Msaken and Bekalta. So, the situation could be represented as follows in Table 1.

Table 1. Personal pronouns in Mahdia, Monastir, Msaken and Bekalta.

Pronoun	Mahdia	Monastir	Msaken	Bekalta
1 p. s.	āni	āni	āni	āni/āni
1 p. p.	aḥna/naḥna	aḥna	aḥna/naḥna	naḥna
2 p. s.	ənti/inti ⁴¹	ənti	ənti	ənti/inti

As for the first-person plural “we”, *aḥna* is the main pronoun used and *naḥna* seldom alternates. Therefore, further studies are needed to check whether this is due to the origin of the speakers interviewed in the videos analysed. In Yoda (2019, p. 65), *ḥna* is also attested when preceded by a vowel such as in the expression *tawwa-ḥna* “now we”.

The first person singular “I” generally shows a final *imāla šadīdā*.⁴² Moreover, Takrouna Arabic shows the pronoun *āni* “I”, which according to Marçais is original to the variety, and also presents the variants *nāya* and *nāy*, which have been imported later (Marçais and Guiga 1925, pp. XXII–XXIII).⁴³ As regards Msaken, Bouhleb (2009, pp. 130–31) indicates *āni* with the meaning of “I” and “me” and the second person singular pronoun *inti* “you”.

3.2.2. Relative Pronouns

In my corpus, in all the towns and villages considered in this study, the only relative pronoun used is *ellī* and the variants *illī* ((Yoda 2019, p. 61) has *illī*) and *lī* when the preceding word ends in a vowel. The relative pronoun *ella*, defined as “villageois” by Marçais, has not been attested at the moment ((Marçais 1950, p. 211); for Chebba, see (D’Anna 2020, p. 90)). Other varieties of Tunisian Sahel, such as the dialect of Sousse, also have the variants *ellī* and *llī*, used when the relative is preceded by a word ending in a vowel (Talmoudi 1980, pp. 146–47).⁴⁴

3.2.3. Numerals

In numbers 3 to 10, the *ism al-ma’ dūd* is plural: *‘ašra yyām* (Yoda 2019, p. 66), *xams iyyām* “five days” and *tlāta yyām* “three days” (Yoda 2019, pp. 60, 63), *tlāta ‘askar* “three soldiers” and *arb’a ‘askar* “four soldiers” ((Yoda 2019, p. 59); for Tunis Arabic, see (Biṭuna 2011)).⁴⁵ As already noticed by Biṭuna for Tunis Arabic, the noun *dīnār* “dinars” is always used as a singular in Mahdia Arabic too (see (Biṭuna 2011) and (Yoda 2019, p. 63)).⁴⁶

In Mahdia Arabic, numerals from 11 to 19 have the same form that they have in Tunisian Arabic in general, that is *ḥdāš* “eleven”, *tnāš* “twelve”, *tluṭṭāš* “thirteen”, *arba’ aṭāš* “fourteen”, *ḥmuṣṭāš* “fifteen”, *səṭṭāš* “sixteen”, *sab’ aṭāš* “seventeen”, *tmūnṭāš* “eighteen” and *tsa’ aṭāš* “nineteen” (some numerals are present in (Yoda 2019): *passim*). One of the main evident characteristics is the disappearance of the pharyngeal /‘/ or its assimilation to the /t/ (Biṭuna 2011, p. 32). These numbers from 11 to 19 are usually followed by a singular, such as in *tnāš ‘ām* “eleven years” and *ḥmuṣṭāš ‘ām* “fifteen years”. It is worth noting the form *tluṭṭāšən el-snā* “eighteen years”, followed by a determined singular noun. Some other examples, taken from Yoda’s texts, are *xmuṣṭāš-in yōm* and *xmuṣṭāš-in yōm* “fifteen days” (Yoda 2019, pp. 65, 66), attested in the same speaker. It is well known that in Maghribi dialects, numerals 11 to 19 show an -n form when in direct annexation with the *ma’ dūd* ((Marçais 1977, p. 178); for Tunis Arabic, see (Biṭuna 2011, p. 32)). In fact, in Sousse, numerals from 11 to 19 also have an -n final shape when in *idāfa* (see (Talmoudi 1980, p. 169)). Concord from 20, in Mahdia Arabic, reflects the rules of *fušḥā* too, such as in the cases of *‘išrīn yōm* “twenty days” ((Yoda 2019, p. 65); for Tunis Arabic, (Biṭuna 2011, p. 34) and *mya ‘askar* “one hundred soldiers” ((Yoda 2019, p. 59); for Maghribi Arabic, see (Marçais 1977, pp. 173–80), for Tunis Judeo-Arabic, (Cohen 1975, p. 232); for Takrouna Arabic see also Marçais; (Marçais and Guiga 1925): *passim*).⁴⁷

The tendency towards weakening or to let the final /n/ drop in tens, in the towns considered in this analysis, has already been dealt with above; some other cases involving dozens are *ḥamsī* “fifty” and *arba‘ī* “fourty”.

3.3. Verbal Morpho-Syntax

3.3.1. An Urban Conjugation

Verbs in the Arabic variety of Mahdia and the surrounding towns, that is Monastir, Msaken and Bekalta, are generally conjugated according to the stem *yimši/yimšīw* and *yaqra/yaqrāw*, traditionally defined as pre-Hilali (Marçais and Guiga 1925, p. 49 and passim; Marçais 1950, p. 209)⁴⁸ as well as in other urban varieties of Sahel, such as Sousse and Sfax (Sellami 2019, pp. 86–87, Sellami 2019; Herin and Zammit 2017, p. 143). Some examples are *yemšīw* “they go”, *naqrāw* “we read”, *nalqāw* “we find”, *netmannēw* “we hope”, *nistennēw* “we wait for” (see also Yoda 2019, pp. 59, 66)⁴⁹.

Even if data for Mahdia Arabic seem to be clear, as Mion (2018, pp. 117–18) pointed out, for Sahel, the issue cannot be reduced to the simple opposition pre-Hilali/Hilali conjugation, since a large part of Tunisian varieties have mixed features with plural perfect conjugation in -āw and imperfect in -ū, i.e., *mšāw/yamšu* (see also (Mion 2015, pp. 272–73)).⁵⁰

3.3.2. Feminine Third Person Singular of Weak-Final Verbs

In Mahdia, Monastir, Bekalta and Msaken, the feminine third person singular of the perfect tense has a long vowel and -ā-t ending, also showing *imāla*, such as in the following cases: *mšēt* “she went”; *ta‘ addēt* “she passed”. The long vowel, preserved as in the so-called pre-Hilali dialects, has been also attested by Yoda (2019, p. 65) in *xdāt* “she took” and *‘tāt* “she gave”. In addition, the surrounding urban varieties have similar features (see Talmoudi 1980, pp. 86–88).⁵¹

3.3.3. Perfect Tense Pattern

Similar to the other urban dialects, Mahdia Arabic does not show gender opposition at the feminine second person singular *anti* of the perfect tense, just like the other urban varieties of Sahel such as Msaken (Bouhleb 2009, p. 131), Sousse (Talmoudi 1980, pp. 77–78) and Sfax (Lajmi 2009, p. 138; see also (Mion 2015, p. 273)).

Gender opposition at the second person singular “you”, in fact, is maintained only in conservative⁵² Bedouin or village dialects in which the feminine person has the desinence -ti (Marçais 1977, p. 36).

3.3.4. Use of the Verb *rā*

In the Arabic of Mahdia, the verb to see is expressed through the verb *šāf* “to see, to look at, to watch”, common to a wide variety of Arabic dialects,⁵³ and *rā* “to see”.

Additionally, in Sfaxi Arabic, the two verbs alternate with a certain predominance of *rā* (Lajmi 2009, p. 140; Zammit 2014, p. 34; Sellami 2019) and, in Sousse, the verb *rā* is used, but with a peculiarity, that is the first person singular of the perfect tense *rūt* (Talmoudi 1980) and not *rīt*, as in Kairouan and Mahdia. Among the examples, we find *rītus?* “have not you seen it?” (Yoda 2019, p. 66). In Mahdia Arabic, *šāf* is very used, as in the cases of *šūf*, *šāf* (Yoda 2019, pp. 60, 63, 66). In Tunis Judeo-Arabic, the verb *rā* is used and it has maintained the diphthongation /ay/ (Cohen 1975, p. 106). From an historical perspective, the verb *rā* replaced *šāf* in Sicilian Arabic, in Sicilian Judeo-Arabic and in Maltese. Moreover, it was very productive in Andalusi Arabic too (La Rosa 2019, p. 259; Burgaretta 2016, p. 89; Corriente et al. 2017, pp. 508–9, for *šāf*, p. 742).

3.3.5. Passive

The passive form of verbs is usually expressed through an initial t- pattern, used in Tunisia and in other Maghribi varieties (see, i.e., (Cohen 1975, pp. 123–25)), “né, sans doute, sous l’influence analogique des réfléchis-passifs à t- initial des thèmes V et VI, procédant

respectivement des thèmes II et V, ce thème procède de verbes du thème fondamental généralement en usage. Il en constitue le réfléchi-passif” (Marçais 1977, p. 66).⁵⁴

The following examples have been attested in Mahdia, Bekalta, Monastir and Msaken: *təbnat* “it was built”; *yitnaḥḥa* “it is removed”, *yitfaḥem* “it is understood”, *təṯḥet fuq al-rās* “it is placed on the head”, *tufṯi* “it switches off”, *mā tufṯiš* “it does not switch off” (for the phenomenon in Sfaxi Arabic, see (Sellami 2019)).⁵⁵

4. Discussion and Conclusions

Marçais (1950, p. 214) stated that the Tunisian Sahel is characterised by sedentary varieties, which break the continuum of Sulaymi dialects and divided Bedouin Tunisian dialects into two main groups: Hilali and Sulaymi. According to this kind of classification, the features of Mahdia Arabic would have some characteristics attributable to the Banū Hilāl, such as the *imāla* of internal vowels, the masculine singular third person pronoun suffix –u, opposed to the suffix –a(h) of the Sulaymi group and the passive in t-. Some Sulaymi features would be the weaker articulation of emphasis and the final *imāla* (see (Marçais 1950, p. 217; Ritt-Benmimoun 2014, p. 354; Taine-Cheikh 2017, pp. 20–21). However, Ritt-Benmimoun’s studies on the Bedouin Arabic of South Tunisia have showed that the categorisation Hilali/Sulaymi is not always definite and obvious and led us to wonder whether “there is a real zone of transition between areas where the S and the H dialects are spoken or if these areas are separated by a more or less well-defined boundary, perhaps corresponding exactly to the settlement area of the different tribes” (Ritt-Benmimoun 2014, p. 358).

In 2017, Taine-Cheikh pointed out that Marçais’ criteria of dialect classification had to be revisited and stated:

Reste, me semble-t-il, une question, celle de la valeur de la distinction entre parlers hilaliens vs sulaymites vs ma’qiliens. S’il existe bien de groupes de parlers plus ou moins différenciés et s’il est nécessaire de leur attribuer un nom, je ne suis pas sûre que ces trois désignations d’origine socio-historique soient d’une réelle précision et donc, d’une véritable secours (Taine-Cheikh 2017, p. 38).⁵⁶

Guerrero (2018) revisited the idea of the villageois category in which the rural Tunisian dialects are generally included, and based on 20 phonological, morpho-syntactic and lexical features, demonstrated that they show important differences from the Algerian and Moroccan varieties, which, instead, constitute a group with consistent features. Benkato (2019, pp. 11–12) pointed out that the classifications of Bedouin dialects, and particularly Sulaymi dialects, derived from W. Marçais’ statements, arose in turn from de Slane and G. Marçais’ assumptions, were “taken as fact”, even if the scholar did not offer any genealogy of these tribes. Benkato added that “these categories were hardly based on sound linguistic argumentation and instead more on the personal experience and reputation for their creators” (Benkato 2019, p. 14).

For all these reasons, Marçais’ traditional criteria of classification therefore cannot be applied *tout court* to the analysis of Mahdia Arabic. That is, it is not possible nor methodologically appropriate to attribute a precise and certain origin to all of its features. In spite of this, the first data presented in this contribution confirm that Mahdia Arabic is a sedentary variety showing some “contact” features.

According to the traditional classification, typical sedentary traits are the voiceless articulation of /q/, the relative pronoun *ellī* and the conjugation pattern *mšēw/yemšēw*. As regards the latter, it is useful to stress what Mion (2018, p. 118) claimed:

Alors que -īw du Maroc à la Tunisie doit être vue comme sédentaire et citadin, un système transversal avec un parfait –/āw/ et imparfait –/ū/ est en réalité bien plus fréquent que ce que la simple opposition méthodologique pré-hilalien/hilalien laisserait entendre. Si ce système doit être conçu comme transitoire au sein d’un *continuum* don’t les deux pôles seraient justement les typologies pré-hilalien/hilalien, alors une bonne partie de la Tunisie (à l’exception de ses métropoles et des variétés Marāzīg) devrait être considérée paradoxalement comme une zone de transition.⁵⁷

Some “contact” features attested in Mahdia Arabic are the final etymological *imāla*, and the opening of /ī/ towards /ē/ in the particle *fī* followed by the personal suffix pronoun *-hā*.

However, the situation does not always seem to be clear and definite since sedentary dialects also share many elements with Bedouin Hilali dialects, such as the following used in Mahdia region: The pronoun for the third person masculine singular *-u*, and the realisation of final /ī/ and /ū/ as /e/ and /o/, when followed by the pronoun *-hā*, which “is not considered as an isogloss marking either Bedouin or sedentary varieties. However, it is attested in dialects not generally considered as ‘rural’” (D’Anna 2020, p. 93). According to Ritt-Benmimoun:

The affinity of the H [Hilali] dialects with Tunisia’s sedentary dialects can be found in Philippe Maarçais’ statement that in the regions with a prevalence of H dialects sedentary dialects had originally been spoken which were subsequently overlaid by Bedouin dialects (Ritt-Benmimoun 2014, p. 355).

As regards the reduction of diphthongs to /ē/ and /ō/, which is generally considered to be a Bedouin feature, is attested by Yoda (2019) in Mahdia Arabic, but is not much attested in my corpus where, instead, a strong presence of /ī/ and /ū/ is noticed. Therefore, further investigation will shed new light on the real attestation and distribution of this feature.

Besides, some of the main sedentary features of Mahdia Arabic are also shared by the so-called *villageois* varieties, such as the voiceless realisation of /q/,⁵⁸ the final etymological *imāla šadīda*, the lack of gender opposition in verbs and pronouns and the long vowel *-ā* in the desinence *-ā-t* of the feminine third person singular of final-weak verbs (Marçais 1950, pp. 207–12).

From a diachronic point of view, some elements pointed out in previous studies offer a sketch of a slightly different variety of Mahdia Arabic. For instance, in 1969, Attia observes that /d/ “remplace l’interdentale /d̪/ qui a disparu du parler de Mahdia” Attia (1969, p. 123) and in 2008, Yoda identifies a series of phenomena, which in my corpus are not so obvious, such as the presence of *fē-ha* or the reduction of diphthongs to /ē/ and /ō/. These and other traits suggest a possible explication, that is that the Arabic of Mahdia is undergoing a change, a sort of linguistic levelling or standardisation towards the Arabic of the capital Tunis and, probably, also through the influence of the main city of Sahel, together with Sfax, Sousse, which is only 60 km away. As already shown above, Gibson’s research goes in this direction⁵⁹ as well as Mion’s observations about a series of phenomena that, in a certain way, contribute to create this situation of standardisation not only in Sahel Arabic, but more generally in Tunisian Arabic. Among these factors, I find particularly interesting the role of the media, and above all, television, that Mion considers a facilitator for the diffusion of Modern Standard Arabic to all the social strata. Television is therefore responsible for the reintroduction of some phonemes in people’s speech, such as the *hamza* and the voiceless articulation of /q/,⁶⁰ and also for the new diffusion of the *-āw/iw* perfect tense endings. Moreover, it helped the spread of the prestigious dialectal Arabic of the capitals, such as the Tunisian of Tunis (Mion 2018, p. 120).⁶¹ The fact that the aged fishermen natives of Burž al-rās, interviewed in the already-mentioned documentary on this quarter, all pronounce the interdental /d̪/ as /d/ confirm Attia’s observations and lead us to consider this feature as a confirmation of the linguistic levelling in progress, above all among the youngest and more educated inhabitants of Mahdia. According to Sayahi (2019, p. 237), the fact that in recent years the institutions often choose to utter their official speeches in Tunisian Arabic fostered the use of the dialect in the public space for public communications and not only for private or semi-private occasions (on this subject, see also (La Rosa 2018)).

Moreover, I believe that the diffusion of the use the smartphone has been crucial, above all after the 2011 Arab Spring(s). In fact, thanks to it, the internet has become more accessible to a large part of the population. Easy and cheap access to the internet has allowed the acquisition and the circulation of information from far-off places within the country and abroad. Thanks to the use of social media, such as Facebook and WhatsApp, or the use of

YouTube, the spread of several forms of oral and written artistic forms and cultures has become more and more easy.⁶² What is most important here is that the linguistic contact between people from different regions of Tunisia, speaking different varieties of Arabic, has become easier too. I think that this has helped the linguistic levelling process in progress by bringing people together, in spite of the geographical distance.

Furthermore, we should also take into consideration that the role of the migration movements inside Tunisia and abroad (see (D'Anna 2017) regarding the Sicilian community of Mazara, mainly consisting of Tunisians from Mahdia and Chebba) for study and work reasons may have had particular relevance in the levelling of some linguistic features contained in the varieties of Sahel. In fact, "migration from the countryside to cities has constituted the most significant demographic change in the last two centuries. Linguistic contact in cities is more intense and involves faster changes than in rural areas. [...] In recent decades, a significant increase in the frequency and variety of types of inter-dialectal and intra-dialectal contact occurred causing faster and greater degrees of levelling" (Vicente 2019, p. 106). Linguistic levelling is also evident in the town, which is probably most affected by migration movements among those mentioned in this study, i.e., Msaken. As Bouhleb (2009, p. 133) already pointed out, in fact, Msaken Arabic underwent a deep change during the last 50 or 70 years, affecting mainly, but not exclusively, lexicon.

In addition to this, as shown above, the medium used for the interviews in my corpus has a crucial role since, to a certain extent and in many cases, it conditions the speaker who knows that he/she is being filmed or recorded and, therefore, tends to use linguistic structures and/or a lexicon belonging to a higher register, that is the standard variety, be it Modern Standard Arabic or Standard Dialect.⁶³

Funding: This paper contains the first results of a research project, titled "*Lahğāt Wilāyat al-Mahdiyya*: proposte di descrizione e classificazione", which is part of the line of research "Starting Grant", of the "Piano di incentivi per la ricerca di Ateneo 2020/2022 (Pia.ce.ri.)", funded by the University of Catania.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

Notes

¹ "We consider here those of the towns and small cities of the Sahel [...] which have not yet been investigated". English translation is mine.

² "The Arabic language of Takrouna is generally consistent with that of the agricultural centres, towns and villages that dot the coastal region of central Tunisia, commonly called Sâhel [...] since the Middle Ages. Separated from each other by differences in detail, these varieties belong, as far as phonetics and grammar are concerned, to the same general type of which takrouni is only a particular variety". English translation is mine.

³ I would like to thank Giuliano Mion for reading these pages and for giving me some precious suggestions. I am also grateful to the anonymous reviewers for their suggestions that allowed me to enrich this paper. Any possible imprecision or mistake is, of course, my responsibility.

⁴ For an analysis and classification of the urban dialects in the Arab world, see (Vicente 2019, p. 106).

⁵ It is what Marçais called "différences de détail" among the dialects of Sahel that we need to describe better, of course, through a team work which will need future studies and the joint commitment of different scholars.

⁶ On the use of this term, see (Benkato 2019, p. 9).

⁷ <http://www.commune-mahdia.gov.tn/en> accessed on 1 August 2021.

⁸ During the Turkish domination in Tunisia (but also in Algeria), this term, originally meaning 'son of a slave' or 'son of a Janissary', indicated the population deriving from the marriage of Turks and local women. Many of them were occupied in agriculture or in the local army. Afterwards, they merged with the local population. See Ed. Kul-Oghlu, in (Bearman et al. 1986, vol. 5, p. 366).

⁹ <https://www.planetware.com/tourist-attractions-/mahdia-tun-md-md.htm> accessed on 1 August 2021.

¹⁰ Regarding the narration of historical sources on Banū Hilāl as destroyers of the Maghrib, see (Benkato 2019, pp. 5–6, n. 8).

- 11 http://www.commune-monastir.gov.tn/web/fr/about_ville accessed on 1 August 2021.
- 12 <http://www.commune-msaken.gov.tn/date.php.html> accessed on 1 August 2021.
- 13 The few studies available on this town are now dated and no map is available in the website of the Commune de Msaken. Even in Medieval sources, the town is less mentioned than Mahdia or Monastir, when it is not mentioned at all.
- 14 It should be noted that for the writing of this contribution I have not included the analysis of speeches performed in a semi-formal style or in mixed Arabic. For Mixed varieties see, among others, (Larcher 2001; Den Heijer 2012), and the Proceedings of the AIMA symposia and relative bibliography: (Lentin and GrandHenry 2008; Zack and Schippers 2012; Bettini and Spisa 2012).
- 15 These elements will be discussed below.
- 16 Baqalta, commonly known as Bekalta, is a village 10 km from Mahdia, belonging to the Wilāyat al-Munastīr.
- 17 I would like to point out that, in this contribution, my main focus will be Mahdia Arabic and I will only make some brief mentions for the surrounding varieties. The focus on the latter will be developed in further studies.
- 18 See the interesting observations of Van-Mol (2010, pp. 70–77) on satellite tv broadcasts and Nguyen (2021) statements on using social media for dialect research which are partially applicable here.
- 19 Yoda (2019) defines the appearance of interdental as classicism. I believe that not only Standard Arabic, but also the Arabic of the capital, Tunis, or of nearby Sousse, influence this realization of interdental (see (Gibson 2002)). It must be pointed out that Saada (1984, p. 23) stated that the Arabic of Tozeur had “conservation des interdental à de rares exceptions [. . .] dans les quartiers des Ulād el Hāḍef et des Žīma’ les interdental sont faibles chez les sujets masculins et féminins” [“retention of interdental with rare exceptions [. . .] in the Ulād el Hāḍef and Žīma’ districts interdental are weak in both men and women speakers”. English translation is mine]. See also the note 2 at the same page.
- 20 Note that I have chosen to transcribe the article *al-* with the lateral /l/ also when it is followed by a solar consonant, such as in the cases of *al-tēnī* (to be read *at-tēnī*), *al-tēnīya* (to be read *at-tēnīya*), and *al-rās* (to be read *ar-rās*).
- 21 Biṭuna (2011, p. 28) notes a certain “instability” in the articulation of the interdental in numerals from 3 to 10. Maura Tarquini gave a paper titled “Le (inter)dentali in arabo tunisino” during the “Prima giornata di dialettologia maghrebina” (Cagliari, 16/05/2019) which has not yet been published.
- 22 See also (Vicente 2019, pp. 107–8).
- 23 Eventual further information about the occurrence of /q/ and /g/ in Mahdia region will emerge from the interviews that I will carry out as soon as possible, when the sanitary emergency allows it.
- 24 For Sfaxi Arabic, see (Lajmi 2009, p. 136). Kerkenna Arabic is a /q/ dialect, except for Mellita variety: see (Herin and Zammit 2017, pp. 140, 142–43).
- 25 As underlined also by other scholars, such as Yoda, the Judeo-Arabic of Tunis shares some features with Mahdia Arabic. Therefore, Cohen’s study on this variety will be taken into account in this paper.
- 26 See <https://tunico.acdh.oeaw.ac.at/dictionary.html?query=flk> accessed on 13 August 2021. It is impossible to indicate here all of the videos analyzed for this research, but it may be useful to point out that these data are taken from a documentary realized by *Menara Fm* in December 2020, dedicated to the quarter of Burž al-rās. It is a 35-min documentary in which examples of many other features described in this paper can be retraced. However, further studies will be needed to clarify the presence and the distribution of the articulation of /k/ as /q/ in Mahdia and the surrounding towns.
- 27 It is well known, however, that in Tunisian Arabic the opposition /r/ and /r̥/ is well established and attested: see (Cohen 1975, vol. 2, pp. 14–15) and (Saada 1984, p. 26).
- 28 Out of my own concern for completeness (as far as possible, since a single contribution is far from being an exhaustive study) of information, I have listened to some videos with speakers from Kairouan, taken from *Radio Monastir*, which have not been inserted in my “official” corpus.
- 29 On the status of emphasis in Arabic see (Cohen 1969; Ghazeli 1981; Gouma 2013).
- 30 As far as I know, there are no systematic studies on the degrees of affrication of dental consonants in Tunisian Arabic. Further studies would allow the phenomenon to be described better.
- 31 “without any doubt stronger than in Muslim Tunis, which would explain the fact that Muslims who want to imitate Jewish speech exaggerate the emphasis as well as the expressive modulations of the sentence”. English translation is mine.
- 32 On the notion of articulatory strength, see for instance (Ladefoged and Maddieson 1996, pp. 95–97). About the influence of the communitarian and identity dynamics on the Arabic language, see (Mion 2017) on Cypriot Arabic. It should be noticed that in Sicily, which showed a complex cultural and religious situation, according to Lentin (2006/2007, p. 76), “Les «emphatiques» semblent avoir été articulée sans (grande) emphase” [“The ‘emphatics’ seem to have been articulated without (much) emphasis”. English translation is mine].
- 33 What Attia (1969, p. 127) stated about /e/ should be noted: “C’est une voyelle semi-ouverte, avant, brève. Elle représente entre /a/ et /i/ un degré d’aperture absent en Arabe classique. Son absence est distinctive, mais sa longueur ne le semble pas. Allongée, elle tend à se confondre avec le degré d’aperture voisin /i/ ou /a/” [“It is a semi-open, front, short vowel. It represents a degree of aperture between /a/ and /i/ which is absent in classical Arabic. Its absence is distinctive, but its length does not

seem to be distinctive. Elongated, it tends to merge with the nearby aperture degree /i/ or /a/. English translation is mine] and about /o/: “C’est une voyelle semi-ouverte, arrière, brève. Elle représente entre /a/ et /u/ un degré d’aperture absent en Arabe classique. Son absence ne semble pas distinctive. Allongée, elle tend à se confondre avec le degré d’aperture voisin /a/ ou /u/” [“It is a semi-open, back, short vowel. It represents a degree of aperture between /a/ and /u/ which is absent in classical Arabic. Its absence does not seem distinctive. Elongated, it tends to merge with the nearby aperture degree /a/ or /u/. English translation is mine]. For Attia, these two phonemes in Mahdia Arabic could only be short (see Attia 1969, pp. 128–29).

34 For the vowel system of Tunisian Judeo-Arabic, see (Cohen 1975, vol. 2, 56–62).

35 Describing in detail the vowel system of Mahdia Arabic is beyond the scope of this article and will be the subject of future studies. I limit myself to observe that, as shown in the examples given in this paper, in my corpus some cases of /ə/ have been found. On the vocalism of Maghribi Arabic see among others the recent observations of Mion (2018, pp. 112–13) and the relative bibliography.

36 In fact, Yoda (2019, p. 61, n. 32) states with certainty: “when the preposition *fi* ‘in’ is followed by the suffix pronoun *-ha*, the vowels of these two phonemes assimilate to each other: *ē* > *fē*-he”. For the phenomenon in the Arabic of Tozeur, see (Saada 1984, pp. 32–33).

37 For the reduction of etymological diphthongs in Tunis Judeo-Arabic, see (Cohen 1975, vol. 2, p. 68).

38 “Like the conspecific dialects of Sâhel, takroûni follows on some points its own paths which, distancing it from the urban dialects of the Regency, bring it closer to certain Bedouin dialects of the eastern Maghreb [. . .]: the old diphthongs *ai* and *au*, accented and not in final position, are generally reduced to *ē* and *ō*”. English translation is mine.

39 In Sfaxi Arabic, diphthongs are reduced to /ī/, /ū/ or /ō/ and in Sellami (2019)’s sample they were retained. The scholar also shows that there is a generational difference in how youngsters and aged speakers treat diphthongs. In Kerkenna Arabic, diphthongs are reduced to /ē/ and /ō/ (Herin and Zammit 2017, p. 140).

40 As for Sahel, Sfaxi Arabic is an exception, especially for medial *imāla*. See (Sellami 2019). The first variety of Arabic of Kerkenna, described by Herin and Zammit, shows some exceptions since medial *imāla* of /ā/ is not raised, but final *imāla* is raised only in monosyllabic words. A second variety shows the raising of final /a/ triggered after non-emphatic front consonants and front vowels. The variety of Mellita raises final /a/. See (Herin and Zammit 2017, pp. 140–41).

41 Sousse Arabic has *inti* (Talmoudi 1980, p. 143).

42 See also (Yoda 2019): *passim*, in which we find *āni* “I”, *aḥna* “we” and *inti* “you”. For Sousse Arabic, see (Talmoudi 1980, p. 143) and for Tunis Judeo-Arabic, which offers a unique form *enti* for the second person singular “you”, (Cohen 1975, pp. 210–11).

43 See in particular p. XXII, n. 3 on the origins of *nāya/nāy*.

44 Also the Judeo-Arabic of Tunis shares the use of *ellī* and variants: see (Cohen 1975, pp. 218–21). In Tozeur, *ellī*, *ella* and allomorphs alternate (Saada 1984, p. 79).

45 See also the comparative study of (Taine-Cheikh 1994).

46 Some other features indicated by Biṭūna are present in the variety of Mahdia and Msaken, such as the use of *zūz* + plural of the noun, feature already used in Medieval Maghribi Arabic (see (La Rosa 2019, pp. 218–19; Corriente et al. 2015, pp. 110–11), for Msaken Arabic see (Bouhleb 2009, p. 128)), the feminine form of the numerals from 3 to 10 when the numeral in question has the function of an isolated substantive. The use of the indefinite numeral *wāḥd al-* has not been attested.

47 The word *‘askar*, “soldier” or “army” as a collective, is considered by the speakers both as a singular and as a plural. As for the variety of Sousse, (Talmoudi 1980, p. 139) points out that it has a different vowel quantity. The *-n* form in numerals from 11 to 19 is also present (Talmoudi 1980, p. 140).

48 The same happens in many Tunisian coastal vernaculars, such as Sfaxi Arabic: see (Sellami 2019), and in Tunis Judeo-Arabic, see (Cohen 1975, p. 104).

49 In Kairouan, we find *nistennu* “we wait for”. Information taken from some videos of the page *Radio Monastir*.

50 In Bekalta, I found the example *nistennēw w-mā nensūhum* “we wait and we do not forget them”, so further research on the varieties spoken in the region will be necessary to sketch a more precise map of the distribution of this feature.

51 The feature, of course, is also shared by Tunis Judeo-Arabic (Cohen 1975, p. 104).

52 “Labelling a dialect as conservative [. . .] means that it exhibits a certain degree of stability over time: certain features are less likely to evolve because there are fewer exogenous factors that could lead to change” (Herin 2019, pp. 94–95).

53 <https://tunico.acdh.oeaw.ac.at/dictionary.html?query=%22to%20see%22> accessed on 13 August 2021.

54 “created, without any doubt, under the analogical influence of the passive-reflexives with initial *t-* of forms V and VI, which precede respectively from forms II and V, this stem proceeds from verbs of the first form generally in use. It constitutes its passive-reflexive”. English translation is mine.

55 For Sousse, see (Talmoudi 1980, p. 101). See also (Saada 1984, pp. 57–58) who finds the phenomenon in Tozeur.

56 “It seems that a question remains: the value of the distinction between Hilali vs Sulaymi vs Ma’qili dialects. If there are indeed groups of more or less differentiated varieties and if it is necessary to give them a name, I am not sure that these three designations of socio-historical origin are really precise and of real help”. English translation is mine.

- 57 [“While -iw from Morocco to Tunisia should be seen as sedentary and urban, a transversal system with a perfect tense in -/āw/ and an imperfect tense in -/ū/ is in fact much more frequent than the simple methodological opposition pre-Hilalian/Hilalian would suggest. If this system has to be conceived as transitional within a *continuum* whose two poles would be precisely the pre-Hilalian/hilalian typologies, then much of Tunisia (with the exception of its metropolises and Marāzig varieties) should paradoxically be considered a transitional zone”. English translation is mine]. The verb *nensū* attested in Bekalta has to be verified and supported with other occurrences to verify whether the varieties near Mahdia have any mixed features.
- 58 The critical issues of the application of this criterion of classification have already been underlined in the relative paragraph in the section devoted to phonology.
- 59 The already mentioned paper by Maura Tarquini seemed to confirm this trend.
- 60 The voiceless realization of /q/ was the norm in Medieval Arab cities and in the recitation of the Koran. For this reason, it must have always been considered a positive element by speakers. This encouraged the spread of the variant in the Arab world via the Maghrib. See (Vicente 2019, p. 109).
- 61 I wonder whether the preservation of the monophthongs /ī/ and /ū/ in the Arabic of Mahdia and the surrounding towns must be included in these elements.
- 62 An example is Nizar Chaari’s novel *Tūnis fi ‘īmayya*, in which the author inserts additional texts and materials downloadable on one’s own smartphone through a QR code. His aim was to involve young people in reading and making it attractive to them, by giving them the opportunity to do it wherever and whenever they want through their smartphone. See (La Rosa Forthcoming).
- 63 Such an analysis on the influence of the medium (radio and video) on the variety of Arabic used by the speakers involved in the videos selected will be the object of future studies. In fact, in the wider framework of the ongoing studies on the features of contemporary Media Arabic, observing how the varieties of Arabic are used in Tunisia could help describe more precisely the complex situation of linguistic variation in the country. On this subject, see for instance the PhD thesis “Arabe Mixte 2.0: la variation syntaxique et stylistique dans les journaux numériques marocains (janvier-décembre 2016)”, defended by Rosa Pennisi in December 2020.

References

- Attia, Abdelmajid. 1969. Le parler de Mahdia. In *Travaux de Phonologie. Parlers de: Djemmal, Gabes, Mahdia (Tunisie), Treviso (Italie)*. Edited by Taieb Baccouche, Hichem Skik, Abdelmajid Attia and Mohamed El Habib Ounali. Tunis: Centre d’études Économiques et Sociales, pp. 116–38.
- Avram, Andrei. 2012. Some phonological changes in Maltese reflected in onomastics. *Bucharest Working Papers in Linguistics* 14: 99–119.
- Bearman, Peri, Thierry Bianquis, Clifford Edmund Bosworth, Emery van Donzel, and Wolfhart P. Heinrichs, eds. 1986. Kul-Oghlu. In *Encyclopaedia of Islam*, 2nd ed. Consulted online on June 9 2021. [CrossRef]
- Benkato, Adam. 2019. From Medieval Tribes to Modern Dialects: On the Afterlives of Colonial Knowledge in Arabic Dialectology. *Philological Encounters* 4: 2–25. [CrossRef]
- Bettini, Lidia, and Paolo La Spisa, eds. 2012. *Au-delà de l’arabe Standard. Moyen arabe et Arabe mixte dans les Sources Médiévales, Modernes et Contemporaines*. Quaderni di Semitistica. Firenze: Università di Firenze, vol. 28.
- Biṭuna, Gabriel. 2011. The Morpho-Syntax of the Numeral in the Spoken Arabic of Tunis. *Romano-Arabica* 8–11: 25–42.
- Boubakri, Hassan, and Ridha Lamine. 1992. Migrations pendulaires autour de M’saken. *Revue Tunisienne de Géographie* 19–20: 11–29.
- Bresc, Henri, and Annliese Nef. 1999. *La première géographie de l’Occident. Présentation, Notes, Index, Chronologie et Bibliographie. Traduction du Chevalier Jaubert, revue par Annliese Nef*. Paris: Flammarion.
- Bouhleb, Ezzeddine. 2009. Le Parler m’sakenien. *Synergies Tunisie* 1: 125–34.
- Bournand, François. 1893. *Tunisie et Tunisiens*. Paris: Taffin-Lefort.
- Burgaretta, Dario. 2016. Un’elegia in giudeo-arabo di Sicilia per il massacro di Noto e Modica del 1474 (Ms. Parm. 1741 della Biblioteca Palatina di Parma). *Sefer Yuḥasin* 4: 7–194.
- Cantineau, Jean. 1960. *Cours de Phonétique Arabe. Notions Générales de Phonétique et Phonologie*. Paris: Librairie Klincksieck.
- Carter, Michael G. 2007. Elision. In *Encyclopedia of Arabic Languages and Linguistics*. Edited by Kees Versteegh. Leiden and Boston: Brill, vol. 2, pp. 16–18.
- Cohen, David. 1969. Sur le statut phonologique de l’emphase en arabe. *Word* 25: 59–69. [CrossRef]
- Cohen, David. 1975. *Le parler arabe des Juifs de Tunis. Tome II. Étude Linguistique*. Paris: Mouton.
- Corriente, Federico, Christophe Pereira, and Ángeles Vicente. 2015. *Aperçu Grammatical Du faisceau Dialectal arabe Andalou. Perspectives Synchroniques, Diachroniques et Panchroniques. Encyclopédie linguistique d’al-Andalus*. Berlin and Boston: De Gruyter, vol. 1.
- Corriente, Federico, Christophe Pereira, and Ángeles Vicente. 2017. *Dictionnaire du Faisceau Dialectal Arabe Andalou. Perspectives Phraséologiques et Étymologique. Encyclopédie Linguistique d’al-Andalus*. Berlin and Boston: De Gruyter, vol. 2.
- Dallaji, Ines. 2017. Grammatical and Lexical Notes on the Arabic Dialect of Nabeul (Northern Tunisia). In *Lisan Al-Arab. Studies in Arabic Dialects. Proceedings of the 10th International Conference of AIDA Qatar University 2013*. Edited by Muntasar Fayez Faris Al-Hamad, Rizwan Ahmad and Hafid I. Alaoui. Lit: Zurich, pp. 151–62.
- D’Anna, Luca. 2017. *Italiano, siciliano e arabo in contatto. Profilo sociolinguistico dela comunità tunisina di Mazara del Vallo*. Palermo: Centro di Studi Filologici e Linguistici Siciliani.

- D'Anna, Luca. 2020. The Arabic dialect of Chebba: Preliminary data and historical considerations. *Zeitschrift für Arabische Linguistik* 72: 80–100. [CrossRef]
- Decret, François. 2003. Les Invasions Hilaliennes en Ifrîqiya. Office National des Statistiques pour la Population et la Démographie. Available online: https://www.clio.fr/bibliotheque/les_invasions_hilaliennes_en_ifriqiya.asp (accessed on 11 August 2021).
- Den Heijer, Johannes. 2012. Introduction: Middle and Mixed Arabic, A New Trend in Arabic Studies. In *Middle Arabic and Mixed Arabic. Diachrony and Synchrony*. Edited by Liesbeth Zack and Arie Schippers. Leiden: Brill, pp. 1–25. [CrossRef]
- Ferrando, Ignacio. 1995. *El dialecto andalusi de la Marca Media. Los Documentos Mozárabes Toledanos de los Siglos XII y XIII*. Zaragoza: Universidad de Zaragoza (Area de estudios Arabes e Islámicos 4).
- Ghazeli, Salem. 1981. La Coarticulation de l'Emphase En Arabe. *Arabica* 28: 251–77. [CrossRef]
- Gibson, Maik. 2002. Dialect Levelling in Tunisian Arabic: Towards a New Spoken Standard. In *Language Contact and Language Conflict in Arabic. Variations on a Sociolinguistic Theme*. Edited by Rouchdy Aleya. Oxon and New York: Routledge, pp. 24–40.
- Gibson, Micheal Luke. 1998. Dialect Contact in Tunisian Arabic: Sociolinguistic and Structural Aspects. Ph.D. thesis, University of Reading, Reading, UK.
- Goitein, Shelomo Dov. 2010. *Studies in Islamic History and Institutions*. Brill Classics in Islam. With an Introduction by Norman A. Stillman. Leiden and Boston: Brill, vol. 5.
- Gouma, Taoufik. 2013. L'emphase en Arabe Marocain: Vers une Analyse Autosegmentale. Ph.D. thesis, Université de Paris 8, Paris, France.
- Guerrero, Jairo. 2018. Les parlers Jbala-villageois. Étude grammaticale d'une Typologie rurale de l'arabe Dialectal Maghrébin. *Dialectologia* 20: 85–105. Available online: <https://hal.archives-ouvertes.fr/hal-02095906> (accessed on 1 August 2021).
- Herin, Bruno. 2019. Traditional Dialects. In *The Routledge Handbook of Arabic Sociolinguistics*. Edited by Enam Al-Wer and Uri Horesh. London and New York: Routledge, pp. 93–105.
- Herin, Bruno, and Martin Zammit. 2017. Three for the Price of One: The Dialects of Kerkennah (Tunisia). In *Tunisian and Libyan Arabic Dialects. Common Trends—Recent Developments—Diachronic Aspects*. Edited by Veronika Ritt-Benmimoun. Zaragoza: Prensas de la Universidad de Zaragoza, pp. 135–46.
- Ladefoged, Peter, and Ian Maddieson. 1996. *The Sounds of the World's Languages*. Oxford: Blackpowell.
- Lajmi, Dhoulha. 2009. Spécificités du dialecte Sfaxien. *Synergies Tunisie* 1: 135–42.
- Larcher, Pierre. 2001. Moyen Arabe et Arabe Moyen. *Arabica* 48: 578–609. [CrossRef]
- La Rosa, Cristina. 2018. Alcune strategie retoriche nel discorso politico tunisino: Uso dei deittici e ripetizione lessicale. *La rivista di Arabilit* 15: 67–92.
- La Rosa, Cristina. 2019. *L'arabo di Sicilia nel Contesto Maghrebino: Nuove Prospettive di Ricerca*. La Sicilia Islamica: Testi, Ricerche Letterarie e Linguistiche. Roma: Istituto Per l'Oriente Carlo Alfonso Nallino, vol. 1.
- La Rosa, Cristina. Forthcoming. Tūnis fi 'inayya by Nizār al-Ša' arī: Al-Dāriġa al-Tūnisīyya between Resistance and Hypertextuality. In *Studies on Arabic Dialectology and Sociolinguistics. Proceedings of the 13th International Conference of AIDA held in Kutaisi in June 10–13*. Edited by GuramGuram Chikovani and Zviad Tskhvediani. Kutaisi: Akaki Tsereteli State University Press, pp. 410–21.
- Lentin, Jérôme. 2006/2007. L'arabe parlé en Sicile était-il un arabe périphérique? *Romano-Arabica* 6–7: 71–84.
- Lentin, Jérôme, and Jacques GrandHenry', eds. 2008. *Moyen Arabe et Variétés Mixtes de l'arabe À travers l'histoire. Actes du Premier colloque International (Louvain-la-Neuve, 10–14 mai 2004)*. Publications de l'Institut Orientaliste de Louvain. Louvain-la-Neuve: Institut Orientaliste de Louvain, vol. 58.
- Maamouri, Mohamed. 1967. The Phonology of Tunisian Arabic. Ph.D. dissertation, Cornell University, Ithaca, NY, USA.
- Ma Mung, Emmanuel. 1984. L'impact des transferts migratoires dans la ville de M'saken (Tunisie). *Revue Européenne des Migrations Internationales* 2: 163–78. Available online: https://www.persee.fr/doc/remi_0765-0752_1986_num_2_1_1001 (accessed on 28 May 2021). [CrossRef]
- Marçais, Philippe. 1977. *Esquisse grammaticale de l'arabe maghrébin*. Paris: Maisonneuve.
- Marçais, William. 1950. Les parlers arabes. In *Initiation à la Tunisie*. Edited by André Basset. Paris: Maisonneuve, pp. 195–219.
- Marçais, William, and Abderrahmâne Guiga. 1925. *Textes arabes de Takroûna. Textes, Transcription et Traduction annotée*. Paris: Imprimerie Nationale, Leroux.
- Marçais, William, and Abderrahmâne Guiga. 1958. *Textes Arabes de Takroûna. Contribution à l'étude du Vocabulaire arabe. Glossaire*. Paris: Imprimerie Nationale, Librairie Orientaliste Paul Geuthner.
- Mion, Giuliano. 2015. Réflexions sur la catégorie des 'parlers villageois' en arabe tunisien. *Romano-Arabica* 15: 269–77.
- Mion, Giuliano. 2017. Un arabo cipriota romanizzato? Distacchi e identità fra variazione scrittoria e confessione religiosa. In *Aspetti della Variazione Linguistica. Discorso, Sistema, Repertori*. Edited by Carlo Consani. Milan: LED, pp. 83–95. [CrossRef]
- Mion, Giuliano. 2018. Pré-hilalien, hilalien, zones de transition. Relire quelques classiques aujourd'hui. In *Mediterranean Contaminations. Middle East, North Africa, and Europe in Contact*. Edited by Giuliano Mion. Berlin: Klaus Schwarz Verlag, pp. 102–25.
- Nguyen, Dong. 2021. *Dialect Variation on Social Media. Pre-print. Forthcoming in Similar Languages, Varieties, and Dialects*. (Studies in Natural Language Processing Series); Edited by Marcos Zampieri and Preslav Nakov. Cambridge: Cambridge University Press.
- Ritt-Benmimoun, Veronika. 2014. The Tunisian Hilāl and Sulaym Dialects. A Preliminary Comparative Study. In *Alf lahġa wa lahġa Lit-Verlag Proceedings of the 9th AIDA Conference*. Edited by Olivier Durand, Angela Daiana Langone, and Giuliano Mion. Münster and Wien: Lit-Verlag, pp. 351–60.
- Saada, Lucienne. 1984. *Éléments de Description du Parler arabe de Tozeur (Tunisie)*. Paris: Geuthner.

- Sayahi, Lotfi. 2019. Diglossia and the normalization of the vernacular. Focus on Tunisia. In *The Routledge Handbook of Arabic Sociolinguistics*. Edited by Enam Al-Wer and Uri Horesh. London and New York: Routledge, pp. 227–39.
- Sellami, Zeineb. 2019. The Dialect of Sfax (Tunisia). In *Studies on Arabic Dialectology and Sociolinguistics: Proceedings of the 12th International Conference of AIDA held in Marseille from May 30th to June 2nd 2017* [online]. Aix-en-Provence: Institut de Recherches et D'études sur les Mondes Arabes et Musulmans. [CrossRef]
- Skik, Hichem. 2000. La prononciation de qâf arabe en Tunisie. In *Proceedings of the Third International Conference of Aida*. Edited by Manwel Mifsud. Malta: pp. 131–36.
- Soucek, Svat. 1993. Monastir. In *Encyclopaedia of Islam*, 2nd ed. Edited by Clifford Edmund Bosworth, Emery van Donzel, Wolfhart P. Heinrichs and Charles Pellat. Leiden: Brill, vol. 7, pp. 227–29.
- Taïeb, Jacques, and Mansour Sayah. 2003. Remarques sur le parler judéo-arabe de Tunisie. *Diasporas. Histoire Et Sociétés* 2: 5–64. Available online: https://www.persee.fr/doc/diasp_1637-5823_2003_num_2_1_1_878 (accessed on 5 May 2021).
- Taine-Cheikh, Catherine. 1994. Les numéraux cardinaux de 3 à 10 dans les dialectes arabes. In *Actes des Premières Journées Internationales de Dialectologie Arabe de Paris*. Edited by Dominique Caubet and Martine Vanhove. Paris: INALCO, pp. 251–66.
- Taine-Cheikh, Catherine. 2017. La classification des parlers bédouins du Maghreb: Revisiter le classement traditionnel. In *Tunisian and Libyan Arabic Dialects: Common Trends, Recent Developments, Diachronic Aspects*. Edited by Prensas Universitarias de Zaragoza. Zaragoza: Prensas de la Universidad de Zaragoza, pp. 15–42.
- Talbi, Mohamed. 1986. "al-Mahdiyya". In *Encyclopaedia of Islam*, 2nd ed. Edited by Clifford Edmund Bosworth, Emery van Donzel, Wolfhart P. Heinrichs and Charles Pellat. Leiden: Brill, vol. 5, pp. 1245–47.
- Talmoudi, Fathi. 1980. *The Arabic Dialect of Sūsa (Tunisia)*. Göteborg: Acta Universitatis Gothoburgensis.
- Van-Mol, Marc. 2010. Oral Media and Corpus Linguistics: A First Methodological Outline. In *Arabic and the Media: Linguistic Analyses and Application*. Edited by Reem Bassiouney. Leiden and Boston: Brill, pp. 63–80.
- Vicente, Ángeles. 2019. Dialect contact and urban dialects. In *The Routledge Handbook of Arabic Sociolinguistics*. Edited by Enam Al-Wer and Uri Horesh. London and New York: Routledge, pp. 106–16.
- Voisins d'Ambre, Anne Caroline. 1884. *Excursions d'une Française dans la Régence de Tunis par Mme de Voisins (Pierre-Coeur)*. Paris: Dreyfous.
- Yoda, Sumikazu. 2008. On the Vowel System of the al-Mahdiyya Dialect of Central Tunisia. In *Between the Atlantic and Indian Oceans. Studies on Contemporary Arabic Dialects. Proceedings of the 7th Aida Conference, Held in Vienna from 5–9 September 2006*. Edited by Stephan Procházka and Veronika Ritt-Benmimoun. Münster and Wien: Lit-Verlag, pp. 483–90.
- Yoda, Sumikazu. 2019. Texts from Mahdiyya. *Journal of Arabic and Islamic Studies* 17: 55–70.
- Zack, Liesbeth, and Arie Schippers. 2012. *Middle Arabic and Mixed Arabic. Diachrony and Synchrony*. Leiden and Boston: Brill.
- Zammit, Martin. 2014. The Sfaxi Tunisian Element in Maltese. In *Perspectives on Maltese Linguistics*. Edited by Albert Borg, Sandro Caruana and Alexandra Vella. Berlin: De Gruyter, pp. 23–44.

Article

Contrastive Feature Typologies of Arabic Consonant Reflexes

Islam Youssef

Department of Languages and Literature Studies, University of South-Eastern Norway,
3833 Bø i Telemark, Norway; islam.youssef@usn.no

Abstract: Attempts to classify spoken Arabic dialects based on distinct reflexes of consonant phonemes are known to employ a mixture of parameters, which often conflate linguistic and non-linguistic facts. This article advances an alternative, theory-informed perspective of segmental typology, one that takes phonological properties as the object of investigation. Under this approach, various classificatory systems are legitimate; and I utilize a typological scheme within the framework of feature geometry. A minimalist model designed to account for segment-internal representations produces neat typologies of the Arabic consonants that vary across dialects, namely *qāf*, *ǧīm*, *kāf*, *dād*, the interdental, the rhotic, and the pharyngeals. Cognates for each of these are analyzed in a typology based on a few monovalent contrastive features. A key benefit of the proposed typologies is that the featural compositions of the various cognates give grounds for their behavior, in terms of contrasts and phonological activity, and potentially in diachronic processes as well. At a more general level, property-based typology is a promising line of research that helps us understand and categorize purely linguistic facts across languages or language varieties.

Keywords: phonological typology; feature geometry; contrastivity; Arabic dialects; consonant reflexes

Citation: Youssef, Islam. 2021. Contrastive Feature Typologies of Arabic Consonant Reflexes. *Languages* 6: 141. <https://doi.org/10.3390/languages6030141>

Academic Editors: Simone Bettega and Roberta Morano

Received: 8 June 2021

Accepted: 18 August 2021

Published: 23 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Modern Arabic vernaculars have relatively large, but varying, consonant inventories. Because of that, they have been typologized according to differences in the reflexes of their consonant phonemes—differences which suggest common origins or long-term contact (Watson 2011a, p. 862). The resulting dialect categories often coincide with various divisions: geographical (eastern–western), lifestyle (sedentary–Bedouin), ethno-religious, social (based on status, age, gender), as well as stylistic and historical. However, using such mixed classificatory devices has always been problematic. Not only do the various factors cross-classify the dialects, but, with persistent exceptions, they exhibit internal inconsistency as well (see Palva 2006 for a discussion of some of these challenges). Moreover, the outcome is largely descriptive. Works that have explored Arabic consonant variation from this perspective include Cantineau (1960), Fischer and Jastrow (1980), Holes (1995), and Kaye and Rosenhouse (1997).

While classifying languages or dialects according to the type of sounds they contain is a recognized approach to phonological typology, it has been criticized for proposing oversimplified groupings with no explanatory value for synchronic or diachronic facts (Kiparsky 2008; Drescher et al. 2018). A more theory-oriented, ‘property-driven’ outlook to typology, advanced by Hyman (2007, 2018), has the individual phonological traits, not language varieties as such, as the primary objects of comparison. In this approach, typology and theory must go hand in hand, and since modern phonological theory is multifaceted and pluralistic in nature, we must admit that any meaningful typology builds on a specific theoretical framework (Kiparsky 2018). This, in turn, means that there will always be several viable options to formulate a typology; hence, there is no such thing as a one-size-fits-all classification system. Within Arabic, theoretically motivated typologies of syllabification phenomena (e.g., Broselow 1992; Kiparsky 2003; Farwaneh 2009) and

of stress placement (e.g., Kiparsky 2000; Watson 2011b) have been more successful than segmental studies.

This article embraces the latter line of research, by which the typologies of Arabic consonants are couched within a theory of representation. So, rather than considering only the existing phoneme reflexes in one inventory as opposed to another, I explore phoneme classes in terms of which of their constituent features are active in the phonology. Representational typologies will be formulated in a minimalist and highly abstract model of feature geometry, which optimizes the use of a minimum number of contrast-relevant features. This model, I argue, affords one possible concrete scheme to correlate the consonant reflexes without resorting to the problematic, long-established categories. It also explains and predicts phonological behavior in a systematic and unambiguous way. I will demonstrate that the typological and traditional classifications can coexist, but only to relate the structural generalizations to what we already know. Apart from that, the two are methodologically incompatible.

The remainder of the paper is organized as follows. Section 2 introduces the theoretical model employed in the analysis. Section 3 develops the typologies of the varying Arabic consonantal phonemes, sketching the geographical distribution of each reflex and justifying its component features in accordance with phonological facts. Sections 3.1–3.7 treat the consonant prototypes *qāf*, *ḡīm*, *kāf*, *dād–dā*, interdental, rhotic, and pharyngeals (in that order). Section 4 discusses various implications of this type of analysis, both for the study of Arabic dialects and for phonological typology in general.

2. A Model for Feature-Based Typology

Typology in general is the classificatory study of languages according to their structural features; and by convention, phonological typology will group them according to the number and type of the phonemes they contain. This traditional view is challenged by Hyman (2007, 2018) who claims that typology is not about classifying languages but rather about characterizing linguistic properties across the linguistic spectrum. When this becomes the primary object of comparison, we move into what he calls property-driven typology. Under this view, the phonologist studying typology should not be interested in how phonological properties are distributed according to extra-linguistic factors. How to analyze the system of variation has been more of a priority for phonology than the ‘where’ question of traditional dialectology (Hyman 2018, pp. 14–15).

By focusing on the ‘how’, I will adopt a line of research that places dialect typologies within theories of phonological representation. Of course, features are the atoms of such representation. They are typically regarded as segment properties and as cross-classifying dimensions that characterize natural or phonologically active classes of segments. Moreover, there is solid evidence that features are arranged in some hierarchical structure, typically under higher-order categories known as ‘class nodes’, such as Place, Manner, and Laryngeal. This understanding of features constitutes the premise for most models of feature geometry (e.g., Clements 1985; Sagey 1986; McCarthy 1988, inter alia). The property-driven analysis of typology in this paper is feature geometric in nature.

Any feature-based theory of typology is potentially undermined by the different assumptions about the nature of the hierarchy or the very set of phonological features upon which it is based (Gordon 2016, p. 71). This should not be a problem, however, if we acknowledge that “there are no theory-neutral grammars, and consequently no theory-neutral typology” (Kiparsky 2018, p. 54). There is no contradiction, Kiparsky argues, that typological generalizations are the product of linguistic theory while they themselves are theory-dependent. The criterion to generate an informed theory-specific typology is thus to ensure that categories are founded on “independently justified linguistically significant representations” (Kiparsky 2018, p. 55). This is a fundamental principle of the framework I am going to employ here.

My analysis of Arabic consonant reflexes is couched in the Parallel Structures Model of feature geometry (PSM; Morén 2003, 2006, inter alia). The PSM is a minimalist framework

in which consonants and vowels have parallel structures and identical, broadly defined features for place and manner articulations. It integrates insights from various other proposals, in particular Unified Place Theory (Clements 1991; Clements and Hume 1996), Element Theory (Harris and Lindsey 1995), and Dependency Phonology (Anderson and Ewen 1987). Features in the PSM are monovalent and exclusively distinctive, i.e., present only if they are necessary to maintain phoneme contrasts and/or are active in the phonology (cf. Clements 2001). In this sense, a PSM analysis is also congruent with the Contrastivist Hypothesis (Hall 2007; Drescher 2009).

How a terminal feature is interpreted in the PSM hinges on its association to a superordinate class node in the hierarchy. As diagrammed in Figure 1, each place or manner feature can be represented under two separate nodes/tiers, with the V-node being dependent on the C-node. This symmetry aims to establish a unified machinery that captures consonant–vowel interactions as well as acoustic/articulatory parallelisms in natural language. To explain their asymmetries, consonants can have both C- and V-features, while vowels can only have the latter. Another architectural mechanism of the model is building segmentally complex structures from simpler ones, which, together with the dependency principle, allows for a high degree of economy in the feature system.

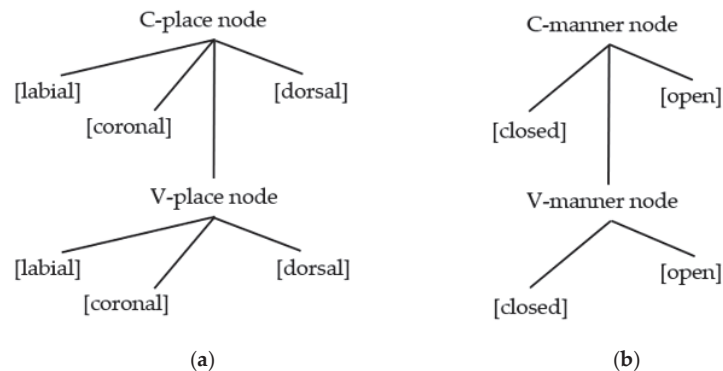


Figure 1. Basic PSM geometry. (a) Place tier; (b) Manner tier.

The Place and Manner tiers in the PSM deserve some attention. Under the Place tier (a), we use the articulator-based features [labial], [coronal], and [dorsal] under the C-place node and its daughter V-node (cf. Clements 1991). Simple consonants have one place feature; complex consonants have multiple features on the same place node; and consonants with secondary articulation have features on both C-place and V-place nodes (Morén 2003, pp. 199, 233). Similarly, under the Manner tier (b), we make use of the loosely defined features [open] and [closed], which can be attached to a C-manner or a V-manner node in arrangements that reflect the relative sonority of segments (Morén 2003, pp. 222–23). As for a Laryngeal tier, it should suffice here to use the feature [voice] to differentiate voiced from voiceless obstruents (see Morén 2003, p. 230).¹

Specifying the above features to a particular segment depends on finding positive evidence in the relevant variety. When varieties are closely related, phonological activity will show major parallels. This, in addition to the universal phonetic properties of speech sounds, means that a given segment will have the same composition across varieties of the same language, unless there is proof to the contrary. Because of this, a contrast-based model like the PSM is a valid tool in drawing typologies, as I will demonstrate in the next section.

3. The Typologies of Arabic Consonant Reflexes

The present study provides feature-based typologies of *q, *ḡ, *k, *d~*ḏ, *θ~*ḥ, *r, and *ħ~*ʕ, which display variation across Arabic dialects. Reflexes of these consonants can be differentiated representationally along the dimensions of place and manner of articulation,

in addition to voicing. As mentioned above, I will draw feature geometric generalizations of these reflexes based on the contrast and phonological activity they exhibit. The facts and data denoting their behavior appear in various studies on individual Arabic dialects as well as in cross-dialectal surveys, as the references indicate.

Before embarking on the analysis, a few remarks are in order. First, when discussing segmental composition in the PSM, it is crucial to separate underlying from surface levels. Our concern here is the major reflexes that have a contrastive phonemic status, i.e., not predictable allophones nor marginal phonemes that exist only in free variation. Although these types of sounds will not be analyzed representationally, they will be mentioned and clearly labeled for what they are, so that confusions are avoided. It is still, however, a challenge to decide which variant, in case of multiple allophones, should be treated as the basic phoneme, and whether sounds confined to loanwords should be included in the phoneme inventory (cf. Gordon 2016, p. 43). There is also the issue of how to deal with several reflexes of one consonant cooccurring in the same dialect.

The answer to the above questions will vary depending on the available evidence in each case: the nature of the environment in which the variants occur, the existence of minimal pairs, the extent and stability of the borrowings, etc. We will see, for example, that /k/ and /dʒ/ are the basic phonemes in dialects that exhibit [k]–[tʃ] and [dʒ]–[j] alternations, simply because the [tʃ] and [j] allophones are restricted to front vowel contexts, while the other two occur elsewhere. We will also learn that many, but not all, of the dialects with /g/ and /ʔ/ reflexes of *q have retained a /q/ phoneme in both stable and more recent loans from Standard Arabic (SA), sometimes leading to minimal pairs or morphological doublets. And even aside from direct borrowing, the diglossic coexistence of dialects with SA often leads to the reintroduction of SA phonemes into their inventories.

In the coming subsections, I will examine reflexes for each of the consonant prototypes listed above, describing their geographical distribution, but more importantly their phonological behavior: the phonemes they contrast with and the processes they participate in. This behavior serves as the basis for assigning their PSM feature-geometric structures, which are the building blocks of the new representational typologies I propose in this work.

3.1. The Qāf Typology

Cantineau (1960, p. 68) states that “the pronunciation of qāf is of utmost importance” in the classification of Arabic dialects. Four major reflexes, /q ʔ k g/, are often named in the literature. A widely accepted generalization is that dialects with the voiceless cognates /q ʔ k/ are spoken by sedentary people, while those with the voiced /g/ are spoken by Bedouin or Bedouin-descended populations (Watson 2011a, p. 859). However, this principle is not without exceptions; for instance, both in North Africa and the Levant we encounter urban dialects with /g/, and in reality, every geographical region has a distinct pattern of variation (see Bahloul 2007). Let us briefly examine each of these four reflexes, in pursuit of a representational typology of *q. For a thorough overview of geographical distributions, see Cantineau (1960, pp. 68–71), Kaye and Rosenhouse (1997, pp. 270–73), Bahloul (2007), and Edzard (2009).

Looking first at the voiceless uvular /q/ reflex, it is most notably attested in the sedentary dialects of Syria and the Maghreb, as well as qəltu Mesopotamian and parts of Oman and Yemen (Fischer and Jastrow 1980, p. 52). Some examples are [trawwaq] ‘he had breakfast’ (Latakia, Syria), [bqa] ‘he stayed’ (Morocco), and [qas¹i:ɪ] ‘short’ (Mosul, Iraq). Based on phonological activity in these dialects, /q/ can be treated as a member of the natural class of primary dorsal segments. It often patterns with velar stops in triggering nasal place assimilation (NPA) toward a back nasal, as in /manqal/ > [manqal] ‘brazier’, and totally assimilates adjacent velar/uvular fricatives in qəltu dialects, as in /ʔaqqas/ > [ʔaqqas] ‘bald’ (Youssef 2019, p. 26).

With no trace of phonological activity that discriminates velars and uvulars, I infer that there is a single natural class of C-place [dorsal] consonants. Of these, uvular /q/ is most suitable for a mannerless segment, i.e., with a bare place feature, since it patterns with

both stops and fricatives. Phonetically, [dorsal] is a fitting feature since /q/'s posterior articulation is known to cause lowering or backing of all immediately adjacent vowels (cf. Al-Ani 1970, pp. 32–33).

By far the most common reflex of *q is the voiced velar stop /g/, which is characteristic of Bedouin dialects (Watson 2002, p. 17). This reflex covers eastern/central Arabian Peninsula and southern Iraq, but also significant pockets in North Africa, Upper Egypt, Sudan, the Levant, and southern Peninsula (Bahloul 2007). As implicated above, /g/ belongs to the class of C-place [dorsal] consonants, as it triggers NPA, e.g., /ji-ngar/ > [jɪŋgar] 'he pecks' (Muslim Baghdadi; Youssef 2013, p. 67); it also partakes in the labialization of /i/ to [u] in certain Iraqi and Levantine dialects (see e.g., Haddad 1984 and Youssef 2015). Further, if we assume that a C-manner [closed] feature indicates a stop constriction in the PSM, then /g/ is specified for this feature as well as [voice], so that it is distinguished from /q/ and /k/.

The second most widespread reflex is the glottal stop /ʔ/, mainly attested in urban centers of the Levant and Lower Egypt, and sporadically in some Maghrebi city dialects (Holes 1995; Bahloul 2007), but also in rural areas especially in Lebanon (Fischer and Jastrow 1980, p. 52); examples: [ʔa:d^si] 'judge' (Cairo); [ʔil-ʔuds] 'Jerusalem' (Beirut); and [rifʔa:t] 'friends' (Damascus). Given that all other stop consonants show contrastive evidence for a place feature, we are left with /ʔ/, the Arabic epenthetic consonant, to assign a single C-manner [closed] feature. From an articulatory standpoint /ʔ/ is simply a stop formed with complete closure between the vocal folds.

A voiceless velar stop cognate, /k/ (sometimes appearing as emphatic /k/), is generally marked as ruralite; and Edzard (2009) notes that it surfaces in those dialects which have affricated the original *kāf* (see Section 3.3 below). It is typical of central Levantine villages, but also in areas of North Africa (Watson 2011a, p. 862). We find, for example, [kalb] 'heart'; [ka:l] 'he said'; and [karji] 'village' (rural Palestinian). Representationally, /k/ is the voiceless counterpart of /g/, and it participates in the same processes: NPA (producing a velar nasal) and labialization (Herzallah 1990). We infer, then, that it is specified for the features C-place [dorsal] and C-manner [closed].

There also exists a number of conditioned variants, which are not included in the analysis because they appear to be the more restricted subsidiary allophones of one of the main reflexes above. For instance, certain eastern Arabian nomadic dialects affricate their /g/ to [dʒ] and further to [dz], but only in front vowel contexts (Johnstone 1967). Alternations such as [ga:l] 'he said' vs. [t^sari:dʒ] 'road' (southern Iraqi) and [t^sari:dz] (Šammari, central Saudi) led Holes (1995, p. 60) to classify the /g/ group into three subtypes of Bedouin dialects.

Another marginal variant is the voiced uvular fricative [ɣ], which appears to be in free variation with [q] in parts of southern Iraq and the Arabian Gulf, e.g., [ɣada]-[qada] 'lunch' and [qit^sa:r]~[ɣat^sa:r] 'train' (Fischer and Jastrow 1980; see also Al-Nassir 1993, p. 40). Moreover, many dialects with one of the major reflexes /ʔ k g/ preserve /q/ in a number of borrowed words from SA, sometimes giving way to semi-contrasts like [wɔr^sga] 'tree leaf'~[worqa] 'sheet of paper' (Moroccan Bedouin; Cantineau 1960, p. 70).

Table 1 offers a restatement of the *qāf* typology in Arabic, with a rough geographical distribution of the four major cognates. Using this feature typology, we can simply refer to dialects with a *q reflex that has all or a subset of the features named. The specifications both reflect and explain each segment's synchronic phonological behavior. And although not the focal point here, historical shifts from one reflex to the other could also be motivated through feature loss or gain.

Table 1. Representational typology of the major *q phoneme reflexes.

	C-place [dorsal]	C-manner [closed]	... [voice]	Geographical Distribution
/q/	✓			Various sedentary: North Africa, Mesopotamian <i>qaltu</i> . . .
/ʔ/		✓		Urban Egyptian and Levantine and sporadic Maghrebi
/k/	✓	✓		Ruralite Levantine dialects
/g/	✓	✓	✓	Bedouin(-origin) dialects

3.2. The Ġīm Typology

Another famously varying consonant is *ġīm*, with the three major reflex phonemes /dʒ ʒ g/. The first two are the most widespread pronunciations, and, broadly speaking, /dʒ/ is characteristic of Bedouin dialects, while /g/ and /ʒ/ are sedentary. Exceptionally, however, /ʒ/ is the predominant reflex in North Africa, irrespective of the sedentary–nomadic split (Cantineau 1960, p. 59). Below I discuss each of the *ġ cognates separately. Detailed geographical typologies are provided in Cantineau (1960, pp. 58–60), Fischer and Jastrow (1980, p. 51), Holes (1995, pp. 61–62), and Zaborski (2007).

The voiced palatoalveolar affricate /dʒ/ is standard “in the majority of eastern Bedouin dialects, in rural dialects of the Levant and Mesopotamia, in the majority of dialects in central Yemen, and in some sedentary dialects in Algeria” (Watson 2011a, p. 863).² Phonologically, /dʒ/ is the voiced counterpart of /tʃ/ in dialects that have developed the latter phoneme through borrowings and historical affrication (Fischer and Jastrow 1980), with minimal pairs like (Baghdadi) [tʃanna] ‘daughter-in-law’–[dʒanna] ‘paradise’ and [ʃuratʃ] ‘brushes’–[ʃuradʒ] ‘he dispelled’; hence, it has [voice].

Two different phonological processes provide evidence that /dʒ/ is coronal. One is that it typically participates in the assimilation of the definite article (L-ass), as one of the ‘sun letters’, e.g., /l-dʒiba:l/ > [dʒ-dʒiba:l] ‘the mountains’ (though not in SA). The other is that it tends to assimilate partially to a following coronal obstruent in onset clusters, producing a fricative [ʒ], with possible devoicing to [ʃ], e.g., /dʒtima:ʔ/ > [ʃtima:ʔ] ‘meeting’ (Iraqi; Youssef 2013, p. 69). Because /dʒ/ is a blocker of emphasis spread (ES) in many dialects, /dʒ/’s coronality is interpreted as a secondary feature, i.e., V-place [coronal], in conflict with the secondary emphatic feature (cf. Davis 1995). Lastly, since affricates behave phonologically as stops, we specify /dʒ/ for C-manner [closed] as well.

The second most frequent reflex is the voiced palatoalveolar fricative /ʒ/, attested in the urban dialects of the Levant (exceptions include Aleppo and most of Jordan, which have /dʒ/) and most urban and non-urban Maghrebi dialects (Zaborski 2007). It is the voiced equivalent of /ʃ/, as seen in the minimal pair [ʒa:j] ‘coming’–[ʃa:j] ‘tea’. It is always a trigger of L-ass, e.g., [ʒ-ʒami:l] ‘the pretty’ (Lebanese), hence coronal, but it also patterns with the ES blockers (see above), hence V-place [coronal].

As discussed above, /ʒ/ results from the assimilation of /dʒ/ to a coronal obstruent. Since /dʒ/ has both V-place [coronal] and [voice], the only way to distinguish it from /ʒ/ is constriction. Parallel to the stops, we may hypothesize that a C-manner [open] feature marks fricative constriction for /ʒ/ and all other consonants with a similar manner of articulation. Mustafawi (2017, p. 15) argues that “the best alternative for /dʒ/ while keeping most of its distinctive features would be /ʒ/”.

A voiced velar stop /g/ is found in Cairo, in rural central and northeastern Delta, and in all urban centers of northern Egypt down to Bani Swēf, but also in various Bedouin dialects of central Arabia and in some Yemenite and Omani dialects (cf. Watson 2002, p. 16; Zaborski 2007, p. 494). The /g/ reflex is often thought to be “the most salient feature of Egyptian speech across the Arab-speaking world” (Holes 1995, p. 61). Phonologically, /g/ has a stop constriction; it contrasts with voiceless /k/, e.g., [gu:ʔ] ‘hunger’–[ku:ʔ] ‘elbow’; and it triggers NPA, e.g., /fiŋga:l/ > [fiŋga:l] ‘coffee cup’ (Cairene; Youssef 2013, p. 35).

We may conclude, then, that /g/ has the following contrastive features: C-place [dorsal], C-manner [closed], and [voice].

As I have noted for *qāf*, there are a few conditioned and marginal variants of *ǧīm*, which, although excluded from the featural analysis, are worth mentioning here. Perhaps the most well known is a palatal approximant [j], found mainly in Bedouin dialects of the Gulf and lower Iraq, which is partly in free variation with [dʒ] and partly lexically conditioned (Zaborski 2007). Despite the variability, e.g., [jarju:r]~[dʒardʒu:r] ‘shark’ or [ʕaj:i:n]~[ʕadʒ:i:n] ‘dough’ (Bahraini), [j] is considered a marker of Gulf speech (Holes 1995, p. 62). Other notable variants include an alveolar stop [d] in some Upper Egyptian dialects in front of liquids and nasals (Behnstedt and Woidich 1985), two affricates: palatoalveolar [tʃ] in Palmyra and alveolar [ts] in the oasis of Suhne (Syria), and a fricative [z] in some Jewish dialects of the Maghreb (cf. Fischer and Jastrow 1980).

Table 2 summarizes and restates the *ǧīm* typology in terms of five contrastive features, which are assigned based on synchronic phonological activity. We now realize that urban Egyptian dialects, which have a glottal stop reflex of *q and a /g/ reflex of *ǧ, have exploited the features C-place [dorsal] and C-manner [closed] to differentiate segments in their inventories. Historically, claims that the Proto-Semitic origin of *ǧ is indeed a velar plosive /g/ (see e.g., Roman 1981) can be also explained by a place of articulation shift from C-place [dorsal] to V-place [coronal] in /dʒ/, while keeping all other features intact.

Table 2. Representational typology of *ǧ phoneme reflexes.

	C-place [dorsal]	V-place [coronal]	C-manner [closed]	C-manner [open]	... [voice]	Geographical Distribution
/dʒ/		✓	✓		✓	Bedouin(-origin) dialects
/g/	✓		✓		✓	Lower Egyptian and sporadic Peninsular
/ʒ/		✓		✓	✓	Urban Levantine and most of Morocco

3.3. The Kāf Typology

This consonant exhibits conditioned and unconditioned variation in modern Arabic dialects. The former type—which concerns us here—affects the *k regardless of neighboring sounds and is due to advancement of /k/'s place of articulation, which makes it prone to affrication and spirantization (Cantineau 1960, p. 66), resulting in /tʃ/. Conditioned alternations produce a [tʃ] or a [ts] allophone of /k/ in the vicinity of front vowels, and [k] elsewhere. As before, we concentrate on phonemic reflexes for our phonological analysis, namely /k/ and /tʃ/, but will also mention the allophonic pattern for the purpose of comparison. Elaborate surveys can be found in Cantineau (1960, pp. 66–67), Johnstone (1967), and Kaye and Rosenhouse (1997, pp. 273–74).

On the one hand, most Arabic varieties from east to west have preserved a velar stop /k/ as the only reflex available. In the east, this is generally viewed as an urban feature, while in Egypt and westwards, the lifestyle factor is insignificant as there is little to no variation observed (Palva 2006, p. 606). Among the dialects with a /g/ phoneme, either as a reflex of *q or *ǧ, /k/ is its voiceless cognate; thus, it has no voicing specification. The contrastive features for /k/ have already been discussed in Section 3.1: a velar point of articulation corresponds to C-place [dorsal], and a stop constriction corresponds to C-manner [closed].

In various ruralite dialects of the Levant, a voiceless palatoalveolar affricate cognate, /tʃ/, is attested, irrespective of the phonological environment (Watson 2011a, p. 873). More specifically, this is the case in central Palestine, a few Syrian villages, and two regions of Algeria, as well as among the Shiites of Bahrain (Fischer and Jastrow 1980, pp. 51–52). Moreover, several Bedouin dialects seem to have regularized affricate /tʃ/

within roots; and although there often remains few [k]–[tʃ] alternations, one can safely pose two phonemes, /k/ and /tʃ/, in contrast. In Muslim Baghdadi (Youssef 2014), for instance, we encounter minimal pairs like [tʃuwa] ‘he scorched’–[kuwa] ‘he ironed’ and [ba:tʃir] ‘tomorrow’–[ba:kir] ‘virgin’. And in some rural Jordanian elderly speech (Cantineau 1960), extensions of [tʃ] to non-front vowel contexts occur as a result of analogy, e.g., [di:tʃ] ‘rooster’ > [dju:tʃ] ‘pl’.

More widespread are the conditioned alternations where either [tʃ] or [ts] occurs in front vowel contexts in complementary distribution with [k], with no morphological repairs, and are thus regarded as allophones of the /k/ phoneme (cf. Holes 1995, p. 60). The [tʃ] variant is attested in the Bedouin north Arabian and related dialects of Jordan and Iraq (Fischer and Jastrow 1980), with alternating examples like [ritʃib] ‘he mounted’–[jirkab] ‘he mounts’.³ The [ts] variant is predominant in central Najdi, among the ‘Anaiza and Šammar tribes (Cantineau 1960, p. 67), e.g., [tsaff] ‘palm of the hand’–[kfu:f] ‘pl’.

Let us now discuss the featural composition of the phonemic /tʃ/ reflex, drawing mainly on Youssef (2014). First, note that all /tʃ/-dialects have a /dʒ/ cognate of *ǧ, the two forming a phoneme pair that differ in terms of voicing (see above); so in other respects, they should have comparable phonological status. On the one hand, /tʃ/ is necessarily coronal because it triggers L-ass, as in [tʃ-tʃa:ku:tʃ] ‘the hammer’. On the other, affricates are stops phonologically, so /tʃ/ is also assigned C-manner [closed].

The proposed feature composition may also reflect the historical development of affricate /tʃ/ in the relevant dialects. If we treat affrication as a shift from velar to coronal that was once motivated by adjacent high vowels /i i:/ or palatal /j/, and if these triggers are specified for V-place [coronal], being blockers of ES, then the output of the assimilation process, namely /tʃ/, must also have the latter feature, while C-manner [closed] remains unchanged (cf. Watson and Dickins 1999). A concise representational typology of *kāf* is given in Table 3.

Table 3. Representational typology of *k phoneme reflexes.

	C-place [dorsal]	V-place [coronal]	C-manner [closed]	Geographical Distribution
/k/	✓		✓	Most Mashriqi and Maghrebi dialects
/tʃ/		✓	✓	-Various ruralite Levantine dialects -Some Bedouin dialects: vs. /k/

3.4. The Interdental Typology

Here we will be dealing only with the plain interdentals *tā* and *dāl*; emphatic *dāl* will be discussed in the next section. A general principle is that Old Arabic /θ ð/ are preserved in Bedouin-type dialects and merged with the corresponding alveolar stops /t d/, and less frequently with alveolar /s z/ or labiodental /f v/ fricatives, in sedentary speech (Cantineau 1960, p. 44). However, this dichotomy encounters numerous exceptions. For example, all dialects in Morocco seem to have shifted to stops (ibid.), while a few city dialects (e.g., Tunis, Mosul, Mardin) have retained the interdentals (Fischer and Jastrow 1980, p. 50). Below, I will individually examine the three pairs of reflexes; for a full overview, see Cantineau (1960, pp. 44–45), Fischer and Jastrow (1980, p. 50), and Mustafawi (2017, pp. 14–15).

What we may call ‘the preservation dialects’ constitute all “Bedouin dialects, dialects of Bedouin origin, the rural sedentary dialects of central Palestine/Jordan, Tunisia and Mesopotamia, and [. . .] all but the western coastal city dialects of the Peninsula” (Watson 2011a, p. 863): an assortment of dialects, if one assumes traditional dichotomies. In all of these, both /θ/ and /ð/ participate in L-ass, e.g., [θ-θo:b] ‘the shirt’ and [ð-ðahab] ‘the gold’, hence C-place [coronal]. Considering that /θ ð/ are non-sibilants, with relatively

weak turbulence, we may propose that they are devoid of manner features (and thus featurally distinct from the sibilants /s z/). Furthermore, the two consonants contrast in voicing, which means that /ð/ is marked for additional [voice].

The majority of urban dialects, as well as many neighboring rural areas, have the dental/alveolar stop cognates /t d/ (Fischer and Jastrow 1980). This vast isogloss covers all of Morocco, all sedentary dialects of Egypt, Hijazi Arabic, and the rest of the Levant (Mustafawi 2017, p. 14). Concerning their featural content, there is good indication that /t d/ are C-place [coronal]. In Cairene (Youssef 2013), for instance, they trigger L-ass, e.g., [ʔit-tiʔi:] ‘the heavy’, [ʔid-de:l] ‘the tail’; and they regressively assimilate to labial and velar stops across word boundaries, e.g., /baʕat kita:b/ > [baʕak kita:b] ‘he sent a book’, /nafab bi-gildu/ > [nafab bi-gildu] ‘he saved his skin’. As stops, they are also specified for C-manner [closed], and /d/ has yet another [voice] feature.

In various northern Mesopotamian dialects, as well as in the Arabic of Afghanistan and Uzbekistan, the development is toward the alveolar sibilants /s z/ (Jastrow 1978), as in [sa:sə] ‘three’, [ʔaxaz] ‘he took’ (Āzəx, Anatolian). These sibilants also tend to replace /θ ð/ in borrowings from SA in the urban dialects of Egypt and the Levant (Mustafawi 2017, e.g., [jisbit] ‘he proves’, [ʔiza:ʕa] ‘broadcasting’ (Aleppo). The /s z/ pair partakes in L-ass, voicing assimilation, and often sibilant assimilation. We can therefore specify them for C-place [coronal], being alveolars, and C-manner [open], being fricatives; with an extra [voice] feature for /z/.

Another known pair of cognates are the labiodental fricatives /f v/, attested in Siirt (southeastern Anatolia), e.g., in [fa:fe] ‘three’ and [vahab] ‘gold’ (Jastrow 1978, pp. 34–39), in some nomadic dialects of the Tell Atlas Mountains, and in Palmyra (Cantineau 1960, p. 45). In the Shiite dialect of Bahrain, only a /f/ reflex of *θ is attested (Mustafawi 2017, p. 15). The /f v/ reflexes form a voiceless-voiced pair; and I further assign them C-place [labial], as they would be expected to trigger NPA, and C-manner [open], which characterizes fricatives.

A crucial point to notice is that in all but the preservation dialects, the change is that of merger with an already existing phoneme—a fact simply built into the feature typology in Table 4. We can also make sense of Cantineau’s (1960, p. 44) observation that the sedentary dialects which pronounce *q as /q/ have retained the interdental. It appears that such dialects have a preference for reflexes with no manner features. Historically, in addition, the cross-linguistically common sound changes /θ ð/ > /t d/ or /s z/ are effortlessly explained as insertion of manner features.

Table 4. Representational typology of *θ–*ð phoneme reflexes⁴.

	C-place [labial]	C-place [coronal]	C-manner [closed]	C-manner [open]	... [voice]	Geographical Distribution
/θ ð/		✓			(✓)	Bedouin(-origin) and few rural dialects
/t d/		✓	✓		(✓)	Sedentary dialects and all of Morocco
/s z/		✓		✓	(✓)	Peripheral and northern Mesopotamian
/f v/	✓			✓	(✓)	Sporadic: Siirt, Tell Atlas, Palmyra ...

3.5. The Ḍād–Ḍā Typology

Next are the emphatic consonants denoted by the Arabic letters *ḏād* and *ḏā*, which in the modern dialects either appear as two distinct phonemes, respectively alveolar stop /d^ḏ/ and fricative /z^ḏ/, or merge into a single interdental fricative /ð^ḏ/. The former is characteristic of sedentary dialects and the latter of nomadic dialects. Historically, these

two sets of dialects have restructured the asymmetrical Old Arabic system in different ways (Holes 1995), as we will see below. Both historical and synchronic surveys are provided in Holes (1995, pp. 57–59), Versteegh (2006), and recently Hamdan and Al-Hawamdeh (2020).

We start with dialects maintaining a contrast between /d^ʕ/ and /z^ʕ/. These coincide unmistakably with city dialects that have neutralized the interdental fricatives /θ ð/, merging them with the corresponding alveolar stops /t d/ (Holes 1995, p. 58). These dialects are said to have a dyadic (binary) system, with voiceless–voiced series for both plain and emphatic consonants, i.e., /t d/, /t^ʕ d^ʕ/, /s z/, /s^ʕ z^ʕ/ (Bellem 2014).⁵ Representationally, /d^ʕ z^ʕ/ are emphatic consonants that trigger long-distance ES, e.g., in [tə-xfid^ʕ-aṭ] ‘discounts’ or [ʕaz^ʕama] ‘greatness’ (Cairene; Watson 2002, p. 273).

Emphatics are distinguished from their plain counterparts by an additional non-primary back articulation (Davis 1995, p. 472). Youssef (2006, 2013) posits V-place [dorsal] to characterize this natural class. This way, [dorsal] alone, on separate tiers, is used to account for velar/uvular and emphatic consonants, which is clearly more economical than introducing an additional [pharyngeal] (McCarthy 1994), [guttural] (Watson 2002), or any other feature proposed specifically for Arabic or Semitic. It is worth mentioning that McCarthy (1994) has also suggested [dorsal] as a redundant feature for emphatics.

The emphatics generally have C-place [coronal] as their primary articulation; /d^ʕ z^ʕ/ do trigger L-ass, e.g., [ʔid^ʕ-d^ʕaj^ʕa] ‘the village’, [ʔiz^ʕ-z^ʕari:f] ‘the pleasant’ (Damascene). Further, /d^ʕ z^ʕ/ are specified for [voice], as they contrast with voiceless /t^ʕ s^ʕ/ in the dyadic system. In terms of manner of articulation, /d^ʕ/ is a stop, with C-manner [closed], and /z^ʕ/ is a fricative, with C-manner [open].

The other group of dialects, where /d^ʕ/ had fallen together with /ð^ʕ/, are mainly Bedouin or have a Bedouin origin, such as *gilit* Mesopotamian, Yemenite, and Peninsular—essentially dialects that have retained the plain interdentals (cf. Embarki 2008, p. 592).⁶ This merger has engendered confusion in defining minimal pairs that used to contrast /d^ʕ/–/ð^ʕ/, e.g., [fa:jjð^ʕ] ‘overflowing/ usury’ and [ð^ʕʕfar] ‘he plaited/ overcame’ (Baghdadi; Youssef 2013, p. 131). These dialects are said to have reduced the asymmetry of the system by developing triadic series, with two three-member sets of voiceless–voiced–emphatic cognates: alveolar plosives /t d t^ʕ/ and interdental fricatives /θ ð ð^ʕ/ (Holes 1995, p. 58; see also Bellem 2014). The featural makeup of /ð^ʕ/ should now be easy to deduce: C-place [coronal], as a trigger of L-ass, V-place [dorsal], as a trigger of ES, and [voice]. And just like the plain interdentals (cf. Table 4) it need not be specified for C-manner.

It is probable that *dād* was historically a voiced lateral/lateralized interdental fricative emphatic (cf. Corriente 1978). A remnant of this is apparently the pronunciation of **d* as emphatic lateral /l^ʕ/ in a few dialects of southern Arabia, such as the Saudi Tihāma (Al-Azraqi 2010) and the Yemeni dialect of Daḥīna (Landberg 1905–1913, cited in Versteegh 2006). I will not pursue an analysis of this marginal reflex here, although my presumption is that it is featurally identical to /ð^ʕ/. Table 5 recapitulates the featural composition of the three major phonemes discussed above: the contrastive /d^ʕ z^ʕ/ and their merged reflex /ð^ʕ/.

Table 5. Representational typology of **d*–**d* phoneme reflexes.

	C-place [coronal]	V-place [dorsal]	C- manner [closed]	C- manner [open]	... [voice]	Geographical Distribution
/d ^ʕ / < * <i>d</i>	✓	✓	✓		✓	Sedentary dialects: vs. /z ^ʕ /
/z ^ʕ / < * <i>d</i>	✓	✓		✓	✓	Same as above: vs. /d ^ʕ /
/ð ^ʕ /	✓	✓			✓	Bedouin(-origin) dialects

3.6. The Rhotic Typology

Most Arabic dialects have a rhotic phoneme /r/, corresponding to the letter *rā*, which is typically realized as a voiced alveolar tap or trill (Younes 1994; Watson 2002). However, two groups of dialects have introduced a phonemic split whereby a new emphatic /r^ʕ/ or uvular fricative /ʁ/ contrasts with a plain /r/ phoneme. A third group has only an emphatic /r^ʕ/ reflex, a fourth has a plain /R/ with a double place of articulation, and a fifth has just a plain /r/. The first four types are thoroughly examined in Youssef (2019, forthcoming); below I provide a synopsis.

Type-I dialects have established two distinct phonemes in contrast, a plain /r/ vs. an emphatic /r^ʕ/, and are therefore dubbed ‘the split-*r* dialects’. They mainly comprise the Arabic dialects of Africa, which include the Maghrebi and Egyptian families, and a few peripheral dialects in sub-Saharan Africa (but also in Anatolia). Minimal pairs are abundant, e.g., [r^ʕɑ:jib] ‘curdled’–[ra:jib] ‘collapsed’ (Moroccan), [ʔar^ʕbaʕ] ‘a Wednesday’–[ʔarbaʕ] ‘he guzzled’ (Egyptian), and [kar^ʕɑ] ‘he was seen’–[kara] ‘he rented’ (Mardin). Additionally, [r^ʕ] and [r] exist partly in the same environments, suggesting that they have parallel distribution.⁷

The phonemes /r r^ʕ/ trigger L-ass, e.g., [ʔr^ʕ-r^ʕɑ:ʒɔl] ‘the man’, [ʔr-razwar] ‘the shaver’ (Moroccan); hence, they are C-place [coronal]. They also trigger coronal sonorant assimilation (CSA), whereby /n l/ assimilate regressively to a following /r r^ʕ/ across word and morpheme boundaries, e.g., /min riglu/ > [mir riglu] ‘from his leg’ (Cairene). The inference is that /r r^ʕ/ are sonorants, for which we may assign a composite of C-manner [open] and V-manner [closed] (see Morén 2006, p. 1210), denoting that sonorants are continuants (open) and vowel-like (sonorous). Finally, emphatic /r^ʕ/ in this group is a trigger of ES, with the same bidirectional, long-range spreading of pharyngealization as other primary emphatics, e.g., [ʔar^ʕabijj-ɑ:t-ak] ‘your cars’ (Cairene). We thus assign it a secondary V-place [dorsal] feature in addition.

Type-II dialects have a single, emphatic /r^ʕ/ phoneme, and incorporate the Levantine dialects spoken in Syria, Lebanon, Palestine, and Jordan. The phoneme has emphatic [r^ʕ] and plain [r] allophones in complementary distribution, and there is no sign of a phonemic split. Distributional evidence that the phoneme is /r^ʕ/ and not /r/ includes the fact that it does not trigger vowel raising (*imāla*), e.g., [dʒɔ:r^ʕɑ] ‘hole’ rather than *[dʒɔ:ri] (rural Palestinian; see Younes 1994 for details). Furthermore, /r^ʕ/ patterns with other emphatics in inducing ES, although it partially differs in its more limited domain and vulnerability to undergo de-emphasis (Younes 1993; Davis 1995); and it participates in L-ass, as well. Therefore, it has C-place [coronal] and V-place [dorsal]. It is a sonorant, as it triggers CSA, e.g., /le:l r^ʕɑ:jig/ > [le:r^ʕ r^ʕɑ:jig] ‘a calm night’ (Jordanian), so we add C-manner [open] and V-manner [closed].

Type-III dialects have one /R/ phoneme, which is underlyingly non-emphatic, yet arguably both coronal and dorsal. They belong to the Peninsular and Mesopotamian *gilit* groups. Here, the /R/ phoneme has fully predictable plain and emphatic realizations; the emphatic allophone causes backing of adjacent low vowels only (Al-Ani 1970, p. 33), which implies low-level coarticulation rather than ES. As expected, /R/ obligatorily triggers L-ass, e.g., [R-Ri:ʔa] ‘the smell’ (Muslim Baghdadi; Erwin 2004); besides, a process of labialization in these dialects shows that /R/ behaves more like velar/uvular than emphatic triggers (cf. Youssef 2015). We infer that /R/ is specified for two primary places of articulation, C-place [coronal] plus [dorsal], but no secondary articulation. In addition, it patterns with the coronal sonorants in CSA; therefore, it also gets the usual manner features for sonorants.

The remarkable type-IV group exhibits two distinct phonemes, an alveolar sonorant /r/ and a uvular fricative /ʁ/, and comprises primarily the Mesopotamian *qəltu* dialects, spoken in various cities in Iraq. In those dialects, the uvular /ʁ/ reflex of **r* coincides and merges with etymological *ḡayn*, whereas /r/ is found principally in loanwords (Blanc 1964; Jastrow 1978). Distributional evidence for two phonemes includes minimal pairs, e.g., [rakkib] ‘he let climb’–[ʁakkib] ‘he assembled’ (Mosul), [farrɑq] ‘he distinguished’–[fɑʁʁɑq] ‘he separated’ (Jewish Baghdadi).

Further phonological processes of assimilation, vocalization, and dissimilation take place to resolve some unusual contacts between uvular /ʁ/ (from *r) and the back consonants /q x ʁ/. If these processes are motivated by an OCP violation, we can propose that /ʁ/ is specified for C-place [dorsal]. Since /ʁ/ also behaves as a fricative and contrasts with voiceless /x/, we can assign additional C-manner [open] and [voice] features. As for the /r/ phoneme, it triggers both L-ass and CSA, so I propose C-place [coronal] together with the two (sonorant) manner features. It does not trigger emphasis spread, though.

We may also add a fifth group for dialects with just a plain /r/ reflex, which contains several Yemeni and Peninsular dialects, as well as peripheral dialects that have lost all emphatic versus plain contrasts in their inventories, e.g., Maltese, Cypriot, Uzbekistani, Juba, and Ki-Nubi. In Sanʿāni, for instance, the allophone [r^ʕ] is only found in proximity of an emphatic obstruent; elsewhere, it is realized as a plain [r], even in words such as [ras] ‘head’ and [ħarr] ‘hot’ (Watson 2002, p. 16). For this group, the /r/ is representationally similar to the /r/ of types I and IV above.

The rhotic typology provides an interesting case where it is hard to rely on surface forms at the expense of actual phonological behavior. This behavior is disclosed in the feature representations of the various reflexes, summarized in Table 6. Variability is due to the general elusive nature of rhotics (see Wiese 2001) and in Arabic, additionally due to the involvement of the notorious emphatic–plain distinction (Youssef, forthcoming). This latter point also relates to the so-called marginal emphatics, with a list consisting of [l^ʕ n^ʕ m^ʕ f^ʕ b^ʕ x^ʕ k^ʕ] (Davis 2009), but since these are often attested in restricted environments, next to other emphatics or to a low vowel, they are arguably not part of the phonemic inventory of most dialects (Youssef 2013, p. 102). If, however, they show contrastive behavior in a dialect, they can be analyzed as having a V-place [dorsal] feature.

Table 6. Representational typology of *r phoneme reflexes.

	C-place [coronal]	C-place [dorsal]	V-place [dorsal]	C-manner [open]	V-manner [closed]	... [voice]	Geographical Distribution
/r/	✓			✓	✓		-Peripheral and some Gulf -Maghrebi and Egyptian: vs. /r ^ʕ / -Mesopotamian qəltu: vs. /ʁ/
/R/	✓	✓		✓	✓		Mesopotamian gilit and most Peninsular dialects
/r ^ʕ /	✓		✓	✓	✓		-Most Levantine dialects -Maghrebi and Egyptian (vs. /r/)
/ʁ/		✓		✓		✓	Mesopotamian qəltu (vs. /r/)

3.7. The Pharyngeal Typology

The voiceless and voiced pharyngeals /ħ ʕ/ have largely been preserved in the modern dialects; however, a weakening of one or both phonemes can be observed in a few outskirts of the Arabic sprachbund (Watson 2002, p. 18). According to Fischer and Jastrow (1980, p. 52), Chadian and Nigerian Arabic have reduced old /ħ ʕ/ to laryngeal /h ʔ/, whereas Tihāma (Yemen) and Şıgo-Sason (Anatolia) dialects have only turned /ʕ/ into /ʔ/.

For /ħ ʕ/, there is no phonological evidence to support a C-place [dor] specification (nor any other place feature). I propose the double C-manner features [closed] and [open], a specification that ties phonetically with the considerable variation in the production of pharyngeals, which have been described as having fricative, approximant, or stop gestures (see McCarthy 1994; Shosted et al. 2017).

In Section 3.1, we assigned a single C-manner [closed] feature to the glottal stop /ʔ/, which surfaced as a reflex of *q in certain other dialects. For the natural class of fricatives, I proposed C-manner [open]. Now, let us posit that /h/ is the (placeless)

segment composed entirely of that feature, considering its tendency to delete in word-final position in modern dialects. Table 7 summarizes the feature representation of these consonants and illustrates that the sound changes /ʕ/ > /ʔ/ and /ħ/ > /h/ involve a simple feature deletion mechanism.

Table 7. Representational typology of *ħ-ʕ phoneme reflexes.

	C-manner [closed]	C-manner [open]	... [voice]	Geographical Distribution
/ħ ʕ/	✓	✓	(✓)	All except a few peripheral dialects
/h/ < *ħ		✓		Peripheral dialects: Chadian, Nigerian ...
/ʔ/ < *ʕ	✓			Same as /h/; plus a few extra

4. Discussion and Conclusions

An important characteristic of the property-driven approach is that it refrains from classifying languages, or for that matter dialects, into types. The latter methodology leads to three false implications, elaborated in Hyman (2018, pp. 10–12), which I will consider in relation to the typologies of Arabic consonant reflexes.

The first is that the resulting categories appear to be mutually exclusive. A good illustration of this is the customary classification of Arabic dialects based on reflexes of *k into /k/ vs. /tʃ/ type dialects. As we saw in Section 3.3, the pure /tʃ/ dialects are relatively few, and many more dialects in fact contrast /tʃ/ and /k/ phonemes. Additionally, with increasing pressure to normalize educated speech toward SA, this phonemic split is expanding or even disappearing in favor of /k/. A division of this sort, therefore, appears simplistic.

A second argument is that the outcomes of such studies pretend to offer unique taxonomies, as if “something has been accomplished” (Hyman 2018, p. 11), while in fact multiple categorizations are often possible. Take the case of stop /t d/ vs. fricative /s z/ reflexes of the interdental in Section 3.4. One typologist may classify, say, the sedentary dialects of Egypt in the /t d/ group (e.g., Fischer and Jastrow 1980) when considering well-established lexical items; another may classify them as /s z/-type dialects (e.g., Embarki 2008) given their rendering of recent SA borrowings into fricatives, never stops.

Another example is the Mesopotamian *qəltu* dialects in the *r typology, which are classified under a separate category as a result of their unique /ʁ/ reflex. However, synchrony alone dictates that they should be part of the ‘plain r’ group since they have a single rhotic phoneme /r/ in loanwords, and since the fricative /ʁ/ reflex of *r has totally merged with an existing phoneme, the etymological *ḡayn*. Rarely are the facts so uncomplicated that we can place a dialect in one or the other category. What really matters in the current approach is that the two categories are structurally delineated so that the phonological behavior of those reflexes can be explained, regardless of which dialect falls under which type.

The final argument advanced by Hyman is that the typological labels are often imprecise and invariably run into exceptions. Let us take, for example, the labels proposed by Youssef (2019) for the *r typology. The so-called ‘split-r dialects’ represent a type that contrasts plain /r/ and emphatic /rˤ/, but the label may equally apply to the ‘uvular-r dialects’, which also happen to split the etymological r into two phonemes, /ʁ/ and /r/. The third type, labeled ‘plain-r dialects’, have a non-emphatic rhotic phoneme which is either doubly marked for C-place [coronal] and [dorsal], /R/, or just [coronal], namely /r/. That is why it is more accurate to divide them into two discrete groups, as I have done

in Section 3.6. Another case of inaccurate labeling is the designation of a ‘type’ with an approximant reflex of *ǧ, i.e., /j/ (cf. Watson 2011a, p. 863), even though [j] is typically a conditioned variant of the /dʒ/ phoneme in such dialects (Section 3.2). We must admit, of course, that labels are useful descriptive tools that help us conceptualize the object of our research; the important thing is that they should not be perceived as explanatory typological devices instead. They are not themselves manifestations of actual phenomena.

So, if typology is not about finding types, what should its goal be? By discarding types and embracing the property-driven view, we were indeed able to make valuable predictions, both empirical (for Arabic) and theoretical. Let us review them one by one.

First, as Hyman (2007, p. 265) states, this approach makes no clear distinction between phonological typology and phonological theory; and in doing so, it affords a range of theoretically informed schemes to typologize and explain variation. The current study appealed to the formal apparatus of phonological representation to account for variation in Arabic consonant phonemes. Here, the raw material of the typological analysis is not the phoneme reflexes per se, but how these reflexes are differentiated by the feature hierarchy (cf. Drescher et al. 2018). This contrastive-feature typology then has explanatory power in that the featural makeup of the various reflexes will correlate with their distinct phonological patterning across varieties of Arabic.

Secondly, the feature typology was generated by a specific model of feature geometry, the PSM. By utilizing a handful of features, which are only operative when distinctive (contrastive and/or active), the PSM provides a minimalist device to account for phonological alternations across languages and dialects. I have illustrated that the PSM is not only sufficient to capture complex typological correlations, but also that the correlations are made transparent by the architectural properties of the model. Crucially, feature economy is maximized and phoneme distribution is accounted for.

A relevant case here is that (the mostly Bedouin) dialects with a /g/ reflex of *q (Section 3.1) are more likely to have a /dʒ/ reflex of *ǧ (Section 3.2). By activating C-place [dorsal] for /g/ and V-place [coronal] for /dʒ/, other features being equal, those dialects make maximal use of the few available distinctive features to express their phoneme inventories (cf. Clements 2003). At the same time, they escape creating a common reflex for the two historical consonants, which would result in a merger (mergers happen typically when the phoneme contrasts have a low functional load, which is not the case here).

Thirdly, although exclusively synchronic in essence, the PSM analysis also sheds light on processes of sound change and phonologization, by offering linguistic explanations for how such processes might have taken place. According to Kiparsky (2008), structural properties (including features), rather than systems of opposition, should form the basis for language change. Typological generalizations then simply follow from recurrent patterns of change. As such, historical changes can provide explanations for closely related dialects, but how is this achieved?

We know, for instance, that partial sound change can eventually lead to a phonemic split. This occurs for several of the consonants under scrutiny where multiple reflexes cooccur in a given group of dialects, e.g., /g/ and /q/ reflexes of *q (Section 3.1), /k/ and /tʃ/ reflexes of *k (Section 3.3), and /ʁ/ and /r/ reflexes of *r (Section 3.6). When there is a single systematic reflex, we have an indication that the change is complete. Additionally, since the reflexes are characterized by minimal feature distinctions, we can often register that phonological change involves the addition or deletion of a few features. Finally, conditioned phonetic variants—as I have pointed out for *q, *ǧ, and *k—can provide clues for the process of phonologization. For a principally diachronic perspective of variation in a range of Arabic consonants, readers may consult Embarki (2008, 2014).

In conclusion, contrastive-feature taxonomies provide an interesting insight into the relations that exist between varieties of the same language, both synchronically and diachronically (cf. Drescher et al. 2018). Having demonstrated that the PSM is well suited to capturing variation in the consonants of genealogically related Arabic dialects, we can also claim, following Kiparsky (2018), that typological generalizations are inevitably theory

dependent. The variety of available theoretical solutions should open up new avenues for dialect categorization, independent of traditional classificatory systems that conflate multiple extra-linguistic factors.

Funding: This research received no external funding.

Acknowledgments: For their thoughtful comments and suggestions, I thank the anonymous reviewers and the audience of the “Semitic Dialectology: Crises and Change” conference, organized by Heidelberg University. All remaining errors are my own.

Conflicts of Interest: The author declares no conflict of interest.

Notes

- ¹ A reviewer has pointed out that the picture of laryngeal distinctions is more intricate. Within the PSM, Morén (2003) proposed the features [closed], [open], and [lax] for the Laryngeal tier; and recently Watson and Heselwood (2016) have advocated Laryngeal [open/closed] to analyze languages (e.g., Mehri) or dialects (e.g., San ‘āni Arabic) in which [voice] does not play an active role in the phonology. With this in mind, I acknowledge that the use of [voice] in this paper is a simplification to help streamline the typological analysis, given that the active laryngeal features may vary across Arabic dialects.
- ² The likely older, palatal plosive pronunciation [j] can be found sporadically in the Arabian Peninsula and in Upper Egypt and Sudan (Fischer and Jastrow 1980, p. 51).
- ³ In addition, all dialects with unconditioned and conditioned [tʃ] use this variant consistently in the second person feminine suffix. In certain dialects of Oman and Yemen, [tʃ] is found only in this pronoun, with other /k/’s remaining intact (see Watson 1992). In northern Yemeni highlands, the form has [ʃ] instead (as in *kaškaša*).
- ⁴ Parenthesized [voice] features in this table, as well as in Table 7, belong exclusively to the second member of the consonant pair to the left, here /ð d z v/.
- ⁵ Bellem (2014) mentions Sunni Baḥraini as an exception, being dyadic while retaining the interdental /θ ð/.
- ⁶ In some Maghrebi dialects, the merge to /ð^ʕ/ is inconsistent; one finds voiceless /t^ʕ/ in free variation with /d^ʕ/ in certain instances: [t^ʕr^ʕab]~[d^ʕr^ʕab] ‘he struck’; [mø:t^ʕāʕ]~[mø:d^ʕāʕ] ‘place’ (Djidjelli; Fischer and Jastrow 1980, p. 50).
- ⁷ Note that, unlike the obstruent emphatics, /r^ʕ/ tends to lose its emphatic feature—and surface as plain [r]—next to front vowels. This distribution, known as /r^ʕ/ de-emphasis, leads to root-internal allomorphy between emphatic [r^ʕ] and plain [r]. Youssef (2019) analyzes de-emphasis as contextual neutralization (Kenstowicz and Kisseberth 1979, p. 211), whereby the underlying /r/~/r^ʕ/ contrast is merged in front vowel contexts but maintained elsewhere.

References

- Al-Ani, Salman H. 1970. *Arabic Phonology*. The Hague: Mouton De Gruyter. [CrossRef]
- Al-Azraqi, Munira. 2010. The ancient Dād in southwest Saudi Arabia. *Arabica* 57: 57–67. [CrossRef]
- Al-Nassir, A. A. 1993. *Sibawayh the Phonologist: A Critical Study of the Phonetic and Phonological Theory of Sibawayh as Presented in His Treatise Al-Kitab*. London: Kegan Paul International.
- Anderson, John, and Colin J. Ewen. 1987. *Principles of Dependency Phonology*. Cambridge: Cambridge University Press. [CrossRef]
- Bahloul, Maher. 2007. Linguistic diversity: The Qaaf across Arabic dialects. In *Perspectives on Arabic linguistics XIX*. Edited by Elabbas Benmamoun. Amsterdam: John Benjamins, pp. 247–65. [CrossRef]
- Behnstedt, Peter, and Manfred Woidich. 1985. *Die ägyptisch-arabischen Dialekte. II. Dialektatlas von Ägypten*. Wiesbaden: L. Reichert.
- Bellem, Alex. 2014. Triads, emphatics and interdental in Arabic sound system typology. In *Arabic Linguistics: Traditional and New Theoretical Approaches*. Journal of Semitic Studies Supplement 34. Edited by Manuela Giofio. Oxford: Oxford University Press, pp. 9–41.
- Blanc, Haim. 1964. *Communal Dialects in Baghdad*. Harvard Middle Eastern Monographs 10. Cambridge: Harvard University Press.
- Broselow, Ellen. 1992. Parametric variation in Arabic dialect phonology. In *Perspectives on Arabic linguistics IV*. Edited by Ellen Broselow, Mushira Eid and John McCarthy. Amsterdam: John Benjamins, pp. 7–46. [CrossRef]
- Cantineau, Jean. 1960. *Cours de Phonétique Arabe*. Paris: Librairie C. Klincksieck.
- Clements, George N. 1985. The geometry of phonological features. *Phonology* 2: 225–52. [CrossRef]
- Clements, George N. 1991. Place of articulation in consonants and vowels: A unified theory. *Working Papers of the Cornell Phonetics Laboratory* 5: 77–123.
- Clements, George N. 2001. Representational economy in constraint-based phonology. In *Distinctive Feature Theory*. Edited by Tracy Alan Hall. Berlin: Mouton De Gruyter, pp. 71–146. [CrossRef]
- Clements, George N. 2003. Feature economy in sound systems. *Phonology* 20: 287–333. [CrossRef]
- Clements, George N., and Elizabeth Hume. 1996. The internal organization of segments. In *The Handbook of Phonological Theory*. Edited by John A. Goldsmith. Cambridge: Blackwell, pp. 245–306.

- Corriente, Frederico. 1978. D-L doublets in Classical Arabic as evidence of the process of de-lateralisation of *dād* and development of its standard reflex. *Journal of Semitic Studies* 23: 50–55. [CrossRef]
- Davis, Stuart. 1995. Emphasis spread and grounded phonology. *Linguistic Inquiry* 26: 465–98.
- Davis, Stuart. 2009. Velarization. In *The Encyclopedia of the Arabic Language*. Edited by Kees Versteegh. Leiden: Brill, vol. 4, pp. 636–38.
- Dresher, Elan B. 2009. *The Contrastive Hierarchy in Phonology*. Cambridge Studies in Linguistics 121. Cambridge: Cambridge University Press.
- Dresher, Elan B., Christopher Harvey, and Will Oxford. 2018. Contrastive feature hierarchies as a new lens on typology. In *Phonological Typology*. Phonology and Phonetics 23. Edited by Larry M. Hyman and Frans Plank. Berlin: Mouton De Gruyter, pp. 273–311.
- Edzard, Lutz. 2009. Qāf. In *The Encyclopedia of the Arabic Language*. Edited by Kees Versteegh. Leiden: Brill, vol. 4, pp. 1–3.
- Embarki, Mohamed. 2008. Les dialectes arabes modernes: État et nouvelles perspectives pour la classification géo-sociologique. *Arabica* 55: 583–604. [CrossRef]
- Embarki, Mohamed. 2014. Evolution et conservatisme phonétiques dans le domaine arabe. *Diachronica* 31: 506–34. [CrossRef]
- Erwin, Wallace M. 2004. *A Short Reference Grammar of Iraqi Arabic*. Georgetown Classics in Arabic Languages and Linguistics series; Washington, DC: Georgetown University Press.
- Farwaneh, Samira. 2009. Toward a typology of Arabic dialects: The role of final consonantality. *Journal of Arabic and Islamic Studies* 9: 82–109. [CrossRef]
- Fischer, Wolfdietrich, and Otto Jastrow. 1980. Phonologie und Morphologie des Neuarabischen. In *Handbuch der arabischen Dialekte*. Edited by Wolfdietrich Fischer and Otto Jastrow. Wiesbaden: Harrassowitz, pp. 49–101.
- Gordon, Matthew K. 2016. *Phonological Typology*. Oxford Surveys in Phonology and Phonetics. Oxford: Oxford University Press.
- Haddad, Ghassan. 1984. Problems and Issues in the Phonology of Lebanese Arabic. Ph.D. dissertation, University of Illinois, Urbana, IL, USA.
- Hall, Daniel Currie. 2007. The Role and Representation of Contrast in Phonological Theory. Ph.D. dissertation, University of Toronto, Toronto, ON, Canada.
- Hamdan, Jihad, and Rose Fowler Al-Hawamdeh. 2020. The Arabic /d^ɟ/ revisited: A critical review. *KEMANUSIAAN: The Asian Journal of Humanities* 27: 17–38. [CrossRef]
- Harris, John, and Geoff Lindsey. 1995. The elements of phonological representation. In *Frontiers of Phonology: Atoms, Structures, Derivations*. Edited by Jacques Durand and Francis Katamba. London: Routledge, pp. 34–79.
- Herzallah, Rukayyah. 1990. Aspects of Palestinian Arabic Phonology: A Non-Linear Approach. Ph.D. dissertation, Cornell University, Ithaca, NY, USA.
- Holes, Clive. 1995. *Modern Arabic: Structures, Functions, and Varieties*. Georgetown Classics in Arabic Languages and Linguistics. Washington, DC: Georgetown University Press.
- Hyman, Larry M. 2007. Where's phonology in typology? *Linguistic Typology* 11: 265–71. [CrossRef]
- Hyman, Larry M. 2018. What is phonological typology? In *Phonological Typology*. Phonology and Phonetics 23. Edited by Larry M. Hyman and Frans Plank. Berlin: Mouton De Gruyter, pp. 1–20.
- Jastrow, Otto. 1978. *Die Mesopotamisch-Arabischen Qāltu-Dialekte*. Band I: Phonologie und Morphologie. Wiesbaden: Steiner.
- Johnstone, T. M. 1967. *Eastern Arabian Dialect Studies*. London Oriental Series 17; London: Oxford University Press.
- Kaye, Alan S., and Judith Rosenhouse. 1997. Arabic dialects and Maltese. In *The Semitic Languages*. Edited by Robert Hetzron. London: Routledge, pp. 263–311.
- Kenstowicz, Michael, and Charles Kisseberth. 1979. *Generative Phonology: Description and Theory*. San Diego: Academic Press. [CrossRef]
- Kiparsky, Paul. 2000. Opacity and cyclicity. *The Linguistic Review* 17: 351–66. [CrossRef]
- Kiparsky, Paul. 2003. Syllables and moras in Arabic. In *The Syllable in Optimality Theory*. Edited by Caroline Féry and Ruben van de Vijver. Cambridge: Cambridge University Press, pp. 147–82. [CrossRef]
- Kiparsky, Paul. 2008. Universals constrain change; change results in typological generalizations. In *Linguistic Universals and Language Change*. Edited by Jeff Good. Oxford: Oxford University Press, pp. 23–53.
- Kiparsky, Paul. 2018. Formal and empirical issues in phonological typology. In *Phonological Typology*. Phonology and Phonetics 23. Edited by Larry M. Hyman and Frans Plank. Berlin: Mouton De Gruyter, pp. 54–106.
- Landberg, Carlo de. 1905–1913. *Études sur Les Dialectes de l'Arabie Méridionale*. *Datīna*. Leiden: E.J. Brill, vol. 2.
- McCarthy, John J. 1988. Feature geometry and dependency: A review. *Phonetica* 45: 84–108. [CrossRef]
- McCarthy, John J. 1994. The phonetics and phonology of Semitic pharyngeals. In *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*. Edited by Patricia A. Keating. Cambridge: Cambridge University Press, pp. 191–233. [CrossRef]
- Morén, Bruce. 2003. The Parallel Structures Model of feature geometry. *Working Papers of the Cornell Phonetics Laboratory* 15: 194–270.
- Morén, Bruce. 2006. Consonant-vowel interactions in Serbian: Features, representation and constraint interaction. *Lingua* 116: 1198–244. [CrossRef]
- Mustafawi, Eiman. 2017. Arabic phonology. In *The Routledge Handbook of Arabic Linguistics*. Edited by Elabbas Benmamoun and Reem Bassiouney. London: Routledge, pp. 11–31. [CrossRef]
- Palva, Heikki. 2006. Dialects: Classification. In *The Encyclopedia of the Arabic Language*. Edited by Kees Versteegh. Leiden: Brill, vol. 1, pp. 604–13.
- Roman, André. 1981. De la langue arabe comme un modèle général de la formation des langues sémitiques et de leur évolution. *Arabica* 28: 127–61. [CrossRef]

- Sagey, Elizabeth. 1986. *The Representation of Features and Relations in Nonlinear Phonology*. Ph.D. dissertation, MIT, Cambridge, MA, USA.
- Shosted, Ryan K., Maojing Fu, and Zainab Hermes. 2017. Arabic pharyngeal and emphatic consonants. In *The Routledge Handbook of Arabic Linguistics*. Edited by Elabbas Benmamoun and Reem Bassiouney. London: Routledge, pp. 48–61. [[CrossRef](#)]
- Versteegh, Kees. 2006. Dād. In *The Encyclopedia of the Arabic Language*. Edited by Kees Versteegh. Leiden: Brill, vol. 1, pp. 544–45.
- Watson, Janet C. E. 1992. Kashkasha with reference to modern Yemeni dialects. *Zeitschrift für Arabische Linguistik* 24: 60–81.
- Watson, Janet C. E. 2002. *The Phonology and Morphology of Arabic*. *The Phonology of the World's Languages*. Oxford: Oxford University Press.
- Watson, Janet C. E. 2011a. Arabic dialects (general article). In *The Semitic Languages: An International Handbook*. Edited by Stephan Weninger, Geoffrey Khan, Michael Streck and Janet Watson. Berlin: Mouton De Gruyter, pp. 851–96. [[CrossRef](#)]
- Watson, Janet C. E. 2011b. Word stress in Arabic. In *The Blackwell Companion to Phonology*. Edited by Marc Oostendorp, Colin Ewen, Elizabeth Hume and Keren Rice. Oxford: Blackwell, vol. 5, pp. 2990–3019. [[CrossRef](#)]
- Watson, Janet C. E., and Barry Heselwood. 2016. Phonation and glottal states in Modern South Arabian and San'ani Arabic. In *Perspectives on Arabic linguistics XXVIII*. Edited by Youssef A. Haddad and Eric Potsdam. Amsterdam: John Benjamins, pp. 3–36. [[CrossRef](#)]
- Watson, Janet C. E., and James Dickins. 1999. The representation of palatal vocoids and palatalization of k in dialects of Arabic. *Roehampton Institute London Working Papers in Linguistics* 1: 97–124.
- Wiese, Richard. 2001. The phonology of /r/. In *Distinctive Feature Theory*. Edited by Tracy Alan Hall. Berlin: Mouton De Gruyter, pp. 335–68. [[CrossRef](#)]
- Younes, Munther A. 1993. Emphasis spread in two Arabic dialects. In *Perspectives on Arabic Linguistics V*. Edited by Mushira Eid and Clive Holes. Amsterdam: John Benjamins, pp. 119–45. [[CrossRef](#)]
- Younes, Munther A. 1994. On emphasis and /r/ in Arabic. In *Perspectives on Arabic Linguistics VI*. Edited by Mushira Eid, Vicente Cantarino and Keith Walters. Amsterdam: John Benjamins, pp. 215–35. [[CrossRef](#)]
- Youssef, Islam. 2006. *Vocalic Feature Assimilation in Cairene Arabic and Buchan Scots*. Master's thesis, University of Tromsø, Tromsø, Norway.
- Youssef, Islam. 2013. *Place Assimilation in Arabic: Contrasts, Features, and Constraints*. Ph.D. dissertation, University of Tromsø, Tromsø, Norway.
- Youssef, Islam. 2014. Affrication in Baghdadi Arabic: Synchrony and diachrony. *Arabica* 61: 746–57. [[CrossRef](#)]
- Youssef, Islam. 2015. Vocalic labialization in Baghdadi Arabic: Representation and computation. *Lingua* 160: 74–90. [[CrossRef](#)]
- Youssef, Islam. 2019. The phonology and micro-typology of Arabic R. *Glossa: A Journal of General Linguistics* 4: 131. [[CrossRef](#)]
- Youssef, Islam. forthcoming. Plain-emphatic R phonemes in Arabic: More evidence for phonetics-phonology mismatches. In *Perspectives on Arabic Linguistics XXXIII*. Edited by Abdel-Khalig Ali and Atiqah Hachimi. Amsterdam: John Benjamins.
- Zaborski, Andrzej. 2007. Jim. In *The Encyclopedia of the Arabic Language*. Edited by Kees Versteegh. Leiden: Brill, vol. 2, pp. 494–96.

Article

Definiteness Systems and Dialect Classification

Mike Turner

Department of World Languages & Cultures, The University of North Carolina Wilmington, Wilmington, NC 28403, USA; turnerm1@uncw.edu

Abstract: In this article I explore how typological approaches can be used to construct novel classification schemes for Arabic dialects, taking the example of definiteness as a case study. Definiteness in Arabic has traditionally been envisioned as an essentially binary system, wherein definite substantives are marked with a reflex of the article *al-* and indefinite ones are not. Recent work has complicated this model, framing definiteness instead as a continuum along which speakers can locate referents using a broader range of morphological and syntactic strategies, including not only the article *al-*, but also reflexes of the demonstrative series and a diverse set of ‘indefinite-specific’ articles found throughout the spoken dialects. I argue that it is possible to describe these strategies with even more precision by modeling them within cross-linguistic frameworks for semantic typology, among them a model known as the ‘Reference Hierarchy,’ which I adopt here. This modeling process allows for classification of dialects not by the presence of shared forms, but rather by parallel typological configurations, even if the forms within them are disparate.

Keywords: definiteness; indefiniteness; specificity; referentiality; determination; article systems

Citation: Turner, Mike. 2021. Definiteness Systems and Dialect Classification. *Languages* 6: 128. <https://doi.org/10.3390/languages6030128>

Academic Editors: Simone Bettiga and Roberta Morano

Received: 13 June 2021
Accepted: 22 July 2021
Published: 28 July 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

To date, most efforts at classifying Arabic dialects have been concerned with grouping dialects on the basis of shared forms. At times, these forms have been phonological, such as the reflexes of **q* that inform the well-known sedentary–bedouin division; at others, they have been morphological, such as the 1SG imperfective prefix *n-* that differentiates western from eastern varieties (see [Palva 2006](#) on these, among others). In this paper I put forth an alternate proposal: that it may be beneficial to look past forms themselves, and add to our toolset the use of semantic typology as a metric for grouping and subgrouping dialects. In doing so, the possibility arises that formally dissimilar features in two or more varieties may actually have more in common than previously thought, at least to the extent that the features in question exhibit the same types of polysemy. This approach is not exclusive of existing classification schemes. Instead, it may be seen as a way to further test and refine previous characterizations, or otherwise break a tie when a classification decision is questionable.

Although the typological approach itself can theoretically be applied to any number of interrelated feature sets, I opt to focus here on the interplay between nominal morphosyntax and a set of semantic notions that I refer to with the umbrella term ‘definiteness’. The choice to use the term holistically follows that of other works, including [Lyons \(1999\)](#), similarly titled *Definiteness*, and presumes [Chafe’s \(1976, p. 39\)](#) definition of the same as “whether I think you already know and can identify the particular referent I have in mind”. Nonetheless, to be clear, in speaking of ‘definiteness systems’, my focus is on a particular range of definite-indefinite meanings, including relevant subcategories, that can accompany common nouns in response to Chafe’s question (whether or not the answer is affirmative). Definiteness is a useful feature set with which to test a typological classification approach for various reasons, among them that (1) it can be modeled with a reasonable degree of precision, (2) Arabic dialects are known to differ in the ways they express it, and (3) sufficient material exists such as to be able to model discrete dialects and compare them, at least on a preliminary basis.

Most discussion of definiteness in the Arabic dialectological literature is, as is often true of other features, primarily concerned with formal representations. These discussions can be subdivided into two primary types, the first being the shape and assimilation patterns of the so-called “definite article” **al-*¹, and the second being the presence and shape of “indefinite articles” in dialects that exhibit them. In the case of the former, the article **al-* typically receives little explicit semantic discussion, as it is usually presumed to indicate true definiteness. Indefinite articles have fared somewhat better, perhaps because they clearly depart from formal expectations imparted by the standard language, and differ within dialects themselves; Mion (2009) provides an excellent survey of these articles, and even provides a preliminary (form-focused) typology, though his paper stops short of placing them into a comparative semantic framework.

The organization of the present paper is as follows: I begin with a theoretical discussion of definiteness and models that can be used to envision it, especially as they apply to the Arabic case. Following that, and in keeping with the overall focus on meaning over form, I provide a tier-by-tier view of the primary semantic categories attested in the above models, providing evidence of variation in Arabic by drawing on material from the dialectological literature. The next section provides more complete models of a sample of discrete Arabic dialects, selected again to exhibit the extent of possible variation, and to allow for side-by-side comparison. Finally, I return to questions of dialect classification, including both how we can construct schemes from the present data and how these schemes might interact with classification proposals previously made.

Because linguistic examples are drawn from various sources, many of which exhibit different conventions, I have adapted them (with the exception of Nubi) into a single transcription system and provided my own interlinear glosses and free translations.² In addition, throughout this paper I follow Dryer’s (2014, p. e234) in adopting an intentionally broad and more semantically oriented definition of the term ‘article,’ which is used interchangeably with ‘marker’ to refer to any morphosyntactic structure that adds referential meaning to a noun. As such, its use here should not be understood as a syntactic judgment of any particular form.

2. Modeling Definiteness

As a starting principle, definiteness (in the holistic sense) is presumed here to be a semantic property of nouns in all human languages, stemming from shared cognitive perceptions of the world, entities within it, and other humans’ knowledge of them. This semantic view is distinct from the grammatical expression of definiteness, which may be realized differently (or not at all) on a language-by-language basis. Dryer (2005a, 2005b) respective overviews of definite and indefinite articles for the *World Atlas of Language Structures* (WALS) underscore this point, showing that common cross-linguistic definiteness systems include formal representation of (1) both definiteness and indefiniteness, (2) definiteness but not indefiniteness, (3) indefiniteness but not definiteness, and (4) neither definiteness nor indefiniteness. Despite the variability of possible arrangements, maps of the same data show that they are not distributed at random, but rather display areal characteristics, often bridging disparate language families that are geographically proximate, but then varying inside a single language family that is geographically distributed. As Arabic falls into the latter category, that it sees variability in the expression of definiteness is a reasonable initial assumption.

Although grammarians often speak of “definiteness and indefiniteness” in binary terms, scholars have nonetheless recognized that definiteness and its expression cannot adequately be envisioned on a bipartite basis. In the past half-century, various models have been offered as visualizations of the cognitive statuses that underlie nominal referentiality, a common component of which has been the subdivision of either the ‘definite’ or ‘indefinite’ categories—often both—into more precise subcategories. These models have also generally recognized the same ordering of categories, which form a sort of continuum along

which formal representations might be distributed. Here I briefly review some of these models and select one for the present task, then move more explicitly into the Arabic case.

2.1. The Wheel Model

Givón (1978, p. 298) proposes a wheel-shaped model that distinguishes six possible nominal statuses, which he identifies as (a) ‘referential definite’, (b) ‘referential indefinite’, (c) ‘referential nondefinite’, (d) ‘nonreferential object’, (e) ‘generic predicate’, and (f) ‘generic subject’, with the first and last categories bordering each other. Figure 1 shows this model as he envisioned it for standard English. The choice of a wheel is motivated by Givón’s observation that, while languages often use a single morphosyntactic strategy (possibly including zero-marking) for two or more statuses at once, their distribution across categories is nearly always contiguous. One notes, for example, that the English ‘indefinite article’ *a* (or *an*) can indicate multiple underlying semantic statuses. Givón’s terms are somewhat clumsy—it not immediately apparent how one would contrast ‘indefinite’ and ‘nondefinite’ without reviewing examples—but they do establish the basic principle of multiple semantic distinctions underlying a single form. He also rightly indicates that plural and singular forms do not have to follow the same patterning, and uniquely carves out space in his model for generic entities.³

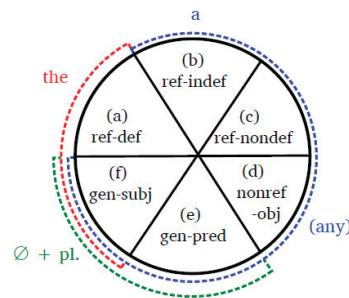


Figure 1. Givón’s Wheel Model, for English (redrawn).

2.2. The Givenness Hierarchy

Gundel et al. (1993) approach the same issue more broadly, framing definiteness as a subcomponent of a larger set of meanings, including those indicated by personal and demonstrative pronouns, that they refer to as ‘givenness’. They propose a ‘Givenness Hierarchy’ (Table 1) consisting of six cognitive statuses, wherein the more discursively ‘known’ or ‘given’ a referent it is, the further to the left of the hierarchy it will be. The three rightmost statuses in the Givenness Hierarchy might be seen as corresponding with the four statuses (a)–(d) of Givón’s Wheel Model, showing a discrepancy in the choice of subdivision despite a general agreement that subdivisions should exist. One contribution of Gundel, Hedberg, and Zacharski is that they provide a formal representation of one of the ‘indefinite’ subcategories by giving informal English *this* as an indefinite article, a use that is further confirmed in Ionin (2006), who calls it a ‘specific’ marker. As it is useful to be able to provide semantically nuanced free translations, I make ample use of indefinite *this* in translations of Arabic examples in this paper.

Table 1. The Givenness Hierarchy (Gundel et al. 1993).

in Focus	>	Activated	>	Familiar	>	Uniquely Identifiable	>	Referential	>	Type Identifiable
<i>it</i>		<i>that</i> <i>this</i> <i>this N</i>		<i>that N</i>		<i>the N</i>		<i>indefinite</i> <i>this N</i>		<i>a N</i>

2.3. The Reference Hierarchy

Drawing together advantages of both the Wheel Model and the Givenness Hierarchy, a more recent proposal by Dryer’s (2014, p. e235) by the name of the ‘Reference Hierarchy’ (Table 2) combines the more limited scope and greater categorical distinctiveness of the former with the hierarchical implications of the latter. Dryer’s model enjoys the unique advantage of having been constructed on the basis of a large corpus of real-world language data, which featured in his (Dryer 2005a, 2005b) work for the WALS database; as such, it is likely to be sufficient for the description of most languages (including Arabic). Like Givón before him, Dryer emphasizes the tendency of articles to be both polysemous and contiguous across a particular range of meanings; meanwhile, like Gundel, Hedberg, & Zacharski, Dryer relies on the notion of a hierarchical relationship whereby nouns that are more ‘known’ or ‘given’ are located further to the left. His choice of five categories is more akin to the Wheel Model, though he leaves out generics and splits ‘referential definites’ into ‘anaphoric definites’ and ‘nonanaphoric definites’. Also like the Wheel Model, the Reference Hierarchy proposes three non-generic indefinite statuses, i.e., one more than the Givenness Hierarchy indicates. Finally, although Dryer’s particular terminologies are lengthy, he does provide a set of 2 to 3-letter abbreviations (in heading of Table 2), which are particularly suitable for in-line reference and interlinear glosses.

Table 2. The Reference Hierarchy (Dryer’s 2014), with abbreviations.

AD	ND	PSI	PNI	SNI				
anaphoric definites	>	nonanaphoric definites	>	pragmatically specific indefinites	>	pragmatically nonspecific (but semantically specific) indefinites	>	semantically nonspecific indefinites

2.4. Applying the Reference Hierarchy

Because it captures the advantages of models before it, was specifically proposed as a response to cross-linguistic data, and allows for abbreviated reference to particular semantic statuses, I opt to use the Reference Hierarchy as the working model for the current paper, and hereby adopt the terms AD, ND, PSI, PNI, and SNI for their respective meanings. These abbreviations are henceforth used liberally in both glosses and prose. It is nonetheless worth pointing out that broad terminological consensus has yet to emerge within this field of inquiry, so I summarize each status as follows, for clarity:

1. Anaphoric definite (AD), which is a subset of both Givón’s ‘referential definite’ and Gundel, Hedberg, & Zacharski’s ‘uniquely identifiable’, refers to the status of a noun that the speaker presumes identifiable to the listener because the referent has already been explicitly introduced or implied in the present discourse. In English it is obligatorily marked with *the*, and optionally with the demonstrative adjectives *this* or *that*.
2. Nonanaphoric definite (ND), which is also a subset of Givón’s ‘referential definite’ and Gundel, Hedberg, & Zacharski’s ‘uniquely identifiable’, refers to the status of a noun that a speaker presumes identifiable to the listener because the referent is available through shared world knowledge. In English it is obligatorily marked with *the*.
3. Pragmatically specific indefinite (PSI), which corresponds with Givón’s ‘referential indefinite’ and is a subset of Gundel, Hedberg, & Zacharski’s ‘referential’, refers to the status of a noun that the speaker can uniquely identify but presumes the listener cannot. It has elsewhere been called ‘specific’, and in English is obligatorily marked with *a(n)*, or in more informal varieties with *this* (Ionin 2006).
4. Pragmatically nonspecific (but semantically specific) indefinite (PNI), which corresponds with Givón’s ‘referential nondefinite’ and is a subset of Gundel, Hedberg, & Zacharski’s ‘referential’, refers to the status of a noun that neither the speaker nor listener can uniquely identify, but which the speaker conceptualizes as being distinct

from others of its type. It has elsewhere been called ‘existential’, and in English is obligatorily marked with *(a)n*, but can also be marked with *some* (Israel 1999).

5. Semantically nonspecific indefinite (SNI), which corresponds with Givón’s ‘nonreferential object’ and Gundel, Hedberg, & Zacharski’s ‘type identifiable’, refers to the status of a noun that is fully unindividuated and is interchangeable with any other of its type. In English it is obligatorily and exclusively marked with *a(n)*.

Using the above definitions, it is possible to build a visual representation of a given language’s definiteness system by representing the Reference Hierarchy as a series of blocks along which corresponding forms can be mapped. Figure 2 gives my interpretation of the system in spoken American English. The articles represented at top, *the* and *a(n)*, are obligatory; meanwhile, the forms at bottom represent auxiliary strategies. This strategy is maintained for other iterations of the model in this paper. The visual model has the added benefit of easing comparison between multiple systems, as is our purpose here, and explored further in Section 4.



Figure 2. Forms represented along the Reference Hierarchy for American English.

2.5. Definiteness in Arabic

A handful of works to date have treated definiteness (or aspects of it) in Arabic specifically. Of these, Brustad (2000, pp. 18–43) is the most immediately relevant in both its focus on spoken Arabic and its comparative approach. She introduces the idea of a ‘definiteness continuum’ that includes not only meanings that are “wholly definite” or “wholly indefinite”, but also exist within an intermediate range that she terms ‘indefinite-specific’. Within the current framework, “wholly” definite and indefinite correspond with the statuses AD/ND and SNI, respectively; meanwhile, the indefinite-specific range that Brustad speaks of seems to cover both PSI and PNI. Looking at Moroccan, Egyptian, Syrian, and Kuwaiti dialects, Brustad identifies common patterns, among them the marking of true definites (AD/ND) with a reflex of **al-*, as well as the zero-marking of non-referential (SNI) nouns. Taken alone as a binary opposition, this initial observation corresponds with the way definiteness in Arabic is often framed.

At the same time, Brustad also establishes the presence of structures that add more nuance than the binary model allows, many of which vary by dialect. Within the indefinite-specific range, she documents use of reflexes of **wāḥid* ‘one’ for all four dialects, observing that it often marks a new topic that is subsequently adopted in the discourse. I qualify such referents as inherently PSI, in that new topics are necessarily known to the speaker—who can therefore expound upon them—but are presumed inaccessible to the listener. Nonetheless, as Brustad notes that **wāḥid* is often restricted to humans (e.g., *wāḥid badwi* ‘a certain bedouin’, p. 20), I am more inclined to read it in such cases as an indefinite pronoun modified by an adjective (i.e., *someone (who is a) bedouin*) rather than a truly inclusive article that can modify any common noun. The exception is in Moroccan, which I discuss more specifically in Section 3.3.

For Moroccan and Syrian varieties, Brustad locates an article *ši*, which she glosses as ‘some (kind of)’ and contends speakers use “to indicate that they have a particular type of entity in mind”. Brustad also raises the possibility of interpreting dialectal *tanwīn* as a sort of indefinite-specific marker, citing Ingham (1994, pp. 47–50) comments on its semantic qualities in Najdi Arabic, and shows how both partitive structures and demonstrative adverbs can have the same semantic effect in Egyptian (Brustad 2000, pp. 30–31). Under the broad definition of ‘article’ used here—which, again, privileges semantic function over syntactic analysis—I consider such structures part of a given dialect’s article system, and specifically include them in below models.

Elsewhere, Brustad complexifies uses of the article **al-*, typically seen to be a marker of true definiteness. Two principal qualifications arise from her data. The first of these is that while true definite (AD and ND) nouns are consistently represented with **al-*, anaphoric definites are often further marked with an unstressed demonstrative adjective (*hād-*, *ha-*, etc.) as a means of increasing their discursive prominence (112–139). I see this common strategy as akin to other auxiliary strategies for marking particular referential meanings, and thus class these as a type of AD marker. The second qualification involves the presence of **al-* in apparently indefinite contexts, which Brustad identifies as a common occurrence in Moroccan (e.g., *xəss-ni l-wəld* ‘I need a son’; p. 36). I interpret this as evidence that the Moroccan reflex of **al-* is distributed over a wider range of referential statuses in general (see Section 4.9).

There are few other holistic studies of definiteness in spoken Arabic. Turner (2018) is comparative and concerned exclusively with spoken Arabic, and employs the same descriptive model as the current paper to explore variability in spoken Arabic; the reader is encouraged to refer to it for additional data presented within the Reference Hierarchy framework. Fassi Fehri (2012, pp. 205–31) provides a more traditional syntactic view of determination in Arabic and Semitic at large, and includes some spoken Arabic data. Remaining studies that have relevance for the study of definiteness in Arabic can be divided into two types. The first are those that focus on single varieties, such as Caubet (1983), Belyayeva (1997), and Fabri (2001) focused and theoretically nuanced descriptions of definiteness in Moroccan Arabic, Palestinian Arabic and Maltese, respectively. The second type of relevant studies are those that examine a single form through a multifunctional semantic lens, and include in turn accounts of its articular functions; among these, Wilmsen’s (2014) expansive account of *šī* across Arabic varieties and Leitner and Procházka’s (Forthcoming) examination of *fard* in the dialects of Iraq and Khuzestan stand out.

3. Points of Variation

Following from Brustad’s observations that structures not traditionally recognized as articles can, on a semantic and pragmatic level, be used to indicate particular referential meanings, a set of metrics for locating these in situ is useful. Even for forms that have been recognized as articles—whether definite or indefinite—in previous literature, a semantics-first view allows us to more specifically delineate the range of meanings that they cover. The aim of this section is, accordingly, to walk through each of the semantic statuses along the Reference Hierarchy, describe how each can be located by discursive context, and identify some points of variation in regard to how each is expressed formally across spoken Arabic varieties.

Because the goal of the section is simply to survey variation, it is more concerned with the fact that a strategy is attested at all than it is with that strategy’s relative frequency of use. Nonetheless, as it is useful for comparative purposes (which follow in Section 4) to establish a baseline measure of how grammaticalized a given strategy is, I do also offer here initial readings of where each falls on a conceptual continuum that ranges between fully ‘obligatory’ and ‘auxiliary’. While obligatory articles are easy to define—they are used by all speakers for all instances of the target meaning—and auxiliary articles can be understood as marked structures that are used by speakers for special emphasis, there is also an intermediate category of markers that are used so frequently with their corresponding meanings such as not to be highly marked, but are still not obligatory in all cases. I refer to these markers as ‘conventionalized’, a placeholder term used with an understanding that truly accurate frequency judgments will require more in-depth semantic study of individual varieties.

3.1. Anaphoric Definites

Anaphoric definites are easily located in extended discourse because they simply involve subsequent reference to an entity that has already been explicitly introduced. In the Hassaniya Arabic sentence given in (1), for example, the narrator introduces a certain

sba' 'lion' as a new referent; when it re-occurs in the text, the referent *sba'* is now necessarily AD, and is accordingly marked with **al-*:

raṣṣaf *'lī-h* *sba'* ... *yraṣṣaf* *'lī-h* *s-sba'* *w-* *yǧūl* *-lu* (1)
 jump.PFV.3MSG upon-3MSG lion.PSI jump.IPFV.3MSG upon-3MSG AD-lion and- say.IPFV.3MSG -3MSG.DAT
 'This lion jumped on him ... the lion is jumping on him and saying to him ... ' (Heath 2003, p. 116)

That the article **al-* is used here as the marker of anaphoric definiteness is not particularly surprising to anyone with knowledge of Arabic, formal or informal, and in most varieties it is indeed the sole obligatory marker of AD nouns. Nonetheless, the point of the example is to highlight contextual expectations. Importantly, when the same sort of discursive context is located elsewhere in the same variety, we find the variation of the type noted by Brustad elsewhere, namely in the auxiliary use of an unstressed demonstrative, as in (2):

w- *Ḍayna* *l-* *fullāni* *-wāḥad* ... *wə-* *dāk* *l-fullāni* *āba* *yarḥal* (2)
 and- come.PFV.1PL to- Fulbe -PSI and- DEM.AD AD-Fulbe refuse.PFV.3MSG leave.IPFV.3MSG
 'We came to this Fulbe man ... the Fulbe man refused to leave' (Heath 2003, p. 78)

While I am not aware of any survey beyond Brustad's, many Arabic varieties exhibit the demonstrative anaphoric reinforcement pattern in one way or another, and because demonstratives themselves vary widely in form, but mirror each other semantically, it is not particularly useful to list off all possible forms here (although see Magidow 2016 for a survey). On a typological level, it is also unsurprising that the demonstrative frequently plays this role, given demonstratives are a frequent source of definite articles in world languages (De Mulder and Carlier 2011). Instead, what is more worthwhile to note in the Arabic case is the degree to which a variety has conventionalized the demonstrative as an AD marker, at which point it might be said be an article of its own. At least some Levantine dialects appear to meet this description, as is evident in the use of *hal-* (etymologically *hā + il-*) in (3), from Baskinta (Lebanon):

'in-na *Ḍār* *ib-* *ḥaḍ-ḍay'a* *b-iḥibb* *an-naʿwim* ... *nāyim* *ḥaz-zalami* (3)
 have-1PL neighbor.PSI in- AD-village IND-sleep.IPFV.3MSG GEN-sleep sleeping.PTCP AD-man
 'We have this neighbor in the village who loves to sleep ... the man was sleeping' (Abu-Haidar 1979, p. 141)

Just how widespread this pattern is in the Levant warrants further study⁴, but for the present purpose it is enough to point out that a dialect that could be shown to *obligatorily* mark AD nouns with a certain structure, but not ND ones, would be typologically distinct from most other varieties at present, and worthy of recognition of such. This phenomenon is also attested in the Nubi Arabic-based creole, wherein a postposed demonstrative reflex *'de* accompanies AD nouns. The major difference is that, in Nubi, the Arabic article **al-* has been lost entirely:

'bas *'uo* *'jowzu* *bī'niya* *'de* (4)
 well 3MSG marry girl AD
 'Well, he married the girl [previously mentioned]' (Wellens 2003, p. 67)

All of these strategies, of course, are overt, and they all incorporate either **al-* or a demonstrative (or a combination of both). The one major exception for AD nouns is the Central Asian cluster of dialects spoken in Uzbekistan (near Bokhara) and northern Afghanistan (near Balkh), which Ingham (2003) has suggested are branches of the same historical group (see also Seeger 2013). These varieties neither have a reflex of **al-* nor have any obligatory compensatory strategy when nouns are AD, as in (5):

fād *mara* *kōnāt* ... *qōlāt* *mara* ... (5)
 PSI woman be.PFV.3FSG say.pfv.3fsg woman.AD
 ‘There was this woman ... the woman said ...’ (Jastrow 2005, p. 138)

That said, even these varieties use demonstratives anaphorically, as in *duk zağīr* ‘the child [previously mentioned]’ (Ingham 2003, p. 33), so they do have at least an auxiliary means of overtly marking AD statuses. In this sense, the Central Asian group shares a typological feature with the larger dialect landscape, even if it is missing a ‘core’ Arabic feature in its lack of **al-*.

3.2. Nonanaphoric Definites

Nonanaphoric definites are uniquely identifiable to both the speaker and listener via world knowledge, and they can be distinguished from AD nouns in extended discourse in that they have not previously been introduced. Common nouns that are ND in most circumstances include ‘the sun,’ ‘the world,’ ‘the country,’ ‘the king,’ and any other for which there is only likely to be one possible interpretation on the part of the listener, despite being new to the discourse; as such, they are relatively easy to locate. This semantic status shows the least variation from dialect to dialect, and is most often represented by **al-* to the exclusion of all other strategies (including demonstrative reinforcement). A typical example is in (6), from the Jazira area of Sudan, where ‘the mayor’ is unique and identifiable as the mayor of the implied town in the narrative despite only being mentioned for the first time:

rawwāḥ *lā-* *l-’umda* *šakā* *-ni* *’alē* ... (6)
 go.PFV.3MSG to- ND-mayor complain.about.PFV.3MSG -1SG.OBJ to.3MSG
 ‘He went to the mayor and complained about me to him’ (Hillelson 1935, p. 48)

The primary exception to this pattern is, predictably, varieties that have lost the article **al-*; in such cases, the ND noun is unmarked. The Afghanistan Arabic utterances in (7), for example, provide first mention of ‘the queen’ with no further modification. Similar unmarked patterns can be identified in Nubi, as in *’hari ta ’shems* ‘the heat of the sun’ (Wellens 2003, p. 67). It is worth noting that these varieties, like others, do not see auxiliary use of demonstratives for ND nouns, even though they allow them for AD nouns.

malika *li-* *zōl* *kasir* *zīn* *kōm* *mi-ššūf* (7)
 queen.ND OBJ- Zal very wonderful be.PFV.3MSG IND-see.IPFV.3FSG
 ‘The queen thought that Zal was very wonderful’ (Ingham 2003, p. 34)

3.3. Pragmatically Specific Indefinites

Pragmatically specific indefinites can be identified in extended discourse as referents that are mentioned for the first time, and not accessible to the listener via world knowledge, but for which the speaker can thereafter be seen to provide specific information. Strategies for marking PSI nouns are the most varied and innovative, particularly if we are to adopt a wide view of what an article is, and many have been under-recognized to date. Most of the “indefinite articles” of the dialectological literature are, in fact, PSI articles, whether exclusively or in a polysemic distribution with the PNI status.

A common source for PSI articles is, as is common in world languages (Heine 1997, pp. 66–83), a numeral **wāḥid* ‘one’ or **fard* ‘one, an individual’. The former of these is best associated with Moroccan and western Algerian varieties, where a reflex of **wāḥid* is typically obligatory for new, pragmatically salient referents of which the speaker has unique knowledge. Unique to this structure, however, is that **wāḥid* accretes with **al-*, yielding a sort of double-marked structure. Caubet (1983, p. 83) gives the Moroccan article as a fused *wāḥad-al*, which is a plausible reading in most cases, but I venture that the article *l-* itself might also be considered a PSI marker, especially as it can be syntactically detached from *wāḥad* but still coincide with a clear PSI meaning, as in (8) from Anjra (Morocco):

hāda wāḥd əl-‘āyəl *u-* l-‘āyla *ma-* *yžəbru* *-ši* *fāyn* *ysəknu* (8)
 EXIST PSI PSI-boy and- PSI-girl NEG- find.IPFV.3PL -NEG REL.LOC live.IPFV.3PL
 ‘There was this boy and this girl who couldn’t find anywhere to live’ (Vicente 2000, p. 221)

The articular use of **wāhid* to mark PSI referents is also attested in eastern varieties of Hassaniya, as spoken in Mali, though here it is suffixed rather than prefixed, and is not obligatory. It has not been explicitly recognized as such, but is regularly apparent in contexts such as (9), recorded in Gao, where further specification of the noun *blad* ‘place’ makes it clear that the speaker has unique knowledge of it. A similar structure is documented in Nubi, e.g., *mas’kin ‘wai* ‘a certain poor man’ (Wellens 2003, p. 64).

dxalna blad wāḥəd *yəngəl* *-lu* *hari-bomo* *ft-l* *s-sbu’a* *yāsriin* (9)
 enter.PFV.1PL country PSI call.PASS.IPFV.3MSG -3MSG.DAT Hari-Bomo in-3MSG SNI-lions many.PL
 ‘We entered this place that’s called Hari-Bomo; there are a lot of lions there’ (Heath 2003, p. 110)

The article **fard*, of similar semantic provenance, is widely recognized in the dialectological literature, where it is most often associated with Mesopotamian varieties. Blanc (1964, 118) locates this article in Baghdad, and describes phonological variants of it associated with particular sectarian groups, but gives limited semantic information, saying “its presence contrasts fairly clearly with that of the article /l/ or other determination marks, but the degree to which it contrasts with absence of any mark is yet to be determined.” Recent work by Leitner and Procházka’s (Forthcoming) significantly expands on the functions of **fard*, showing that it is a polyfunctional lexeme with multiple senses, one of which is to mark a noun that is “new for the hearer and important for the subsequent discourse.” This quintessentially PSI sense for **fard* is attested throughout Iraq and Khuzestan, as in (10), from Basra, where the speaker starts a story by introducing a particular *ṭālib* ‘student’:

fadd *yōm* fadd ṭālib *rāḥ* *li-* *l-madrasa* *mit’axxir* (10)
 PSI day PSI student go.PFV.3MSG to- ND-school late
 ‘One day this student went to school late’ (Denz and Edzard 1966, p. 78)

Mion (2009) locates reflexes of **fard* in other Arabic varieties, too, including those of Mardin and Tunis, but in most of these cases the reflex is less apparently referential and simply implies ‘one, the same’ (though potential for future reanalysis remains). Nonetheless, it is attested with a clear PSI meaning in Central Asian varieties, as in *fad mara* ‘a [certain] woman’ in (5), above.

These are the only structures regularly called ‘articles’ in the literature, to my knowledge, that meet the semantic parameters of PSI, but under the broad definition we can easily expand the field of extant PSI articles. The first sort of novel article is derived from the demonstrative adverb, but has the same pragmatic effect of indicating a referent that is identifiable to the speaker, but not the listener. Brustad offers this interpretation of *kida* in Cairene (*šuft ḥāga kida* ‘I saw this thing . . . ’), a view that is supported by numerous examples in Woidich (2006, p. 236). The same function can also be located elsewhere in Egypt, as in (11), from Bani Swayf:

bi-ni’mil *-laha* ḥuwəyza sḡayyaṛa kida *‘ala* *‘addi-ha* (11)
 IND-do.IPFV.1PL -3FSG.DAT box small.F PSI upon size-3FSG.POSS
 ‘We make this little box for it, in its size’ (Behnstedt and Woidich 1988, p. 16)

Furthermore, there is evidence for a parallel strategy in some Yemeni varieties, which use the demonstrative adverb *hākaḏāha* (and similar; see Watson and ‘Amri 1993, pp. 418–19) to the same effect; in (12), for example, the speaker introduces a *bug’ah* ‘place’ and immediately provides more information, a hallmark of a PSI noun:

ḥna fi'il-na l-'iris ... fi bug'ah hākaḏāhā nisammī -ha mafraj (12)
 1PL do.PFV.1PL AD-wedding in place PSI call.IPFV.1PL 3SG.OBJ mafraj
 'We had the wedding ... in this place that we call a *mafraj*' (Watson and 'Amri 2000, p. 242)

Another structure that qualifies as a PSI marker on the basis of its semantic associations is the so-called 'dialectal *tanwīn*' (DT) of the dialectological literature. Even though the origins of this marker remain an object of debate, its functions are relatively similar across varieties. Stokes (2020, p. 637) summarizes DT as "the morpheme, typically realized as *in* or *an*, that is suffixed to a morphologically indefinite noun, primarily when followed by some type of adnominal adjective or clause". The fact that DT is restricted to indefinites is alone sufficient to establish that it has some relationship with the semantics of referentiality; in addition, that it typically proceeds an adnominal element—which, on a pragmatic level, individuate the noun as distinct from others of its type—calls for a PSI or PNI reading of the resulting phrase. As such, it is not surprising that it can be located with nouns that clearly meet the parameters of a PSI referent, as in (12) from the Jezira (Sudan):

qibēl ḡa fōq-i gana-yan šukri rākib -lu ba'ir-an ḥūri (13)
 earlier come.PFV.3MSG up-1SG boy-PSI Shukri riding.PTCP -3MSG.DAT camel-PSI yellow
 'Earlier this Shukri boy came up to me riding this yellow camel' (Hillelson 1935, p. 60)

While DT can accordingly be read as a sort of PSI article, in most cases it is still syntactically conditioned, in that it depends on the presence of an adnominal attribute (regardless of the speaker's ability to uniquely identify the referent). There is nonetheless evidence that some varieties have moved toward fully semanticizing DT, as in Najdi, for which Ingham (1994, p. 50) gives examples such as *liḡēt bēt-in* 'I've found a [certain] house'. It is also possible to locate varieties in which a reflex of DT (which only occurs in this sense) accretes with another PSI article such as **wāḥid*, as can be seen in (14), from Tillo (Anatolia):

yehkaw baḥs āv ḡt-tattūn ḥakkoyḡt -ən-wāḥde (14)
 tell.IPFV.3PL about DEM AD-tobacco story -PSI-PSI
 'They tell this story about that tobacco' (Lahdo 2009, p. 229)

Finally, it is worth pointing out that in many varieties, underlying PSI referents are simply unmarked. Such nouns have the same underlying semantic properties, but are not overtly marked as such, either because a marker is unavailable or the speaker chooses not to use it. A typical example is in (15), from Baskinta (Lebanon).

'in-na ḡār ib- haḡ-ḡay'a b-iḡibb an-naḡm ... (15)
 have-1PL neighbor.PSI in- AD-village IND-sleep.IPFV.3MSG GEN-sleep
 'We have this neighbor in the village who loves to sleep' (Abu-Haidar 1979, p. 141)

3.4. Pragmatically Nonspecific Indefinites

Pragmatically nonspecific indefinites are neither uniquely identifiable to the speaker nor the listener, but are conceived of by the speaker as being distinct from others of their type in the world at large. Though a speaker of a variety that marks overtly PNI nouns can signal them as such in any desired context, from an observer's perspective this semantic status is most easily located where the speaker speculates about the potential nature of a unique referent not yet located; as such, it is often the object of verbs such as 'find', 'obtain', and 'make'. The most easily identifiable PNI article is *ši*, conventionalized in Levantine (16) and Moroccan (17), and which carries this sense exclusively when used as an article:⁵

kān ya'mil -lu šī mašhra w- išil išwayyit faḥm (16)
 be.PFV.3MSG make.IPFV.3MSG -3MSG.DAT pni kiln and- take.out.IPFV.3MSG bit charcoal
 'He would make himself some sort of kiln and produce a bit of charcoal' (Abu-Haidar 1979, p. 145)
 ma- gal -līya -š smiyt-u, gal -li kāyn šī filam māzyān (17)
 NEG- say.PFV.3MSG -1SG.DAT -NEG name-3MSG.POSS say.PFV.3MSG -1SG.DAT EXIST PNI film good
 'He didn't tell me its name; he told me there's some good film [playing]' (Caubet 1993, p. 338)

The article **fard*, described above as a conventionalized marker of PSI statuses, is also attested with a PNI meaning, making the form itself polysemous, as in (18), from Baghdad. The *bayt* 'house' in question here is semantically specific, but the speaker has not located it yet. Reflexes of **fard* are used comparably in Central Asian varieties, as in *fad- ʾord* 'some place' (Ingham 2003, p. 34).

dā-dawwir 'ala fād bayt l- il-'iḡār āni w- zawīḡt-i (18)
 ASP-search.1PL.IPFV for PNI house for GEN-rent 1SG and- wife-1SG.POSS
 'I'm looking for some house or the other for me and my wife to rent' (McCarthy and Raffouli 1965, p. 17)

Exhibiting similar polysemy, if we are to read it as a type of article, is dialectal *tanwīn*, which can also indicate a PNI meaning. This is evident in (19), from the Jezira (Sudan), where the speaker has no particular *arnab* 'rabbit' in mind, but implies God might:

allāh yaḏḏī -na -lēna arnab -an nit'ašša b-a (19)
 God bring.IPFV.3MSG -1PL.OBJ 1PL.DAT rabbit -PNI have.dinner.IPFV.1PL with-3FSG
 'May God bring us some rabbit that we can have for dinner' (Hillelson 1935, p. 46)

Beyond these articles, I am not aware of any other regularly occurring PNI markers, and most varieties simply leave PNI nouns unmarked, as in (20) from Sanaa. This is not to rule out that partitive-like structures, in particular, might sometimes bridge into this meaning; Sanaani itself does, for example, occasionally use a form *zārat* with plurals or as part of the SNI indefinite pronoun *zārat wāḥid* (Watson and 'Amri 2000, p. 114).

hānāk tilgā maktab illī hum yixarrijū -k bi- s-siyāha (20)
 there find.IPFV.2SG office REL 3PL take.out.ipfv.3pl -2msg.obj for GEN-tourism
 'There you'll find some office or the other that can take you out for tourism' (Watson and 'Amri 2000, p. 26)

3.5. Semantically Nonspecific Indefinites

Semantically nonspecific indefinites are, by definition, interchangeable with any other entity of their type, and cannot be discursively prominent. As such, they are nearly always the object of a verb or preposition and not typically modified. Across Arabic varieties, SNI nouns are most commonly unmarked. The word *ḥbal* 'rope' in the Hassaniya example in (21) is typical:

garraṇna lā-ḥmūr ... kull tlatā fā- ḥbal (21)
 bind.PFV.1PL AD-donkeys every three to- rope
 'We bound the donkeys . . . each three with a rope. (Heath 2003, p. 110)

As a general rule, articles that fulfill the PSI or PNI function are not used to indicate SNI entities, though pragmatic considerations may occasionally let PNI markers bridge into this meaning.⁶ In the case of *tanwīn*, which is both semantically and syntactically conditioned, the fact that SNI nouns are unmodified means there is no syntactic impetus for it to appear with them, and I am not aware of any examples that show it being used alone with any sense other than the PSI one noted in Section 3.3.

The primary exception to the general tendency of Arabic varieties to leave SNI nouns unmarked is, perhaps unexpectedly, in varieties that instead mark them with **al-*, at least in some circumstances. Moroccan is most notable for this, as in (22) and (23), where both the *tūr* 'bull' and *sālḥām* 'cloak' are non-referential, being mentioned only once in passing:

dbəḥ t-tūr, 'rəḍ 'la n-nās (22)
slaughter.PFV.3MSG SNI-bull invite.PFV.3MSG PREP ND-people

'He slaughtered a bull, invited people over ...' (Brustad 2000, p. 37)
lyūm *huwa* *lābəs* s-səlhām (23)
today 3MSG wearing.PTCP SNI-cloak
'Today he is wearing a cloak' (Harrell 1966, p. 190)

That this pattern is attested and permissible is sufficient to call the view of **al-* as a *universal* definite article in Arabic into question.⁷ That said, within Moroccan it is possible to find SNI nouns both with **al-* and with no marking at all. I have elsewhere argued that the marked pattern is more common with type-focused uses of SNI nouns and that the unmarked one is mostly reserved for delineating a specific quantity (Turner 2018, pp. 184–88). It is probably not prudent to call **al-* obligatory in this sense, but it is frequent.

4. Systems in Comparison

Taking the above data into account, it seems fair to say that there are a wide variety of strategies for expressing discrete definiteness values in Arabic dialects. This observation alone has implications for descriptive practice, as being aware of extant diversity within a linguistic group is always helpful in delineating which grammatical categories one should check for in fieldwork and comment on in publications. The greater promise of explicitly collecting such data, that said, is that it opens the door for new comparative approaches. In this section, I provide provisional sketches of the overall arrangement of definiteness systems in a sample of ten Arabic varieties, in addition to the Nubi Arabic-based creole, allowing for side-by-side comparison, before moving into the final discussion of how we might use such characterizations for classification. The rough order of sketches here is from more simplex systems to more complex ones, as I estimate them to be.⁸

4.1. Libyan

Libyan Arabic dialects, including those spoken in the eastern Benghazi area (Elfitoury 1976; Owens 1984) and Tripoli further west (Grand'Henry 2000; Yoda 2005), show a very strict binary division between definite (AD and ND) nouns, marked with *(i)l-*, and indefinite (PSI, PNI, and SNI) nouns, which are invariably unmarked. A review of texts in Grand'Henry (2000) confirms this impression, and I am not able to locate any regular auxiliary strategies. Figure 3 gives the distribution of forms in Libyan.

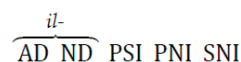


Figure 3. Reference Hierarchy for Libyan.

4.2. Egyptian

Egyptian varieties show the same basic pattern of obligatorily marked definite (AD and ND) nouns, and Brustad (2000, p. 140) specifically notes “the absence of an anaphoric demonstrative article in Egyptian.” Brustad’s data are from Cairo, but texts from Behnstedt and Woidich (1988) show the same patterns elsewhere in Lower Egypt. Although it does not have any obligatory means for indicating indefinite meanings, speakers of Egyptian do have the auxiliary marker *kida* for PSI referents (see Section 3.3). Figure 4 gives the distribution of forms in Egyptian, with the obligatory *il-* represented at top and the auxiliary *kida* at bottom.

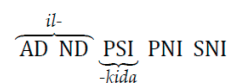


Figure 4. Reference Hierarchy for Egyptian.

4.3. Kuwaiti

Kuwaiti Arabic (Figure 5) also shows the formal distinction between true definites marked with *il-* and unmarked indefinites, but also allows for regular auxiliary marking of AD nouns with an unstressed anaphoric demonstrative *ha-* (Brustad 2000, pp. 120–21), which accretes with the definite article. Brustad does not identify any Kuwaiti structures that would express meanings in her ‘indefinite-specific’ range (i.e., PSI and PNI), and I am likewise unable to locate any in her texts.

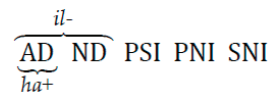


Figure 5. Reference Hierarchy for Kuwaiti.

4.4. Hassaniya

Hassaniya Arabic varieties are found across a wide expanse of western Africa; Cohen (1963) provides a description of the Hassaniya of southwestern Mauritania, Heath (2003) a collection of texts from further east in Mali, and Aguadé (1998) a brief overview of speech in southern Morocco. The latter shows features more similar to Moroccan (below), so I do not consider them here. More western varieties (Figure 6), including those in Mauritania and Gao, show a relatively simplex distribution of forms that looks much like Kuwaiti, i.e., an obligatory definite marker *il-* and auxiliary marking of AD referents with a demonstrative *dāk* or *dīk* (inflected for gender). Malian varieties around Gao (Figure 7), however, exhibit additional complexity in that they have a relatively frequent PSI marker *wāḥīd* (see Section 3.3). Heath (p. 8) asserts that “the grammar of Malian Hassaniya differs little from that of Mauritanian dialects,” but the current framework does raise the question of whether grammatical marking of PSI nouns might be a useful metric for internal classification of Hassaniya.

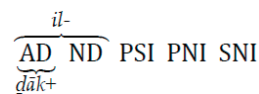


Figure 6. Reference Hierarchy for western Hassaniya varieties.

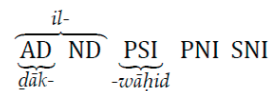


Figure 7. Reference Hierarchy for eastern Hassaniya varieties.

4.5. Sanaani

There is so much linguistic diversity in Yemen that I am hesitant to make broad pronouncements about “Yemeni,” and thus base my judgements here only on Watson and ‘Amri (2000) texts from Sanaa. In them, Sanaani (Figure 8) can be seen to obligatorily mark AD and ND statuses together with *il-*, like other varieties above, and also allows for auxiliary marking of AD with a preposed demonstrative *ḍayyik* (etc.).⁹ In addition, Sanaani has an auxiliary strategy, described in Section 3.3, wherein PSI referents can be further differentiated with what is elsewhere a demonstrative adverb *hākaḍā(yā)*. This marker is similar in function to the Egyptian PSI marker *kida*.

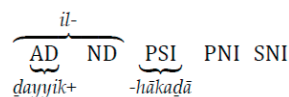


Figure 8. Reference Hierarchy for Sana’ani.

4.6. Levantine

Levantine varieties again show the pattern of marking AD and ND nouns with *il-*, and allow for additional delineation of AD nouns with an unstressed demonstrative *ha-*, but differ from varieties above in that they have a conventionalized article *ši* that denotes PNI referents (see Section 3.4). As discussed in Section 3.1, varieties of the Levant also make particularly productive use of anaphoric *ha-*, some perhaps to the extent that the resulting fused marker *hal-* should be considered its own, exclusive marker of AD statuses. Figure 9 gives a more conservative interpretation of the distribution of forms in Levantine, and Figure 10 offers the secondary analysis.

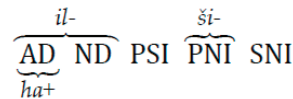


Figure 9. Reference Hierarchy for Levantine varieties.

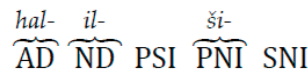


Figure 10. Possible reading for some Levantine varieties.

4.7. Iraqi

Arabic varieties in Iraq (Figure 11) have been described as having an indefinite **fard* (Blanc 1964, p. 118), and Leitner and Procházka’s (Forthcoming) focused semantic analysis supports the notion that this polyfunctional lexeme acts as a conventionalized PSI/PNI article in most Iraqi dialects (see Sections 3.3 and 3.4). Texts in Iraqi varieties also regular show the use of demonstrative *ha-* as an auxiliary AD marker alongside the obliterary definite marker *il-*, as is common elsewhere.

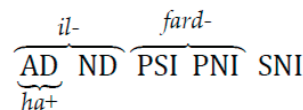


Figure 11. Reference Hierarchy for Iraqi varieties.

4.8. Najdi

The expression of definiteness in Najdi Arabic (Figure 12), as described in (Ingham 1994), somewhat parallels the formal distribution given for Iraqi above. For AD and ND nouns, *il-* is the obligatory article, with auxiliary marking of AD nouns possible with *ha-*. As a dialect that has so-called dialectical *tanwīn*, PSI and PNI nouns that are adnominally modified with adjectives, relative clauses, or prepositional phrases obligatorily have the marker *-in*. There is also evidence, described in Section 3.3, that at least some Najdi speakers can use DT on a purely semantic basis, i.e., without the noun being followed by any sort of modifier.

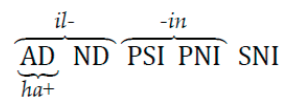


Figure 12. Reference Hierarchy for Najdi varieties.

4.9. Moroccan

Moroccan varieties (Figure 13) represent a relatively complex case, the main complications of which are that (1) the article *l-* is not restricted to definite (AD and ND) nouns

and (2) both PSI and PNI meanings are uniquely distinguished with overt, highly conventionalized articles. While the reflex of **al-* in all the above varieties is restricted and can thus truly be considered a definite article, in Moroccan it is conventionally extended to PSI referents (see Section 3.3) and is frequently used with SNI nouns as well (see Section 3.5). For PSI nouns, *l-* accretes with an article *wāḥad*, which is similar in function to the optional article found in eastern Hassaniya (Section 4.4); meanwhile, for PNI nouns, an article *ši-*—identical in form and meaning to that attested in the Levant (Section 4.6)—is used. Moroccan also allows for auxiliary indication of AD nouns with the proximal and distal anaphoric demonstratives *hād-* and *dāk-*, the former of which is uninflected.

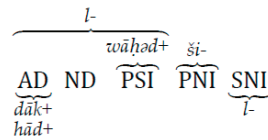


Figure 13. Reference Hierarchy for Moroccan varieties.

4.10. Central Asian

Central Asian varieties combine known strategies from elsewhere in Arabic with the unique feature of not having a reflex of **al-*; among others, this latter feature has probably played a role in these varieties being characterized as “metatypized” (Ratcliffe 2005), particularly given other nearby languages also lack true definite articles. There is evidence that Central Asian Arabic varieties can, like many others, use unstressed demonstratives for anaphoric (AD) reference (see 3.1). In addition, these dialects also show a reflex of **fard* that has the same PSI/PNI semantic scope of **fard* in Iraqi varieties (4.7). Central Asian also shows its own reflex of dialectal *tanwīn*, which sees the same syntactic conditioning as elsewhere (i.e., before adnominals), but has a wider semantic range because it can also occur with true definites.¹⁰ It is not attested with SNI nouns, but considering these are unlikely to be adnominally modified in the first place (see Section 3.5), it would not be unreasonable to say that DT in Central Asian has fully lost its referential dimensions, and can be envisioned purely as a syntactic linker, hence the question mark in Figure 14.

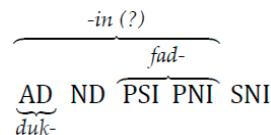


Figure 14. Reference Hierarchy for Central Asian varieties.

4.11. Nubi

Finally, while Wellens (2003), among others, has classified Nubi (Figure 15) as an Arabic-lexifier creole rather than a “true” Arabic variety, it is worthwhile to consider points of overlap with the above varieties in its expression of definiteness. Like Central Asian, Nubi has lost the article **al-*, differentiating it from the greater body of Arabic; nonetheless, also like Central Asian, the markers it does use have commonality with strategies attested in Arabic at large. The “definite article” *de* that Wellens identifies is, in my reading, primarily an AD article, and shares semantic scope with the many other demonstrative forms that mark anaphoric definiteness in Arabic dialects. In addition, the apparently polysemic PSI/PNI article *wai* has clear parallels with the postposed use of *wāḥid* in Hassaniya (Section 4.4).

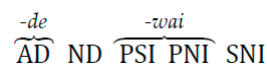


Figure 15. Reference Hierarchy for Nubi.

5. Definiteness and Classification

In theory, if the definiteness systems of Arabic dialects can be modeled, they should be relatively easy to classify. In practice, various complications arise that mean any attempt at classification will necessarily be subject to caveats and in need of ongoing refinement. As indicated more than once above, some of the systems themselves need more focused study to confirm how fully applicable the provisional models I have provided are to the dialect group as a whole. Scholars of Levantine Arabic, for example, face an open question as to just how close the unstressed anaphoric demonstrative complex *hal-* has come to acting as an obligatory article; similarly, scholars of Moroccan and Iraqi dialects may be able to further quantify uses of their respective indefinite articles in the same way by looking at them through a primarily semantic lens.

A related question is the concept of ‘obligatory’ vs. ‘auxiliary’, which I have attempted to frame here as a sort of continuum, the intermediate range of which might be described as ‘conventionalized’. For the purpose of grouping and classification, it seems that obligatory articles—those that are required when a speaker wants to denote a particular referential meaning—should take priority, as they represent a sort of linguistic consensus on the part of the speaker community that is not present for other markers. Nonetheless, is not always immediately clear what ‘obligatory’ means. It seems unwise to treat it as an absolute notion that only a single contrasting token would disqualify, especially when diglossic practices allow speakers to switch between registers (and their respective definiteness systems) at will. Instead, it seems more reasonable to look at the preponderance of the evidence: what forms *most often* arise in everyday conversation between native speakers of the variety in question? I suggest that these highly conventionalized strategies should also be prioritized for the purposes of classification.

This is not to say, either, that less frequent auxiliary strategies have no value, else I would not have included them here. To the contrary, it does appear worthwhile to point out that a majority of Arabic varieties optionally use unstressed demonstratives for anaphoric definite meanings, and that both varieties that do not (such as Egyptian) and varieties that oblige them (such as some in the Levant) are the outliers. It does seem relevant to note that not just one, but at least two, Arabic varieties (Egyptian and Sanaani) show the same typological pattern of co-opting a demonstrative adverb as a marker of specific indefinites, even if these are not required or even all that frequently used, statistically speaking, to express that meaning. Most importantly, although these are synchronic patterns, all fully crystallized innovations were presumably in flux at one time, so for the historical record alone it is worth noting that such strategies exist.

With these qualifications in mind, then, we can approach the question of classification more directly. I propose that there are two primary methodologies for grouping dialects when looking at a set of interrelated semantic features, as is the case with definiteness. The first is a ‘single-tier’ approach, meaning we simply limit our view to a particular type of meaning within the Reference Hierarchy, survey the forms that are attested for it, and order them into groups. This approach is not particularly distinctive from the survey I provided in Section 3, and can be useful as a starting point for hypotheses, especially because it is suitable for identifying outliers. The Central Asian group, for example, clearly stands out in that it does *not* obligatorily mark definite (AD/ND) nouns (see Sections 3.1 and 3.2), and Moroccan clearly stands out in that it *can* mark full indefinite (SNI) nouns (see Section 3.5). Nonetheless, while this approach might be initially useful for looking beyond forms and toward semantic function—e.g., for noting that *šī* and **fard* have at least partial semantic overlap—it is not particularly useful for comparing systems as whole.

Instead, I offer that a preferable approach is to look at the distribution of forms holistically, in what might be called a ‘multi-tier’ approach. It is still necessary, of course, that we prioritize some features over others as a means of subgrouping, but as a general principle I hold that each primary subgroup should be selected to describe as many varieties as possible while whittling away the outliers. One possible schema, based on the

comparative systems given in Section 4 (minus Nubi), and taking into account the above points about obligatory and conventionalized forms, as is follows:

1. Dialects with strict formal distinction between true definites and indefinites . . .
 - a. No highly conventionalized marking of indefinites . . .
 - i. No attested auxiliary strategies: **Libyan, Kuwaiti**
 - ii. Attested auxiliary strategies: **Egyptian, Hassaniya, Sanaani**
 - b. Highly conventionalized marking of some indefinites . . .
 - i. Marking syntactically determined: **tanwīn dialects; Najdi**
 - ii. Marking semantically or pragmatically determined . . .
 1. Single marker for specific (PSI) and existential (SNI) indefinites: **Iraqi**
 2. Marker for existential (SNI) indefinites only: **Levantine**
2. Dialects with lax formal distinction between true definites and indefinites . . .
 - a. Marked definites: **Moroccan**
 - b. Unmarked definites: **Central Asian**

There are admittedly other ways in which this same set of metrics could be ordered, and the varieties in question consequently be grouped, but this one has a few advantages. The first is that the present classification does give some credence to traditionalist views of Arabic as having a normative system where **al-* is a “definite article,” while leaving room for exceptions and, at the same time, expanding the profile of what a “normative” dialect is by showing that a majority of these do have at least some means of marking indefinite referents, a pattern that stretches from the Atlantic to the Gulf. A second advantage is that the classification serves to group together varieties that might not necessarily share features, but which do share basic semantic patterns, in turn opening the door for diachronic questions, especially when these varieties are geographically distant from each other. I do not mean to imply by this a hereunto undiscovered genetic relationship between Moroccan and Central Asian varieties, but I do mean to point out that both groups have seen the strict categorical distinction between definites and indefinites unravel, and they are both at the far ends of the Arabic-speaking world.

Interpreted this way, the definiteness data align most closely with a ‘core-periphery’ classification model, in that a strict formal distinction between definites and indefinites is maintained across a large, contiguous cultural area and frays only at its edges. Within the core area, there is frequent variation in the particular means of marking referential indefiniteness, and somewhat of a northern–southern split as one moves from unmarked or optional marking strategies of Egypt, Yemen, and the Gulf to the more conventionalized strategies of the Levant and Mesopotamia, but the strict and exclusive association of **al-* with definiteness goes unchallenged. Meanwhile, on the geographic fringes of this core, dialects break away typologically by either (1) extending **al-* to indefinite meanings or (2) detaching it from definite meanings.¹¹ The concept of peripheral dialects has been explored in volumes such as Owens (2000) and Anghelescu and Grigore (2007), and even though such varieties are just as often defined by what they are not than what they have in common, the addition of definiteness as a metric does at least support the idea of the ‘core’ against which they are defined as a viable linguistic entity.

Other classification proposals do not align as well with a scheme based on definiteness systems. The oft-proposed east–west division of dialects (see Palva 2006) is not easily evident here, especially given that the minimal expression of indefiniteness in Hassaniya varieties fall into the same general pattern as dialects much further east, including those of Egypt, Yemen, and Kuwait. The bedouin–sedentary division (again see Palva) is tenable only on the basis of the *tanwīn* feature, which is largely limited to bedouin-type varieties and is unique among indefinite markers in that it is conditioned by syntactic factors in addition to semantic ones. Nonetheless, in a purely typological sense, the presence of a conventionalized indefinite marker actually places DT-expressive bedouin varieties such as

Najdi closer to the indefinite-marking sedentary dialects of the Levant and Mesopotamia than it does to other bedouin varieties that lack it, such as western Hassaniya or Kuwaiti. Finally, one may consider whether, within the sedentary dialects, an urban-rural division is relevant; this too seems unlikely, given the systems found in a given geographic region *do* tend to be contiguous across urban and rural areas. The Levantine PNI article *šī*, for example, is used by speakers both in Beirut and small mountain villages in the same way that the Moroccan PSI article *wāḥad* is found both in the old cities and rural countryside.

In summary, the system-level configuration of definiteness marking does ultimately seem to be an areal pattern, and even minor differences between systems might consequently be useful for further subdividing clusters of geographically adjacent dialects. This possibility has already been raised for eastern vs. western Hassaniya (Section 4.4), as well as Levantine (Section 4.6) varieties. I also offer the observation that somewhere between central Algeria and Tunisia, dialects see an abrupt shift from complex, Moroccan-like systems (Section 4.9) to simplex, Libyan-like systems (Section 4.1). Precisely where these lines may lie—and why—is a question for future studies to address. Many of the systems in question seem to be the product of innovation, whether via semantic extension or leveling, and whether prompted by contact or otherwise. As it seems reasonable to expect that groups that innovate together, along the same timeline and to the exclusion of nearby groups, are indeed more likely to share history and social ties, further studies on definiteness and referentiality in spoken Arabic will be of value to the larger project of dialect classification.

6. Conclusions

In this paper I have outlined the process of building a novel classification scheme for Arabic dialects, using semantic typology as a metric for grouping rather than relying on the presence of forms alone. Taking definiteness as a case study, I discussed a selection of possible models, and adopted Dryer's (2014) 'Reference Hierarchy' as the most suitable of these for the task of envisioning definiteness systems in Arabic. I thereafter showed that, for expression of each semantic status along the Reference Hierarchy, the dialectological literature attests multiple strategies across the Arabic-speaking world. This variability can be made more useful for classification by modeling the semantic distribution of forms for discrete dialects holistically and then placing those models side by side, in turn allowing us to look past the forms themselves and instead class the dialects by shared typological characteristics. Key metrics that emerge are whether varieties maintain a strict formal delineation between true definites and indefinites, whether they overtly distinguish referential indefinites, and whether the latter is subject to syntactic conditions beyond the semantic ones. This particular classification approach does not align well with some traditional proposals, such as an east–west or bedouin–sedentary split, but it does lend some credence to the idea of a 'core' dialect area that contrasts with a 'periphery'.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: I would like to acknowledge Kristen Brustad, Mahmoud Al-Batal, Pattie Epps, and Cinzia Russi, all of whom served on the committee for the dissertation in which many of these ideas were developed. I would also like to thank the two anonymous reviewers for their valuable insights and suggestions.

Conflicts of Interest: The author declares no conflict of interest.

Notes

- 1 For the sake of simplicity, I use **al-* to refer to this article and all its various phonological realizations in the dialects. The same is true of other markers that have a common etymological source, such as **wāḥid* and **fard*. The precise shapes of their reflexes are not particularly relevant to a semantic analysis, and are already well documented.
- 2 Readers should refer to the original sources, cited alongside the examples, for further context. As I draw some conclusions independent of those of the original authors, errors of interpretation are my own.
- 3 In the present article I do not treat generics or plurals, although there is strong evidence for variation among them as well, with likely diachronic implications; see Turner (2018), pp. 232–35).
- 4 Rosenhouse (1984, p. 82) notes the same pattern in Bedouin varieties of northern Israel, stating that “often this attachment is so strong that it seems to lose the demonstrative function and serve only for definition of the noun”. To this I would add the caveat that it highlights *anaphoric* definition specifically.
- 5 The lexeme *ši* itself is polyfunctional, as explored in detail in Wilmsen’s (2014). Wilmsen (51–53) calls this particular use ‘partitive *ši*’ and notes its “indefinite determiner function as marking a quality somewhere between indefinite and definite,” as is descriptive of PNI in the current framework.
- 6 For example, ‘*ʿandāk šī šīlu?*’ ‘do you have some sort of pen?’ is often used in the sense of ‘do you have a pen [I can borrow]?’ in Moroccan speech. Even though from the speaker’s perspective there need be nothing particular about the ‘pen’ in question, allowing that there might be is a polite deferral to the listener. Leitner and Procházka’s (Forthcoming) call this discursive strategy “mitigation” and locate it as a use of **fard* in Iraq and Khuzistan.
- 7 Although using **al-* with singular SNI nouns is not possible in most varieties, a much greater number allow it with unquantified SNI plurals; see, for example, *s-sbūʿa* ‘lions’ in example (9). In this light, Moroccan might be seen as having simply leveled a more widespread plural paradigm to singulars.
- 8 For this and the following models, I use a hyphen [-] to indicate the syntactic position of the marker in relation to the noun, and a plus sign [+] to indicate both the marker’s syntactic position and that it accretes with other markers in the same semantic range. Like in Figure 2 (for English), forms given at top are either obligatory or highly conventionalized, whereas forms at bottom represent more marked auxiliary strategies.
- 9 Demonstratives in Sanaʿani are highly variable (see Watson and ʿAmri 2000, p. 20); they appear to be used interchangeably in this sense, and are inflected for gender.
- 10 For example, *duk parvardigōr-in ki lā-yi fi raḥim umm-i ḥāvī-ni* ‘the protector who protected me in my mother’s womb’ (Ingham 2003, p. 36).
- 11 Similar “fraying” of the definiteness system occurs in Arabic varieties of southern Iran (Matras and Shabibi 2007) and southern Turkey (Akkuş 2016), where unmarked definite head nouns are attested, albeit under different syntactic constraints. In Maltese, the strict association of **al-* with definites has been lost for adjectival attributes; see Fabri (2001).

References

- Abu-Haidar, Farida. 1979. *A Study of the Spoken Arabic of Baskinta*. Leiden and London: E. J. Brill.
- Aguadé, Jordi. 1998. Relatos en hassaniyya recogidos en Mhamid (valle del Dra, sur de Marruecos). *Estudios de dialectología norteafricana y andalusí* 3: 203–15.
- Akkuş, Faruk. 2016. The Arabic Dialect of Mutki-Sason Areas. In *Arabic Varieties: Far and Wide: Proceedings of the 11th International Conference of AIDA, Bucharest, 2015*. Edited by Gheorghe Grigore and Gabriel Biţună. Bucharest: Editura Universităţii din Bucureşti, pp. 29–40.
- Angelescu, Nadia, and George Grigore, eds. 2007. *Peripheral Arabic Varieties*. Romano-Arabica, VI–VII. Bucharest: Center for Arab Studies.
- Behnstedt, Peter, and Manfred Woidich. 1988. *Die ägyptisch-arabischen Dialekte: Texte. Niltaldialekte, Oasendialekte*. Wiesbaden: Reichert, vol. 3.
- Belyayeva, Dina. 1997. Definiteness Realization and Function in Palestinian Arabic. In *Perspectives on Arabic Linguistics: Papers from the Annual Symposium on Arabic Linguistics, Salt Lake City, 1996*. Edited by Mushira Eid and Robert R. Ratcliffe. Amsterdam: John Benjamins Publishing, vol. X, pp. 47–67.
- Blanc, Haim. 1964. *Communal Dialects in Baghdad*. Cambridge: Harvard University Press.
- Brustad, Kristen. 2000. *The Syntax of Spoken Arabic: A Comparative Study of Moroccan, Egyptian, Syrian, and Kuwaiti Dialects*. Washington, DC: Georgetown University Press.
- Caubet, Dominique. 1983. *La détermination en arabe marocain*. Paris: Université Paris 7, Dép. de recherches linguistiques, Laboratoire de linguistique formelle.
- Caubet, Dominique. 1993. *L’arabe Marocain: Syntaxe et Catégories Grammaticales, Textes*. Paris/Louvain: Peeters.
- Chafe, Wallace L. 1976. Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View. In *Subject and Topic*. Edited by Charles N. Li. New York: Academic Press, pp. 159–82.
- Cohen, David. 1963. *Le Dialecte Arabe Ḥassāniyya de Mauritanie, Parler de La Gbela*. Paris: C. Klincksieck.

- De Mulder, Walter, and Anne Carlier. 2011. The Grammaticalization of Definite Articles. In *The Oxford Handbook of Grammaticalization*. Edited by Bernd Heine and Heiko Narrog. Oxford: Oxford University Press.
- Denz, Adolf, and Dietz Otto Edzard. 1966. Iraq-Arabische Texte Nach Tonbandaufnahmen Aus al-Hilla, al-Afač Und al-Basra. *Zeitschrift Der Deutschen Morgenländischen Gesellschaft* 116: 60–96.
- Dryer, Matthew. 2005a. Definite Articles. In *The World Atlas of Language Structures*. Edited by Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie. Oxford: OUP Oxford, pp. 154–57.
- Dryer, Matthew. 2005b. Indefinite Articles. In *The World Atlas of Language Structures*. Edited by Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie. Oxford: OUP Oxford, pp. 158–61.
- Dryer, Matthew. 2014. Competing Methods for Uncovering Linguistic Diversity: The Case of Definite and Indefinite Articles (Commentary on Davis, Gillon, and Matthewson). *Language* 90: e232–49. [[CrossRef](#)]
- Elfitoury, Abubaker Abdalla. 1976. A Descriptive Grammar of Libyan Arabic. Unpublished Ph.D. thesis, Georgetown University, Washington, DC, USA.
- Fabri, Ray. 2001. Definiteness Marking and the Structure of the NP in Maltese. *Verbum* 2: 153–72.
- Fassi Fehri, Abdelkader. 2012. *Key Features and Parameters in Arabic Grammar*. Amsterdam: J. Benjamins Pub. Co.
- Givón, Talmay. 1978. Definiteness and Referentiality. In *Universals of Human Language: Vol. 4, Syntax*. Edited by Joseph Harold Greenberg, Charles Albert Ferguson and Edith A. Moravcsik. Palo Alto: Stanford University Press, pp. 291–330.
- Grand'Henry, Jacques. 2000. Deux Textes Arabes de Benghazi (Libye). *Oriente Moderno* 19: 47–57. [[CrossRef](#)]
- Gundel, Jeanette K., Nancy Hedberg, and Ron Zacharski. 1993. Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69: 274–307. [[CrossRef](#)]
- Harrell, Richard S. 1966. *A Dictionary of Moroccan Arabic: Arabic-English*. The Richard Slade Harrell Arabic Series: No.9; Washington: Georgetown University Press.
- Heath, Jeffrey. 2003. *Hassaniya Arabic (Mali): Poetic and Ethnographic Texts*. Leipzig: Otto Harrassowitz Verlag.
- Heine, Bernd. 1997. *Cognitive Foundations of Grammar*, 1st ed. New York: Oxford University Press.
- Hillelson, Sigmar. 1935. *Sudan Arabic Texts*. Cambridge: Cambridge University Press.
- Ingham, Bruce. 1994. *Najdi Arabic: Central Arabian*. Amsterdam: John Benjamins Publishing Company.
- Ingham, Bruce. 2003. Language Survival in Isolation: The Arabic Dialect of Afghanistan. In *AIDA Proceedings: Fifth International Conference of the Association Internationale de Dialectologie Arabe*. Cádiz: Universidad de Cádiz Publications.
- Ionin, Tania. 2006. This Is Definitely Specific: Specificity and Definiteness in Article Systems. *Natural Language Semantics* 14: 175–234. [[CrossRef](#)]
- Israel, Michael. 1999. Some and the Pragmatics of Indefinite Construal. *Proceedings of Berkeley Linguistics Society* 25: 121–32. [[CrossRef](#)]
- Jastrow, Otto. 2005. Uzbekistan Arabic: A Language Created by Semitic-Iranian-Turkic Linguistic Convergence. In *Linguistic Convergence and Areal Diffusion: Case Studies from Iranian, Semitic and Turkic*. Edited by Éva Ágnes Csató, Bo Isaksson and Carina Jahani. Hove: Psychology Press, pp. 133–39.
- Lahdo, Ablahad. 2009. *The Arabic Dialect of Tillo in the Region of Siirt (South-Eastern Turkey)*. Acta Universitatis Upsaliensis Studia Semitica Upsaliensia 26. Uppsala: Uppsala University, Department of African and Asian Languages.
- Leitner, Bettina, and Stephan Procházka. Forthcoming. The polyfunctional lexeme /fard/ in the Arabic dialects of Iraq and Khuzestan: More than an indefinite article. *Brill's Journal of Afroasiatic Languages and Linguistics* 13.
- Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press.
- Magidow, Alexander. 2016. Diachronic Dialect Classification with Demonstratives. *Al-Arabiyya* 49: 91–115.
- Matras, Yaron, and Maryam Shabibi. 2007. Grammatical Borrowing in Khuzistani Arabic. In *Grammatical Borrowing in Cross Linguistic Perspective*. Berlin: Mouton de Gruyter.
- McCarthy, Richard. J., and Faraj Raffouli. 1965. *Spoken Arabic of Baghdad, Part Two (A): Anthology of Texts*. Beirut: Librairie Orientale.
- Mion, Giuliano. 2009. L'indétermination nominale dans les dialectes arabes. Une vue d'ensemble. In *Miscellanea Arabica 2009*. Edited by Angelo Arioli. Rome: Edizioni Nuova Cultura, pp. 215–31.
- Owens, Jonathan. 1984. *A Short Reference Grammar of Eastern Libyan Arabic*. Wiesbaden: Otto Harrassowitz.
- Owens, Jonathan. 2000. *Arabic as a Minority Language*. Berlin: Walter de Gruyter.
- Palva, Heikki. 2006. Dialects: Classification. In *Encyclopedia of Arabic Language and Linguistics*. Edited by Kees Versteegh. Leiden and Boston: Brill, vol. 1, pp. 604–13.
- Ratliffe, Robert. 2005. Bukhara Arabic: A Metatypized Dialect of Arabic in Central Asia. In *Linguistic Convergence and Areal Diffusion: Case Studies from Iranian, Semitic and Turkic*. Edited by Éva Ágnes Csató, Bo Isaksson and Carina Jahani. Hove: Psychology Press, pp. 141–51.
- Rosenhouse, Judith. 1984. *Bedouin Arabic Dialects: General Problems and a Close Analysis of North Israel Bedouin Dialects*. Weisbaden: Otto Harrassowitz.
- Seeger, Ulrich. 2013. Zum Verhältnis Der Zentralasiatischen Arabischen Dialekte Mit Einem Bisher Unveröffentlichten Text Aus Südchorasán. In *Nicht Nur Mit Engelszungen: Beiträge Zur Semitischen Dialektologie-Festschrift Für Werner Arnold Zum 60. Geburtstag*, 1st ed. Edited by Renaud Kutay, Ulrich Seeger and Shabo Talay. Wiesbaden: Otto Harrassowitz, pp. 313–22.
- Stokes, Phillip W. 2020. A Fresh Analysis of the Origin and Diachronic Development of 'Dialectal Tanwīn' in Arabic. *Journal of the American Oriental Society* 140: 637–64. [[CrossRef](#)]
- Turner, Michael. 2018. Definiteness in the Arabic Dialects. Ph.D. dissertation, The University of Texas at Austin, Texas, USA.

- Vicente, Ángeles. 2000. *El Dialecto Árabe de Anjra (Norte de Marruecos)*. Zaragoza: Universidad de Zaragoza.
- Watson, Janet C. E., and 'Abd al-Salām 'Amri. 1993. *A Syntax of Ṣan'ānī Arabic*. Leipzig: Otto Harrassowitz Verlag.
- Watson, Janet C. E., and 'Abd al-Salām 'Amri. 2000. *Wasf Ṣan'ā: Texts in Ṣan'ānī Arabic*. Wiesbaden: Harrassowitz.
- Wellens, Inneke Hilda Werner. 2003. *An Arabic Creole in Africa: The Nubi Language of Uganda*. Ph.D. dissertation, Radboud University, Nijmegen, The Netherlands.
- Wilmsen, David. 2014. *Arabic Indefinites, Interrogatives, and Negators: A Linguistic History of Western Dialects*. Oxford: Oxford University.
- Woidich, Manfred. 2006. *Das Kairenisch-Arabische. Eine Grammatik*. Leipzig: Otto Harrassowitz Verlag.
- Yoda, Sumikazu. 2005. *The Arabic Dialect of the Jews in Tripoli (Libya): Grammar, Text and Glossary*. Leipzig: Otto Harrassowitz Verlag, vol. 35.

Article

Interrogating the Egypto-Sudanic Arabic Connection

Thomas A. Leddy-Cecere

Bennington College, Bennington, VT 05201, USA; thomasleddycecere@bennington.edu

Abstract: The Arabic dialectology literature repeatedly asserts the existence of a macro-level classificatory relationship binding the Arabic speech varieties of the combined Egypto-Sudanic area. This proposal, though oft-encountered, has not previously been formulated in reference to extensive linguistic criteria, but is instead framed primarily on the nonlinguistic premise of historical demographic and genealogical relationships joining the Arabic-speaking communities of the region. The present contribution provides a linguistically based evaluation of this proposed dialectal grouping, to assess whether the postulated dialectal unity is meaningfully borne out by available language data. Isoglosses from the domains of segmental phonology, phonological processes, pronominal morphology, verbal inflection, and syntax are analyzed across six dialects representing Arabic speech in the region. These are shown to offer minimal support for a unified Egypto-Sudanic dialect classification, but instead to indicate a significant north–south differentiation within the sample—a finding further qualified via application of the novel method of Historical Glottometry developed by François and Kalyan. The investigation concludes with reflection on the implications of these results on the understandings of the correspondence between linguistic and human genealogical relationships in the history of Arabic and in dialectological practice more broadly.

Keywords: dialect classification; subgrouping; Sudanic Arabic; Egyptian Arabic

Citation: Leddy-Cecere, Thomas A. 2021. Interrogating the Egypto-Sudanic Arabic Connection. *Languages* 6: 123. <https://doi.org/10.3390/languages6030123>

Academic Editors: Roberta Morano and Simone Bettega

Received: 10 June 2021
Accepted: 16 July 2021
Published: 23 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

This investigation intends a twofold contribution to the advancement of Arabic dialect classification. In the finer grain, I present an empirical evaluation of the frequently asserted macro-level classificatory grouping comprising the Arabic dialects of Egypt and those of the greater Sudanic region (Fischer and Jastrow 1980; Kaye and Rosenhouse 1997; Dickins 2011; Versteegh 2014). At a broader scale, I seek to interrogate the principal theoretical premise invoked in support of this proposed classification: that the shared (human) genealogical history of speech communities constitutes a reliable a priori basis for the classification of those communities' dialects in terms of diachronic relatedness and/or synchronic similarity. While the former priority will primarily engage specialists in Arabic and related languages, it is hoped that the latter will provide reflection pertinent both within and beyond the Arabist sphere, and initiate mutually informative conversations with colleagues of diverse foci, perspectives and expertise.

Dialectological description of Arabic dialects spoken in the Egyptian and Sudanic areas is remarkable for its unevenness. This broad, contiguous zone extends from the Mediterranean in the north to the Sudan–South Sudan border region in the south, and from the Red Sea westward to the Libyan Desert and, further, the vicinity of Lake Chad in Central Africa—the north–south stretch of the Nile Valley constituting an organizing “spine” and center of geographic and demographic gravity. Arabic varieties spoken in this region are utilized by a combined speaker population well in excess of 100 million (Eberhard et al. 2021). Knowledge of dialect diversity in the Egyptian portion of the zone has benefitted immensely from the achievement of Behnstedt and Woidich's (1985–1999) multivolume dialect atlas, text collection and glossary, and analysis of the dialect of Cairo has been particularly thorough (esp. Woidich 2006b). In comparison, Fischer and Jastrow could write of the vast Sudanic Arabophone territory as late as 1980 that “[w]ir haben

zwar aus diesen Raum eine Anzahl Texte und einige Lehrbücher sowie Vokabulare, aber nicht eine einzige halbwegs moderne grammatische Monographie“ [We have from this area a number of texts and some textbooks, as well as vocabularies, but not a single halfway modern grammatical monograph] (Fischer and Jastrow 1980, p. 31). The state of scholarship has improved somewhat since, with the publication of two key book-length treatments of varieties local to the east (Reichmuth 1983) and far west (Owens 1993a) of the Sudanic region alongside thematic analyses of structural phenomena in, respectively, urban and semi-nomadic lects of Sudan’s center (Dickins 2007a, 2009, 2010) and west (Manfredi 2014, 2018). Even so, the differential in scholarly attention to the Egyptian and Sudanic dialect areas remains severe, and the two, respectively, contain some of the best and least described speech varieties of modern Arabic.

The relevance of this imbalance is heightened when taken in combination with the fact that dialects of the combined Egypto-Sudanic zone are commonly associated with one another in discussions of Arabic dialect classification and subgrouping, frequently culminating in their collective classification as an identifiable dialectological unit superordinate to more localized groups. Illustrative articulations of this view are, among others, Kaye and Rosenhouse’s assertion that “[a]s a whole, Sudanese dialects, at least those in the north, form one macro-grouping with the Egyptian dialects” (Kaye and Rosenhouse 1997, p. 265), and Fischer and Jastrow’s positioning of the dialects of central and eastern Sudan as “[d]ie südliche Fortsetzung der oberägyptischen Dialekte” [the southern continuation of the Upper Egyptian dialects] (Fischer and Jastrow 1980, p. 31). To some extent, this lumping may stem from the shared failure of a number of varieties in both the Egyptian and Sudanic areas to clearly align with either of two primary classificatory dichotomies espoused by Arabic dialectologists, the Bedouin vs. sedentary split and the Eastern Arabic vs. Western Arabic split (cf. Heikki Palva 2006): Fischer and Jastrow, for instance, describe the collected dialects of Egypt and the Sudan as taking “eine Sonderstellung zwischen denen des Ostens und des Maghrib” [a special position between those of the East and those of the Maghrib] (Fischer and Jastrow 1980, p. 29). Such negative characterizations, however, framed on these dialects’ incongruity with external typologies, do little to positively establish dialectal unity within the Egypto-Sudanic region. In this regard, analysts like the latter authors instead place particular emphasis on the identification of Egypt as the primary source for the historic in-migration of Arabic speakers to the greater Sudan (Fischer and Jastrow 1980, p. 22). It is this second criterion—the putative common genealogical history of the Egyptian and Sudanic Arabophone speech communities—which has most frequently and most prominently featured as the anchoring factor of proposed classificatory relationships between Egyptian and Sudanic Arabic dialects.

Present in the influential early work of Kaye (1976), reliance on genealogical connection persists as the dominant narrative of more recent scholarship linking Egyptian and Sudanic Arabic. This reasoning is encapsulated in Dickins’ position that, “[r]eflecting the fact that the major penetration route of Arabic speakers was from Upper Egypt, through Nubia into Central Sudan, CUSA [Central Urban Sudanese Arabic] is more closely related to Egyptian Arabic—and particularly the Ša’īdī [Upper Egyptian] dialects, than any other non-Sudanese dialects” (Dickins 2011, p. 936). Likewise, Versteegh, in describing varieties of the combined Egypto-Sudanic area under the heading “Egyptian dialects,” frames his account with the assertion that “[f]rom Egypt, the Arabic language was brought along the Nile to the South, into Sudan and Chad” (Versteegh 2014, p. 205). Certainly, the correlation of linguistic isoglosses to paths of migration and human movement remains a venerable and valuable practice in the Arabist tradition (Behnstedt and Woidich 2005; Heikki Palva 2006) and dialectology more generally (Chambers and Trudgill 2004; Britain 2016). In the Egypto-Sudanic case, however, the practice has not precisely been realized. Likely connected to the comparative lack of reliable dialectological description in the Sudanic portion of the area, observations of genealogical links between Egyptian and Sudanic speech communities have most often been proffered in place of a detailed accounting of shared linguistic features, rather than alongside one—thus positioning common genealogy as a direct *indicator* of

dialectal classificatory relationship, not an *explanans* to be utilized in the interpretation of a relationship separately established on linguistic grounds. On the whole, it is observed that macro-level co-classifications of Egyptian and Sudanic dialects have tended to proceed from the extra-linguistically founded premise of shared population origin to treat the collective body of Sudanic Arabic, definitionally, as “originally a dialect of an Egyptian dialect of Arabic” (Kaye 1976, p. 177), and to subsequently adduce linguistic evidence of this—linguistic—relationship only in a secondary, corroborating fashion (if at all).

Gratefully, a small number of exceptions to this general pattern are to be found. Owens (1993b) undertakes a thorough investigation of the dialectological relationships of Nigerian Arabic, which he situates within concentric spheres of affiliation incorporating, successively, other West Sudanic varieties, Sudanic Arabic writ large, and (primarily Upper) Egypt. In a further (2003) work, the same author presents a focused and convincingly argued account of the migratory dispersal of a particular feature, inflection of the first person imperfect, between specified subregions of the Egypto-Sudanic area. These contributions prove marked advancements in the understanding of dialectal inter-relationships within the region, and stand out for their reliance on concrete linguistic data. However, given their targeted framing and methodological emphasis on “patchwork” features which typify particular pairings/subsets of Egypto-Sudanic varieties but not the area as a whole (Owens 1993b, p. 158), these studies are not positioned to stand as full corrective or confirmation to the more broadly construed claims of macro-level classificatory unity so often advanced elsewhere in the literature. Approaching that task in more direct yet far more perfunctory fashion is Reichmuth, who in the introductory pages of his descriptive grammar of the East Sudanese dialect of the Šukriyya (Reichmuth 1983, pp. 24–29) sketches the extra-Sudanic incidence of several isoglosses characteristic of that variety as a baseline evaluation of its compatibility with the proposal of an Egypto-Sudanic subgroup, among other potential affiliates. Though he does identify a degree of isoglossic overlap between Šukriyya and Egyptian forms, he deems the comparison inconclusive and unable to demonstrate a direct taxonomic dependency. Valuable as Reichmuth’s work may be in conception, the preliminary state of his evaluation and its conscious limitation to the focal point of the Šukriyya variety unfortunately constrain its usefulness as a linguistically anchored counterpoint to the genealogy-centered accounts of Egypto-Sudanic subgrouping that continue to dominate Arabist discourse.

It is in relation to this lacuna that I frame the present contribution: a linguistic investigation of the validity of the proposed linking of the Arabic dialects of the Egypto-Sudanic region as a macro-level classificatory unit, as has been prominently and repeatedly proposed in the Arabist literature on the nonlinguistic grounds of shared genealogical history. As described in detail in the following subsections, I shall present data from a selection of Sudanic and Egyptian varieties for analysis via both conventional and more innovative dialectological methods to determine whether their common classification as a macro-level dialect grouping is linguistically justified—or whether, in Reichmuth’s words, “[s]o bleibt nur die Annahme gemeinsamer Ursprünge übrig” [all that remains is the assumption of common origins] (Reichmuth 1983, p. 29).

2. Methods and Sources

Consistent with the framing described just above, this investigation does not seek to re-litigate the historical basis of shared genealogies and migration paths that dialectologists and others have considered to bind Arabic speakers of the Egypto-Sudanic region. That these have their root in the first major demographic influx of Arabs westward into Egypt in the seventh century, thence southward into the Sudanic area—incipient as early as the tenth century, more saliently from the fourteenth onward (with prominent place given to tribal entities including the Juhayna and the Ja’aliyyīn)—is largely accepted in the historical literature and has not substantively fluctuated over the previous century of scholarship (Holt and Daly 2011; and cf. MacMichael 1922). Though this consistency does not elevate the accepted narrative of events or its central tenets beyond any question or criticism

(see, e.g., Spaulding 2000), it does make it likely that any meaningful revision of these understandings needs be based in a specialist comprehension of historical demography and supported by the advent of novel or reinterpreted historical data—neither of which I claim here. Instead, the present inquiry, true to its conception, centers on the evaluation of the linguistic relationship purported to mirror these historical genealogical linkages and connect the region’s speech varieties in a manner worthy of reflection in macro-level schemes of Arabic dialect classification.

To accomplish this, I compare dialectological data from a sampling of Arabic varieties local to the proposed Egypto-Sudanic dialect area, in order to establish the definition and incidence of isoglosses which might weigh for or against the identification of a region-wide dialectal unity. I have selected six dialects to serve as core sources of data for this inquiry: three from the Egyptian portion of the zone and three from the Sudanic. The choice of the latter, especially, is constrained on the basis of available descriptive material. Thus, I have opted for the two varieties of Sudanic Arabic most comprehensively documented via book-length descriptive grammars—the dialect of the Šukriyya of eastern Sudan’s Buṭāna region, as described by Reichmuth (1983), and that of Arabic speakers living in northeastern Nigeria’s Borno state, documented by Owens (1993a)—in addition to the dialect of Khartoum, here mainly reflecting the grammatical sketches of Dickins (2007b, 2011), as occasionally supplemented by material from Bergman (2002) and Hillelson (1935). Together, these exemplify the West Sudanic type (Nigerian) and both traditional (Šukriyya) and urban (Khartoum) speech forms of the core Sudanic area. The three varieties representing the Egyptian portion of the region comprise, from north to south, those of Cairo (Woidich 2006b), Qift (Nishio 1995) on the east bank of the Nile in Upper Egypt, and the il-Bī’erāt territory on the Nile’s west bank opposite Luxor (Woidich 2006a). Drawn from a larger pool of available descriptive material, these Egyptian varieties have been selected to provide a focus on the Nile Valley, due to its centrality in existing discussions of dialectal interrelationship within the Egypto-Sudanic sphere (Owens 2003; Versteegh 2014). While the dialects of Qift and il-Bī’erāt are spoken quite near to one another in absolute terms, each is recognized as belonging to a distinct dialectal subregion of Upper Egyptian Arabic (cf. Behnstedt and Woidich 2018). This sampling of six varieties is not intended to be comprehensive, but rather sufficiently representative to establish the minimum viability of a proposed Egypto-Sudanic dialect classification—in the view that any isogloss with the potential to support a unified Egypto-Sudanic grouping should provide a detectable signal in *at least* this subset of six dialects, and that the artificial reduction in dialect diversity this (or any) sampling entails is more likely to overestimate the incidence of globally unifying features than to ignore them.

As to the nature of such potential features, this inquiry will address variation across the six dialects examined in the areas of phonology (segmental phonology and synchronic phonological processes), pronominal systems (personal, demonstrative, relative and interrogative), verbal inflectional morphology (agreement and tense-aspect-mood marking), and selected areas of syntax (negation, analytic possession, and demonstrative and interrogative word orders) in the attempt to identify shared features which might serve to join all or most of the six in support of a unified Egypto-Sudanic dialect grouping. The first three of these domains, and the features within them, have been chosen for (a) their consistently important roles in existing frameworks of Arabic dialect classification and (b) their attestation via comparable qualities of data across the individual dialect descriptions consulted. The fourth domain, that of syntax, is less commonly relied upon than these first three in general Arabic dialectological surveys,¹ but is included here due to its prominence in discussion of Egypto-Sudanic varieties, specifically (e.g., Versteegh 2014). Coverage within each domain will strive to be inclusive of all potentially relevant variables but will generally restrict discussion to features which typify two or more of the speech varieties under examination, favoring a focus on cross-dialectal commonality rather than individually defining features.

Once the data pertaining to each of these domains have been presented and appropriately described, the global results will be evaluated to determine their consistency with an Egypto-Sudanic classificatory unit proposed on the basis of shared genealogical history of the region’s speech communities. Consistent with the migration-based narrative’s inherent implication of diachronic linguistic relatedness, elements of these findings will also be assessed in the light of directly attested historical data, as well as comparative and internal reconstructive analyses. Following this, further insight will be derived via application the novel model of Historical Glottometry developed by François and Kalyan (François 2014; Kalyan and François 2018), which will be shown to offer interpretatively relevant perspective on the complex data at hand. Following discussion of these points, I will reflect on their implications for direct reliance on shared genealogical history in the shaping of linguistic classificatory schemes—in the Egypto-Sudanic case, in Arabic at large, and, by extension, as a practice adopted by students and scholars of dialectology more generally.

3. Results

The following subsections present results of the investigation of phonological, pronominal, verbal inflectional, and syntactic variables in the six dialects of the Egypto-Sudanic area currently under consideration. Unless otherwise specified in the text or via a table note, data for each dialect are derived from the descriptive source mentioned in association with that variety in Section 2, above. As relevant, the incidence of a given dialect feature in Arabic varieties spoken outside the immediate study area will also be noted.

3.1. Phonology

3.1.1. Segmental Phonology

This section describes the variable realization of consonantal and vocalic segments in the six studied varieties. In terms of consonants, these variables comprise the reflexes of Old Arabic */g/ (<ج>), */q/, the interdental series */θ, ð, ð̣/, and */ṭ/ (a subscript dot indicating the phonemic feature of “emphasis”, the phonetic quality of which has been variously described as pharyngealization, verlarization, uvularization, or dorsalization—cf. discussion in Jongman et al. 2011). Vowels examined include reflexes of the Old Arabic diphthongs */ay, aw/ and short vowels */i, a, u/. Results are summarized in Table 1.

Table 1. Segmental Phonology.

Variable	Cairo	Qift	B ̣eri	Khartoum	Šukriyya	Nigeria
*/g/	/g/	/ɟ~ d/	/j~ ḍ/	/j/ ¹	/j/	/ɟ~ j/
*/q/	/ʔ/	/g/	/g/	/g/	/g/	/g/
*/θ/	/t/	/t/	/t/	/t/	/t/	/t/
*/ð/	/d/	/d/	/d/	/d, ḍ/ ²	/d, ḍ/	/d, ḍ/ ³
*/ð̣/	/ḍ/	/ḍ/	/ḍ/	/ḍ/	/ḍ/	/ḍ/ ³
*/ṭ/	/ṭ/	/ṭ~ ṭʔ/	/ṭʔ/	/ṭ/	/ṭ/	/ḍ/
*/ay, aw/	/ē, ō/	/ē, ō/	/ē, ō/	/ē, ō/	/ē, ō/	/ē, ō/
*/i, a, u/	*/i, a, u/	*/i, a, u/	*/i, a, u/	*/i, a, u/	*/i, a, u/	*/i ~ u, a/ ³

¹ Bergman (2002). ² Hillelson (1935). ³ Owens and Hassan (2009).

Following Behnstedt and Woidich (2018, pp. 69–70) in identifying the Old Arabic articulation of <ج> as [g], rather than the received [ɟ] of the Classical Arabic tradition, we may conservatively view the /g/ realization of */g/ in Cairo as a retention. Outside Cairo, more fronted realizations are evidenced. The palatal articulation /j/ dominates in the core Sudanic region represented by the dialects of Khartoum and the Šukriyya, and is variably present in Nigeria and in B ̣eri Arabic, the southernmost of the three Egyptian varieties examined.² Realization as an alveopalatal affricate /ɟ/ is variably attested for Nigeria and Qift, and increasingly alveolar articulations /ḍ/ and /d/ are additionally observed in the B ̣eri and Qift varieties, respectively. None of these realizations, then, is ubiquitous. Palatal /j/ is perhaps of high salience, given its comparative rarity outside

this region (primarily also known from a limited number of varieties of the Arabian Peninsula—Ingham 1971; Zaborski 2007), but cannot be described as a typifying feature of the collected Egypto-Sudanic dialects as whole, or even of a substantial majority.

Perhaps of greater potential in this sense is the voiced reflex /g/ of Old Arabic */q/, robustly characteristic of all varieties in the sample outside of Cairo. Typical of all six dialects inclusive of Cairo is the merger of the Old Arabic interdental series */θ, ð, ð̣ / with corresponding alveolar stops */t, d, ḍ/. Setting apart the three varieties of the Sudanic area, however, is an additional emphatic reflex /ḍ/ of */ð/, the conditioning of which vis à vis plain /d/ is not immediately clear, but which is clearly and consistently attested across all three varieties and must logically have preceded the more general merger of */ð/ > /d/. Neither the voicing of */q/ nor the fortition of the interdentals is unique to the Egypto-Sudanic region, as these features are pervasive throughout the modern Arabic-speaking world. The coincidence of the two is perhaps more noteworthy, breaking as it does from the oft-discussed Bedouin/sedentary dichotomy which associates voiced reflexes of */q/ with the preservation of interdentals and voiceless realizations with their loss. Taine-Cheikh (2000), however, illuminates in detail the more general co-occurrence of these two isoglosses across a wide, northeast African geographic zone stretching from western Libya to points in the Sinai Peninsula and easternmost Hijaz, thereby rendering the coincidence less unusual in the Egypto-Sudanic area's immediate geographic context.

Far scarcer, but not unknown, in broader dialectological light are glottalic/glottalized realizations of */ṭ/, comprising the Nigerian implosive /ɗ/ alongside the glottalized articulation /t̚/ typical of B-əri and variably noted for Qift—all viewed similarly here for the conspicuous involvement of the glottis in the production of each (downward retraction of the glottis in /ɗ/, closure and release of the glottis in /t̚/). Similar realizations are noted outside the Egypto-Sudanic area in some Moroccan varieties as well as in scattered locations in the Levant and southern Arabia (cf. Zeroual 2006); these remain minority forms cross-dialectally, however, and may therefore be indicative of a linkage between the three specific Egypto-Sudanic dialects that display them. These realizations do not, however, serve to typify Egypto-Sudanic varieties as a whole, and neither is their status as innovation or retention—a crucial distinction in this instance—immediately clear.

In terms of vocalism, we may note the ubiquitous monophthongization of inherited diphthongs */ay, aw/ to long mid vowels /ē, ô/. This feature links all six members of the present sample, though it does not meaningfully distinguish them from neighboring dialects to the east (Kaye and Rosenhouse 1997) or immediate west (Owens 1984). Retention of all three Old Arabic short vowels has been invoked as a more distinctive regional feature in the case of Egyptian varieties (Versteegh 2014), and this generalization bears out in the current sample for all dialects save that of Nigeria, in which reflexes of Old Arabic */i, u/ are largely noncontrastive. This consistency is noteworthy in the context of widespread merger of */a, i/ to the west of the Egypto-Sudanic area and of */i, u/ to its north and east. Retention of all three vowels is not unknown outside the area, however—attested, for example, in Yemen and other portions of the Arabian Peninsula (Behnstedt and Woidich 2018)—and the nature of the feature as a common inheritance rather than a shared innovation limits its utility in support of a diachronically oriented, migration-based model of dialect classification, as will be discussed in Section 4.1 below.

3.1.2. Phonological Processes

Our review of phonological features also includes three synchronically active phonological processes: the raising of /a/ > /e/ in word-final position (often referred to as final *imāla*); the elision of unstressed /i, u/ in nonfinal open syllables following a vowel; and the shortening of phonemically long vowels in unstressed position. The incidence of these processes is summarized in Table 2 (<+> denoting the presence of a given process in each dialect and <-> its absence).

Table 2. Phonological Processes.

Variable	Cairo	Qift	B'ëri	Khartoum	Šukriyya	Nigeria
/a, ā/ > /e/ / _#	–	+	+	–	–	+
/i, u/ > Ø / VC_CV [-stress]	+	+	+	+	+	–
/V̄/ > [V] / [-stress]	+	+	+	–	–	–

A process which may be broadly described as word-final /a/-raising, affecting reflexes of both Old Arabic */a, ā/, is indicated for the B'ëri, Qift and Nigerian dialects. Cross-dialectally, processes with similar phonetic outcomes may be identified in a number of Levantine varieties, alongside looser correlates in Mesopotamia, Arabia and elsewhere in the Arabic-speaking world (cf. Levin 2007). However, significant differences in conditioning complicate the co-identification of the three Egypto-Sudanic processes as a single, shared feature, either within or without the sample: raising in Qift is reported to occur word-finally (though it would appear from Nishio's data that the rule is variably applied), B'ëri Arabic raises in pausal position, and in Nigerian Arabic /a/ is raised word-finally as triggered by the presence of a front vowel in the preceding syllable. Regardless of this feature's ultimate (dis)unity, it is not sufficiently widespread in the sample to be considered characteristic of a potential Egypto-Sudanic dialect grouping, occurring as it does in three dialects at most.

Elision of /i, u/, however, presents a different picture. Five of the six dialects in the sample display a similar form of conditioned deletion affecting the two short high vowels to the exception of their low counterpart, the core environment of which involves occurrence in a nonfinal, unstressed open syllable preceded by a vowel. In Nigerian Arabic, all short vowels, including /a/, are potentially subject to elision processes, and the conditioning environment is somewhat distinct from and more limited than that observed elsewhere in the set, requiring the presence of a preceding long vowel or sonorant (see Owens 1993a, pp. 33–36). The consistent occurrence of the elision feature across the remainder of the dialects surveyed is noteworthy, though not necessarily distinctive, as it further typifies an extensive array of additional dialects spoken across the Levant, Northwest Arabia, and elsewhere (frequently identified under Cantineau's traditional designation *parlers différentiels* for their distinct treatment of high and low short vowels under these conditions).

Finally, the shortening of unstressed long vowels is observed to occur across the sample's three Egyptian varieties, and is in fact commonly referenced as a distinctive phonological process of that area. While this generalization is borne out for Egyptian varieties by the current data, it would not seem to extend to the Sudanic contingent of the dialects examined, all three of which maintain vocalic length distinctions in both stressed and unstressed positions.

3.2. Pronominal Morphology

3.2.1. Personal Pronouns

Table 3 summarizes the independent personal pronoun paradigms for the six dialects under review. Discussion here will primarily focus on these morphologically free forms, used in subject function, though mention of their enclitic counterparts utilized in object and possessive roles is also made as relevant below. Note that the Qift forms cited ending in /a/ vary with equivalents ending in /e/ (see discussion of /a/ > /e/ raising in Section 3.1.2), and that one speaker of this variety attests a 3.pl form *humman*.

On the whole, the observations arising from comparative review of these paradigms tend toward the identification of distinct Egyptian and Sudanic norms over pan-regional unity. The first such generalization that can be made is the association of 1.pl forms lacking initial /n/ (typically viewed as an innovative) with the dialects of the Egyptian portion of the area, and forms maintaining it with those of the Sudanic portion—here considering Nigerian *anīna* an /n/-ful form, perhaps remodeled by analogy with 1.sg *ana*, and additionally recognizing the occasional occurrence of /n/-ful reflexes in Egyptian territory, as described variably for Qift. In the second place, we may also observe the distinct distri-

butions of “short” and “long” forms of the third person pronouns, with short, monosyllabic forms (e.g., Šukriyya *hū, hī, hun, hin*) typical of the Sudanic portion of the region and long, disyllabic forms (e.g., Cairo *huwwa, hiyya, humma*) typical of Egyptian territory. In the singular, both short and long forms may be considered innovations from earlier **huwa, *hiya* (which appear to be variably retained in Khartoum and Qift alongside innovative short and long forms, respectively). For the plural, typically reconstructed as **hum, *hinna*, the long masculine and short feminine forms may be seen as innovations and the short masculine and long feminine forms as retentions (Fischer and Jastrow 1980; Procházka 2014). The geographic distribution of third person patterns is somewhat complicated by the “mixed” composition of the Nigerian paradigm, presenting short forms in the singular and long forms in the plural, and the existence of both short and long variants of the B̄eri singulars, but overall the general principle of bifurcation between the two subregions—rather than commonality across them—is maintained. Also consistent with this pattern are shifts of */nt/ > /tt/ in second person forms and 3.m.pl **hum* > *hun* in the dialects of the Šukriyya and Khartoum, the latter perhaps deriving via analogy with feminine *hin*. An exception is the shift of initial */a/ > /i/ in all six dialects’ second person forms, an innovation common to the majority of modern Arabic varieties.

Table 3. Independent Personal Pronouns.

Pronoun		Cairo	Qift	B̄eri	Khartoum	Šukriyya	Nigeria	
3rd	Sg.	m.	<i>huwwa</i>	<i>huwa ~ hūwa</i>	<i>hū ~ hūwa</i>	<i>hu ~ huwa</i> ¹	<i>hū</i>	<i>hu</i>
		f.	<i>hiyya</i>	<i>hiya ~ hīya</i>	<i>hī ~ hīya</i>	<i>hi ~ hiya</i> ¹	<i>hī</i>	<i>hi</i>
	Pl.	m.	<i>humma</i>	<i>humma</i>	<i>huḡḡma</i>	<i>hum ~ hun</i> ¹	<i>hun</i>	<i>humma</i>
		f.	–	–	<i>hinna</i>	[<i>hin</i>]	<i>hin</i>	<i>hinna</i>
2nd	Sg.	m.	<i>inta</i>	<i>inta</i>	<i>inta</i>	<i>inta ~ itta</i> ¹	<i>itt</i>	<i>inta</i>
		f.	<i>inti</i>	<i>inti</i>	<i>inti</i>	<i>inti ~ itti</i> ¹	<i>itti</i>	<i>inti</i>
	Pl.	m.	<i>intu</i>	<i>intu</i>	<i>intu</i>	<i>intu ~ ittu</i> ¹	<i>ittu</i>	<i>intu</i>
		f.	–	–	<i>intan</i>	[<i>intan ~ ittan</i> ¹]	<i>ittan</i>	<i>intan</i>
1st	Sg.	<i>ana</i>	<i>ana</i>	<i>ana ~ āna</i>	<i>ana</i>	<i>ana</i>	<i>ana</i>	
	Pl.	<i>ihna</i>	<i>ihna ~ naḡna</i>	<i>ihna</i>	<i>niḡna</i>	<i>niḡna</i>	<i>aniḡna</i>	

¹ Bergman (2002) [transcription of final vowel length regularized for comparability].

The loss of the masculine/feminine distinction in second and third person plurals, with consequent generalization of the inherited masculine form, may be noted in Cairo and Qift and would appear to be currently progressing in the dialect of Khartoum, where distinctive feminine plural forms appear obsolescent and are sociolinguistically associated with rurality (Dickins 2007b, p. 561). Indeed, given the relative population structures of the speech communities under discussion, such a rural/urban dichotomy may underlie the distribution of this feature in the current sampling more meaningfully than would the geographic divide described in relation to the previous two, though it should not escape notice that the two geographically atypical cases, Khartoum and B̄eri, represent the northernmost Sudanic and southernmost Egyptian varieties sampled, respectively. In either case, the picture is once again one of heterogeneity rather than conformity of personal pronoun systems across the Egypto-Sudanic zone.

Turning briefly to the bound personal pronoun forms, not presented in Table 3, three distinctive features are observed. The unusual retention of 2.sg.f *-ki* and the innovation of 3.m.sg *-a* (< **-hu*) serve to bind Nigerian and B̄eri, while the remaining varieties instead attest the innovative forms *-ik* and *-u* (though Qift *-o*) near-ubiquitous in modern Arabic, likely as pre-diasporic developments (cf. Behnstedt and Woidich 2005; Owens 2006). Thirdly, the three Sudanic dialects of Nigeria, Khartoum and (more marginally) the Šukriyya share with one another the variable loss of initial /h/ in third person bound forms, and display similar interactions of this feature with stress assignment.

3.2.2. Demonstrative Pronouns

The proximal and distal demonstrative pronoun series of the six dialects examined are presented in Table 4. This presentation summarizes a highly diverse array of available data, particularly as pertains to the three Egyptian varieties. Intra-dialectally varying forms deemed to represent progressive degrees of reduction from a common source etymon have been simplified with a single representation here, and those displaying singular sporadic phonetic developments or synchronically predictable pausal realizations are likewise not shown; for a full accounting of all variants, relevant to this analysis and otherwise, see Woidich (2006b, pp. 44–46, 303), Nishio (1995, p. 190), and Bergman (2002, p. 43). When a sole plural form is indicated, its use comprises both masculine and feminine values, its position within the table selected on the basis of cognacy.

Table 4. Demonstrative Pronouns.

Pronoun		Cairo	Qift	B ēri	Khartoum	Šukriyya	Nigeria	
Near	Sg.	m.	<i>da ~ dawwa</i>	<i>de</i>	<i>da</i>	<i>da</i>	<i>da</i>	
		f.	<i>di ~ diyya</i>	<i>dey ~ diye</i>	<i>di</i>	<i>di</i>	<i>di</i>	
	Pl.	m.	<i>dōl(a)</i>	<i>dōl ~ dowal</i>	<i>dōl(a)</i>	[<i>dōl</i>] ¹	<i>dōl</i>	<i>dōl(a)</i>
		f.	–	<i>dōle</i>	<i>dēl(a)</i>	<i>dēl</i>	<i>dēl</i>	<i>dēl(a)</i>
Far	Sg.	m.	<i>dukha</i>	<i>dakka ~ dāk</i>	<i>dukkāti</i>	<i>dāk</i>	<i>dāk</i>	<i>qāka</i>
		f.	<i>dikha</i>	<i>dikke</i>	<i>dikkīti</i>	<i>dīk</i>	<i>dīk</i>	<i>qīke</i>
	Pl.	m.	<i>dukham</i>	<i>dokkum</i>	<i>dukkumma</i>	[<i>dōlak</i>] ¹	<i>dōlak</i>	<i>qōlak(a)</i>
		f.	–	–	<i>dikkinma</i>	<i>dēlak ~ dēk</i>	<i>dēlak</i>	<i>qēlak(a)</i>

¹ Bergman (2002).

As a point of departure for analysis, it seems likely that the full array of forms presented here (with the possible exception of the distal plurals, as discussed below) ultimately originates in a paradigm similar to that attested for the Šukriyya variety in Table 4. Relevant features at a broad level of Arabic demonstrative classification involve the leveling of initial /d/ (<*/ð/) across all members of the paradigm, the absence of a reflex of the Old Arabic presentative particle *hā-, and the use of vowel alternation to indicate gender distinction in both the singular and plural while, for the most part, simultaneously maintaining consonantal marking of plurality (for further discussion of these traits in cross-dialectal context, see Magidow 2013). Taken individually, none of these characteristics is restricted to the Egypto-Sudanic area; their converging incidence, however, largely is, identifiable elsewhere in comparable fashion only at scattered points in southwestern Arabia and possibly the central Levant (Magidow 2016). The essential tenets of this shared basic paradigm, then, together rise as a potentially significant piece of linguistic evidence supporting the common classification of the dialects of the Egypto-Sudanic area.

At the same time, substantial secondary divergences parallel the north–south splits between Egyptian and Sudanic varieties already observed in relation to several personal pronoun forms. Primary among these is the rise of forms etymologically comprising a demonstrative element of the type witnessed above supplemented by the incorporation of a following independent personal pronoun. With the exception of the Qift variant *dāk*, these forms have entirely supplanted the presumably unsupplemented original distals in the three Egyptian varieties (e.g., Cairo *dukha* < *dāk huwwa, *dikha* < *dīk hiyya, etc.), and occur variably in the proximal series of the two northernmost Egyptian varieties as well (e.g., Qift *dīye* < *di hīye). Magidow (2013, p. 400) has previously proposed that these composite forms evolved from an original presentative structure of the same composition, on the basis of like presentatives in use in other dialects, such as Ḥassāniyya. This assessment is support by the presence of presentatives of this type closer to home in the current sample, in the form of Nigerian *dawa* < *da huwa ‘here he is ...’, *qākwa* < *dāk huwa ‘there he is ...’, etc. (cf. Cairo *dawwa* ‘this (m.sg)’, *dukha* ‘that (m.sg)’). This development—to which Nigerian shares the precursor, but in which it does not participate—thus serves to differentiate the three Egyptian dialects of the sample from their Sudanic counterparts.

Following from analysis of Egyptian distals in this manner is the further insight that the Cairo, Qift and B'ēri paradigms may display a distinctive, vowel-alternating mode of plural formation. In contrast to the distal plurals of the three Sudanic varieties, which are transparently formed via the addition of the distal morph *-k* to the existing proximal plural, the Egyptian forms do not contain any visible reflex of the proximal plural's distinctive /l/—instead, we encounter a plural marking back vowel of the type *dukham*, *dokkum*, *dukkumma*. It is certainly possible that an original /l/ of the plural form has simply vocalized, or that, given the occurrence of /u/ in some Egyptian m.sg forms (e.g., B'ēri *dukkāti*), these plurals contain reflexes of a generalized singular *dāk. It is also plausible, however, to connect the vowel-alternating inflection attested in, e.g., Qift *dakka*, *dikke*, *dokkum*, to that known from a number of North African varieties, as in Ḥassāniyya *ḍāk* (m.sg), *ḍūk* (f.sg), *ḍūk* (pl.) (Taine-Cheikh 2007). If this development is indeed reflected in the Egyptian forms, it would mirror that encountered in the Khartoum variant *dēk*, which may in turn have a counterpart in the initial element of B'ēri f.pl *dikkinna*. Such plurals are not the norm in the three Sudanic varieties, however, which instead maintain the /l/-marked plural intact (including in the Nigerian presentative set perhaps cognate to the Egyptian distals, m.pl *ḍolakkahumma*, f.pl *ḍelakkahinna*), thereby presenting a further potential north–south distinguishing feature among the dialects examined.

The final secondary development of note in relation to the demonstrative pronouns is the occasional loss of gender distinction in the plural, accompanied by the generalization of a single plural form to encompass both gender values. In the Egyptian varieties that have lost their original gender distinction, an original masculine form has generalized, whereas in Khartoum, when gender distinctions are lost, it is an original feminine form that has done so (compare Šukriyya m.pl *dōl*, f.pl *dēl* with Cairo c.pl *dōl(a)*, Khartoum c.pl *dēl*). Qift, with m.pl *dōl*, f.pl *dōla* ~ *dōle*, would seem to have initially followed Cairo in generalizing a masculine form, but subsequently reallocated originally variable *dōl* ~ *dōla* to distinct gender values (perhaps via analogy with the f.sg nominal marker *-a*). Given the differentiated pathways taken in the generalization of formerly gendered forms in the dialects of Cairo and Qift, on the one hand, and Khartoum, on the other, it is perhaps advisable to view these two developments as parallel yet independent.

3.2.3. Relative Pronoun

Rather than unifying the Egypto-Sudanic zone, relative pronoun forms further perpetuate the previously witnessed divide between the three Egyptian varieties of Cairo, Qift and il-Bi-erāt and the three Sudanic ones of Nigeria, Khartoum and the Šukriyya. The former set all display the identical relative form *illi*, reflecting a development near-ubiquitous across modern Arabic varieties (Vicente 2009). The dialects of the Sudanic area, on the other hand, all present the identical form *al-*, which has for all intents and purposes functionally merged with the definite article (Dickins 2009). This latter development is far less common in comparative scope, but is also apparent in a small number of dialects of the northern Fertile Crescent area (Vicente 2009).

3.2.4. Interrogative Pronouns

Table 5 summarizes the interrogative pronouns ‘who?’, ‘what?’ and ‘which?’ attested for the six dialects under investigation. Cautiously excluded here are forms transparently mirroring Classical Arabic *ʔayy* ‘which?’ noted for Cairo and Nigeria, on the grounds that this is frequently identified in the modern Arabophone world as a diglossic import, not indicative of these varieties’ inter-dialectal relationships but rather of their individual connections to a shared acrolect (cf. Woidich 2006b, p. 35). In the case of Qift, such a form is the only one given for ‘which?’ by Nishio (1995); interpretation of this is fact discussed below.

Table 5. Interrogative Pronouns.

Pronoun	Cairo	Qift	B ēri	Khartoum	Šukriyya	Nigeria
‘who?’	<i>mīn</i>	<i>mīn</i>	<i>mīn</i>	<i>minu</i> ¹	<i>minū</i> ¹ , <i>min</i>	<i>mine</i>
‘what?’	<i>ē(h)</i>	<i>ē</i>	<i>ē(h)</i>	<i>šinu</i> ¹	<i>šinū</i> ¹ , <i>šin</i>	<i>šinu</i> ~ <i>šumu</i>
‘which?’	<i>anhi</i> ¹ ~ <i>āni</i>	[<i>ayy</i>]	<i>innhi</i> ¹	<i>yātu</i> ¹	<i>yātū</i> ¹	<i>yēnu</i> ¹ ~ <i>yatu</i> ¹

¹ Demonstrates agreement phenomena.

Replicating the geographic patterning now familiar from other aspects of the pronominal system, forms for ‘who?’ in the Egyptian portion of the Egypto-Sudanic region are unified in displaying an innovative, sporadic long vowel /ī/. This long vowel is not present in any of the three more southerly varieties, though these likewise agree with one another in the inclusion of an original personal pronoun, incorporated alongside inherited *min as a marker of gender and number agreement. Such inflectional behavior is maintained in its full form in the dialect of the Šukriyya (m.sg *minū*, f.sg *minī*, m.pl *minun*, f.pl *minin*), alongside an uninflecting form *min* (the distinct syntactic behavior of which is treated below in Section 3.4.4). Inflecting forms are noted as well in older descriptions of the speech of Khartoum (Hillelson 1935), though modern sources (Dickins 2007b; Bergman 2002) indicate that an invariant (originally m.sg) *minu* at least alternates with these, if it has not replaced them entirely. The latter outcome would seem to have been the case for Nigerian *mine*, which does not inflect for number or gender but appears to display the reflex of an earlier incorporated pronoun. As isoglosses, both short and long vocalic reflexes, as well as personal pronoun incorporation, are well known outside the Egypto-Sudanic region.

Forms for ‘what?’ follow a similar north–south division: those of the three Egyptian dialects feature a reflex of earlier *ēš < *ʔayy šayʔ ‘which thing?’, while those of the Sudanic varieties seem to ultimately reflect a version of a similar etymological source phrase with the inclusion of nunation: *šin < *ʔayy šayʔin ‘which thing?’. As observed for the Sudanic ‘who?’ forms, ‘what?’ forms of this area also display the incorporation of personal pronouns, with comparable patterns of productivity in agreement inflection to those described just above. Though neither the nunated nor the non-nunated derivation serves to unify the study area, both are widespread in modern Arabic more broadly.

Pronouns meaning ‘which?’ may additionally be distinguished into northern and southern blocks within the Egypto-Sudanic zone, though along a slightly different boundary. Complicating evaluation in the context of the present study is the recording of a single form *ayy* for Qift, which, as has been noted, likely represents a borrowing from Classical Arabic *ʔayy* (more transparently so in the case of Cairo *ʔayy*, which displays an initial glottal stop regularly lost in the variety). It is probable that Qift also includes (or included, until recently) a form cognate with Cairo *anhi*, B ēri *innhi*, perhaps similar to the *ʔinhi* reported for nearby ʔzbat al-Būša (Khalafallah 1969). Regardless, it would appear that forms of this type, reflecting Old Arabic *ʔayyun (or Aramaic *aynā*) combined with an etymological personal pronoun, are typical of the Egyptian portion of the area, and likely also include the Nigerian variant *yēnu*. The southern dialects, including Nigerian via its variant *yatu*, are instead distinguished by reflexes of earlier *ʔayyat (plus incorporated pronoun). Products of both etymologies inflect for agreement when following a modified noun (e.g., B ēri m.sg *innhū*, f.sg *innhī*, pl. *innhumma*, Khartoum m.sg *yātu*, f.sg *yāti*, pl. *yātum*), but occur invariantly when preceding one—the Sudanic varieties fixing an original masculine singular form in this usage, the Egyptian ones more often an original feminine. Forms of the *ʔayyun type are well known beyond the confines of the Egypto-Sudanic region. Reflexes of *ʔayyat are much more unusual, known elsewhere only from a few locations in western (and especially northwestern) Arabia (cf. Reichmuth 1983, p. 118).

3.3. Verbal Inflectional Morphology

3.3.1. Agreement Inflection

Table 6 summarizes the major distinctive elements of verbal agreement inflection across the Egypto-Sudanic varieties surveyed. The feature “f.pl” refers to the presence of

distinct masculine and feminine agreement morphemes in the second and third person plurals of all conjugation paradigms; those dialects that do not display this feature have generalized inherited m.pl forms across both contexts. The remaining features relate specifically to either the perfect or the imperfect conjugation of Form I sound verbs, as indicated in the table.

Table 6. Verbal Agreement Inflection.

Pronoun	Cairo	Qift	B'ēri	Khartoum	Šukriyya	Nigeria
f.pl	–	–	+	[+]	+	+
Perfect						
1.sg	-t	-t	-t	-ta	-t ~ '_(-t)	'_(-t)
3.f.sg	-it	-at	-at	-at	-at	-at
3.m.pl	-u, -ū-	-u ~ -ow, -ū- ¹	-aw, -ō-	-u, -ō- ²	-u, -ō-	-o, -ō-
Imperfect						
Prefix V	Ci-	Ci-	Ci-, Ca-	Ca-	Ca-	Ca- ~ Ci-
1.sg	a-	a-	n-	a-	a-	(b)a- ~ n-
1.pl	n-	n-	n...-aw	n-	n-	n ~ n...-u

¹ Behnstedt and Woidich (1985–1999, Map 207). ² Bergman (2002).

Parallel to the pronominal development described in Section 3.2.1, above, the dialects of Cairo and Qift do not retain a gender distinction in plural agreement morphology, and this distinction appears to be fading from use in the dialect of Khartoum. This development may thus be seen as a source of differentiation within the six varieties sampled, perhaps reflecting a rough north–south geographic divide, perhaps on the basis of difference between urban and rural populations. All dialects which distinguish the feminine plural do so in a formally identical manner, via use of a suffix *-an* (3.f.pl)/*-tan* (2.f.pl).

In assessing agreement markers of the perfect conjugation, features distinctive of this dataset in the pan-Arabic view include the conjugation of the 1.sg (identical in all cases to the 2.m.sg), the 3.f.sg, and the 3.m.pl. The 1.sg forms of the three dialects of the Egyptian area show the expected *-t* (< **-tu*) typical of the great majority of modern Arabic varieties. Among the three dialects of the Sudanic area, however, we view a pair of innovative local developments. In Khartoum, the 1.sg agreement value is marked with the suffix *-ta*, the /a/ of which likely represents the morphologized product of a former paragogic vowel, following an earlier development **-tu* > **-t* (similar to the 3.m.sg *-a* of geminated verbs in the same dialect). In Nigeria, we witness the loss of earlier 1.sg *-t* and consequent rise of contrastive stress distinguishing 1.sg *ka'tab* from 3.m.sg *katab*; the original *-t* resurfaces prevocally, as when preceding a bound object suffix or occasionally in connected speech. The same inflectional pattern is recorded among the Šukriyya, there alternating with more standard *-t*. The 3.f.sg suffix is *-at* in all dialects save that of Cairo, where it is *-it*. The Cairene reflex is innovative; retention of inherited *-at* thus typifies the rest of the group, though as a feature it does not serve to differentiate these dialects from neighboring varieties of Libya or the Arabian Peninsula.

South of Cairo, one encounters lowered realizations of the inherited 3.m.pl suffix **-ū*. These begin marginally in Qift, in a minority variant *-ow* of more general *-u*. In B'ēri, this suffix is *-aw*, with an allomorph *-ō-* in nonfinal position (i.e., when followed by an additional suffix), and Nigerian Arabic likewise appears to show lowered realizations, *-o* and *-ō-*, in both conditions—though definitive interpretation of the Nigerian data is potentially confounded by the influence of vowel harmony (Owens 1993a, p. 105). In the dialects of Khartoum and the Šukriyya, a lowered realization only emerges as a nonfinal allomorph, contrasting final *-u* with nonfinal *-ō-*. The universally lowered reflexes identified in B'ēri and Nigerian are a traditionally acknowledged “Bedouin” feature characteristic of a wide array of Arabic varieties from North Africa to the Arabian Peninsula to Mesopotamia. The conditioned lowering exemplified in the speech of Khartoum and the Šukriyya is of far more limited distribution, though it does also occur in the dialect of Mecca and the Jewish communolect of Baghdad (Reichmuth 1983, p. 28). Were these two types of

lowering to be identified as a single dialectal feature, then they would serve as an additional southerly isogloss linking the three Sudanic varieties, as well as the southernmost Egyptian variety—however, it is not clear that it is warranted to overlook the potentially significant allomorphic differences between the two.

Turning to the imperfect conjugation, the six dialects do not pattern uniformly with regard to the quality of the vowel utilized in the formation of imperfect agreement prefixes with Form I sound verbs. As noted in relation to several previous features, the general shape of the distribution would seem to be one of a north–south divide: the varieties of Cairo and Qift, aligning with the majority of modern Arabic varieties, show a prefix vowel /i/, while those of Khartoum and the Šukriyya have /a/. The dialect of Nigeria offers variation on this count, speakers utilizing both /i/ and /a/ reflexes. In B'ēri, the prefix vowel shows harmony with the theme vowel of the inflected verb, thus manifesting as /i/ or /a/ in predictable fashion. All of these patterns may be considered innovative in relation to the oldest reconstructable state of this variable in Arabic, which has been proposed to consist of alternation between /a/ and /i/ in inverse relation to the height of the imperfect theme vowel, in accordance with the Barth-Ginsberg Law (Bloch 1967; Pat-El 2017). Thus, salient isoglosses within the Epyto-Sudanic area, none of these developments are confined to the zone: generalization of /a/ is known in Western Arabia and the Yemeni Tihama, while that of /i/ dominates elsewhere, and harmonization of the B'ēri type is also known in North Africa (cf. Behnstedt and Woidich 2005, pp. 12–13).

The innovative first person agreement marking scheme 1.sg *n-*/1.pl *n-...-u*, typical of North Africa west of the Epyto-Sudanic area, also appears in our data as an inflectional norm in B'ēri and as an available variant in Nigerian. Though its presence is of dialectological note, this feature does little to clarify broader understandings of a potential macro-level Epyto-Sudanic dialect classification, as heterogeneity on this point is already well established in both the Egyptian and Sudanic portions of the area. For excellent discussion of this development's history and distribution in the region, see Owens (2003).

3.3.2. Tense, Aspect, Mood and Voice Inflection

Beyond agreement, Table 7 summarizes additional verbal inflectional morphology utilized in the expression of tense, aspect, mood and voice. The prefix of the imperative mood is provided first, followed by the passive marker. Next, an array of “preverbal” modifiers are included which indicate a complex (and often varying) set of tense, aspect and mood values, details of which will be explicated as part of the following discussion.

Table 7. Tense, Aspect, Mood and Voice Inflection.

Pronoun	Cairo	Qift	B'ēri	Khartoum	Šukriyya	Nigeria
Imperative Prefix	<i>i-</i>	<i>i-</i>	<i>i-</i>	<i>a-</i> ¹	<i>a-</i>	<i>a-</i>
Passive Prefix	<i>it-</i>	<i>it-</i>	<i>it-</i>	<i>it-</i>	<i>in-</i>	<i>an-</i>
TAM modifiers						
*bi-	<i>bi-</i>	[<i>ba-</i> ²]	–	<i>bi-</i>	<i>bi-</i>	<i>b-</i>
*ʕammāl	ʕammāl	–	ʕa- ~ ʕama-	–	–	–
*rāh	<i>ha- ~ ha-</i>	<i>ħa-</i>	<i>raħ ~ raħa ~ ha-</i>	<i>ħa-</i>	–	–

¹ Bergman (2002). ² Behnstedt and Woidich (1985–1999, Map 221).

Vowel qualities of the imperative prefix display a similar north–south differentiation to that previously noted for the prefix vowel of the imperfective, and in synchronic terms these two traits are likely not systemically independent; total convergence of this kind is likely best interpreted as innovative in each case, the probable product of analogy (cf. Bar-Asher 2008). A prefix *i-* is thus encountered in the three dialects of Egypt, while the form *a-* is found in the three varieties of the Sudanic portion of the region. The latter is known outside the area in the same limited distribution described for the *Ca-* prefix vowel (Section 3.3.1), while the former occurs in modern Arabic more widely.

The passive morpheme splits the area latitudinally in a similar manner, though in this instance Khartoum, the northernmost variety of the Sudan, is seen to pattern with the body of Egyptian varieties in displaying *-it*. The dialects of Nigeria and the Šukriyya, on the other hand, share in presenting /n/-based forms. Both features are generally considered to be innovations on the Old Arabic type, and each shares a wide distribution in the modern Arabophone world more broadly.

The first TAM modifier to be discussed, (shallowly) reconstructable to *bi-, is the most widely distributed in the present sampling with detectable reflexes in all but one of the six varieties—not being recorded for B'ëri. In Nigerian, this morpheme has partly been subsumed into the person marking system, occurring as a quasi-fixed component of originally vowel-initial agreement prefixes of the imperfect conjugation; elements of productive use do remain, though their precise functions in Nigerian remain far from clear (see discussion in Owens 1993a, pp. 106–10). Values of *bi-* in Khartoum and among the Šukriyya more plainly include continuous (ongoing, repetitive, habitual) aspect, and futurity. Cairene *bi-* echoes the former of these, though notably not the latter, and adds a meaning of general realis or indicative mood (Brustad 2000, pp. 246–47). Little information on preverbal modifiers is provided as part of Nishio's descriptive materials for Qift. Behnstedt and Woidich's (1985–1999) immediately neighboring sample point of il-Barāhma, however, attests a "Verbmodifikator Präsens" *ba-*, which, given its treatment in the atlas, likely expresses semantics similar to those of Cairo *bi-*. While the functions of these various items are thus differentiated across the dialects examined, their simple exponence as a feature does unite the greater part of the area. Reflexes of innovative *bi- are, of course, well known outside the Egypto-Sudanic zone as well—most especially in the Levant, the Arabian Peninsula, and Libya. Semantically, the functional range described for the core Sudanic varieties is the more typical cross-dialectally, which in some cases leans even more heavily toward future and volitional readings.

An additional continuous aspect marker is found in Egypt, reconstructable to *ʕammāl. This item is reflected in Cairene *ʕammāl*, the meanings of which are far more narrowly defined than those of *bi-* and express a notion of intensity, iterativity, and repetition. B'ëri *ʕa- ~ ʕama-* is of more generalized usage, and is reported to carry functions largely comparable to those filled by reflexes of *bi- in other dialects of the sample. Though *ʕammāl is not entirely absent in Sudanic territory (cf. Hillelson 1935), it is a definitive rarity there, and does not occur in the three Sudanic dialects sampled. It is elsewhere known outside Egypt from the Levant and scattered points in southwest Arabia. Future tense markers reflecting *rāh are attested in all three Egyptian dialects and in Khartoum. In the latter location, *ḥa-* is reported by both Dickins (2007b, p. 569) and Bergman (2002, p. 38) as a recent Egyptianism. In light of its absence in the other Sudanic varieties sampled, and the lack of a clear dialect-internal grammaticalization chain (*maša* largely outcompeting lexical *rāh* as the general term for 'go' in Sudanic dialects), this attribution is likely correct. Regardless of its ultimate originality in Khartoum, this innovative feature serves to differentiate the two southernmost varieties of the sample from the four northernmost, which join a wide array of modern dialects to attest products of this development from Algeria to Mesopotamia (cf. Leddy-Cecere 2020).

3.4. Syntax

The following subsections address a selection of syntactic features relevant in the evaluation of a potential Egypto-Sudanic dialectal unity, namely negation strategies, analytic genitive structures, the ordering of adnominal demonstratives, and WH-question formation. Though all elements of the ensuing discussion will likely be familiar to Arabic dialectologists, it bears note that, while syntactic variation of these types has frequently been addressed in formal (Aoun et al. 2010), comparative (Brustad 2000) and diachronic light (Wilmsen 2014), it has less often been the stuff of broad-based efforts toward top-tier Arabic dialect classification. In the Egypto-Sudanic case specifically, however, shared syntactic features—particularly the latter two considered in this section—are consistently among

the few concrete pieces of linguistic evidence invoked in support of the identification of a unified subgroup (cf. Versteegh 2014, p. 209); as such, they merit a full treatment here.

3.4.1. Negation

Rather than uniting the six Egypto-Sudanic dialects surveyed, negation strategies are seen once again to divide the region into northern and southern camps, reminiscent of geographic patterns previously established in relation to numerous phonological and morphological variables already considered. The three Egyptian varieties surveyed display a “split” negation system typical of both modern and historical forms of Arabic, whereby two distinct strategies exist for the negation of verbal and nonverbal predicates. In all three dialects, the first of these involves a discontinuous negation structure, the second a unitary particle deriving diachronically from a negated third person singular pronoun (having since shed such morphological specification). The following examples from Cairene, showing verbal *ma ...-š* and nonverbal *miš* (~*muš*), are typical:

1. *ma katab-š*
NEG wrote.3MSG-NEG
'He did not write.'
(Cairo: Woidich 2006b, p. 335)
2. *da miš kwayyis*
that NEG good
'That is not good.'
(Cairo: Woidich 2006b, p. 334)

Equivalent markers in Qift and B'ēri are *ma ...-š/muš* and *ma ...-(i)š/miš*. B'ēri stands out for allowing at least a limited application of verbal *ma ...-(i)š* to nonverbal predicates (e.g., *ma zēn-iš* 'not good', Woidich 2006a, p. 303) alongside more standard *miš*, although potential pragmatic specificities of such usage remain undescribed (cf. discussion in Brustad 2000, pp. 291–94).

In the three Sudanic varieties of the sample, by contrast, no such verbal/nonverbal distinction in negation strategies exists, and both predicate types are negated by a unitary operator with no discontinuous element. Consider, from Khartoum:

3. *ma btafham ṣarabi*
NEG understand.2MSG Arabic
'You don't understand Arabic.'
(Khartoum: Dickins 2007b, p. 570)
4. *inta ma kwēyis*
you NEG good
'You are not nice.'
(Khartoum: Dickins 2007b, p. 570)

Negator *ma ~ mā* is used similarly in the dialects of the Šukriyya and Nigeria. In the latter variety, alongside *ma* we also encounter a generalized negator *mi* grammaticalized from an earlier negated third person pronoun and used in nonverbal negation, thus analogous in origin to Egyptian *miš ~ muš* but with no sign of an original discontinuous element -š. In sum, then, we find negation dividing the sampled dialects into Egyptian and Sudanic camps on two fronts. In the first place, the three Egyptian varieties are defined by the presence of discontinuous negation, and the three Sudanic varieties by its absence; in the second, the Egyptian dialects characteristically comprise distinct verbal and nonverbal negation strategies while the Sudanic dialects do not—B'ēri and Nigerian each demonstrating a degree of variable “slippage” from these otherwise generalizable norms. On the first count, the innovative Egyptian trait is broadly typical of dialects of the Arabic-speaking West, the more conservative Sudanic one those of the East. On the second count, it is Sudanic which stands out against the general backdrop of modern Arabic in utilizing a single strategy for the unmarked negation of both verbal and nonverbal predicates, though such may in fact represent a retention of inherited properties of Old Arabic *mā* (cf. discussion in Brustad 2000, pp. 277–83; Ouhalla 2008).

A third negation type, that of a negated personal pronoun paradigm fulfilling what has often been described as a negative copular function, is also in evidence in dialects of the Egypto-Sudanic area; however, a paucity of coverage in descriptive sources renders

a comprehensive evaluation here impossible. On the basis of those dialects for which sufficient data are available (those of Cairo, Nigeria and the Šukriyya), it seems likely that a north–south split of the dimensions already described characterizes treatment of this negation strategy as well. This would be true both in terms of the pragmatic markedness of such usage (largely unmarked in the two Sudanic varieties, while in Cairene indicating the negation of a presupposition) and in terms case assignment (the negative structure generally triggering accompanying accusative pronouns in the two Sudanic varieties but nominative ones in Cairene, e.g., Šukriyya *māk*, Cairo *mantāš* ‘you (m.sg) are not’). Though thus not inconsistent with the geographic division outlined in relation to the better known strategies, more definitive analysis of this third negation type awaits further descriptive information.

3.4.2. Analytic Genitive

All six dialects of the Egypto-Sudanic area examined present use of an analytic genitive structure alongside the inherited Old Arabic synthetic (juxtaposed) genitive. Such structures as a general scheme are a widespread innovation in modern Arabic, though individual forms and properties vary widely from dialect to dialect (Behnstedt and Woidich 2005; Eksell Harning 1980). The essential components of the construction are a possessum, which governs a following genitive exponent, which in turn governs a following (nominal or pronominal) possessor, on the model of the following:

5. *al-kutub haġgat al-madrasa*
 DEF-books GEN DEF-school
 ‘the books of the school’ (Khartoum: Dickins 2007b, p. 570)
6. *ik-karafaġtāt bitāš-ik*
 DEF-ties GEN-you.FSG
 ‘your ties’ (Cairo: Woidich 2006b, p. 59)

Beyond the existence of the general schema, which all six dialects attest, the overall picture of analytic genitive structures across the varieties sampled is one of both formal and functional diversity. In the first place, a wide array of different exponents occur, of diverse etymology. The most widely spread are those reconstructable to *bitāš, ultimately < *matāš ‘property’. Reflexes of the latter are distributed broadly from Morocco to the southern Levant, but known with the sporadic mutation of initial */m/ > /b/ in the eastern portion of this region only (Egypt and the Sudan, alongside some Levantine attestations). Such forms are instantiated in Cairo *bitāš* and Qift *bitāš* ~ *ibtāš*, the sole genitive exponents reported at these locations, and in variation with products of other etymologies in the dialects of il-Bīerāt (*ibtāš*), Khartoum (*bitāš*), and (more marginally) the Šukriyya (*bitāš* ~ *butāš*), thus leaving Nigerian the sole dialect sampled not to attest a reflex. There, the genitive exponent is instead *hana* < *hana ‘thing’, which is also reflected in B’ēri *ihnīn*, and encountered outside the region in the interior northern Levant. Alongside *bitāš*, Khartoum sports an exponent *haġg* < *haqq ‘property, right’, well known from dialects of the Arabian Peninsula, which is also attested as a marginal variant *haġg* among the Šukriyya. The primary exponent in this last variety is *hīl* (likely < *hū li- ‘it (3.m.sg) [is] for’), which may also be reflected in its f.sg guise *hīl* as a suppletive variant member of the Nigerian *hana* paradigm: m.sg *hana*, f.sg *hīl* ~ *hīnta*. Šukriyya further attests yet another variant *allīl* (< *allī li- ‘which [is] for’), also known from dialects of southern Egypt not included in this sample (Behnstedt and Woidich 2005).³ Thus, the picture which emerges is one in which reflexes of *bitāš are typical and (to a degree) distinctive of the bulk of the varieties surveyed, though not to the exclusion of other exponents in use in the same speech communities. Meanwhile, reflexes of *hana, *hūl and *haqq serve to unite pairs of dialects within the sample, but do not broadly typify the set as a whole.

As far as syntactic behavior and semantic functions are concerned, the information provided by descriptive sources is uneven, but the following generalizations may be made. In all cases save that of Qift, for which Nishio does not specify, exponents are observed to agree in gender and number with their governing possessum (e.g., B’ēri m.sg *ihnīn*,

f.sg *ihnīt*, m.pl *ihniyyīn*, f.pl *ihniyyāt*); such inflection in fact reveals underlying variation in the use of *bitāʿ*, which inflects for m.pl as *bitāʿim/ibtāʿim* in the Khartoum, B'eri and Šukriyya dialects, but as *bitūʿ* in Cairo. These agreement properties are not unique to the Egypto-Sudanic area, though nor are they universal cross-dialectally. Cairene *bitāʿ* constructions have been demonstrated to show a strong dispreference for the governing of indefinite and/or nonspecific possessors, outside of an idiomatic meaning of 'one who likes ...' (Brustad 2000, pp. 80–82); information is lacking for Qift and B'eri, but Dickins (2007b) description of Khartoum *bitāʿ* and *hagg* would seem to indicate a similar state of affairs. In the Šukriyya and Nigerian dialects, however, such uses are noted, in Nigerian even extending as far as fully nonreferential classificatory function:

7. *nāga hīl šukriyya*
 camel GEN Šukriyya
 'a Šukriyya woman's camel' (Šukriyya: Reichmuth 1983, p. 112)
8. *maraba hīl lēs*
 cloth GEN lace
 'lace cloth' (Nigeria: Owens 1993a, p. 64)

Examples like these indicate a clear heterogeneity of analytic genitive functional properties within the Egypto-Sudanic zone, and mirror potential correlates in dialects as far-flung as Morocco and Kuwait (cf. Brustad 2000). Though a geographic, social, or other ordering may ultimately underlie these patterns, information is insufficient to offer such a determination at the present time.

3.4.3. Adnominal Demonstrative Order

The etymological form and paradigmatic organization of demonstratives has been described above (Section 3.2.2) as a potentially strong instance of innovative uniformity across members of a proposed Egypto-Sudanic dialect classification. In addition to these commonalities noted in the morphological dimension, the syntactic properties of demonstratives in adnominal usage also display distinctive and uniform characteristics across dialects of this region—a fact which has arisen in the Arabist literature as one of a small number of concrete linguistic traits identified as definitive of a macro-level Egypto-Sudanic grouping. Specifically, demonstratives in all six dialects sampled occur post-nominally, as in (9) and (10), thus opposed to the typical Arabic pre-nominal pattern exemplified by Moroccan in (11):

9. *ar-rājil da*
 DEF-man DEM.MSG
 'this man' (Nigeria: Owens 1993a, p. 45)
10. *ir-rijjal dowal*
 DEF-men DEM.MPL
 'these men' (Qift: Nishio 1995, p. 190)
11. *hād n-nās*
 DEM DEF-people
 'these people' (Moroccan: Brustad 2000, p. 118)

While available as a pragmatically marked alternative to the pre-nominal position in many dialects, as well older forms of Arabic, utilization of the post-nominal structure as an unmarked norm, without a genuinely productive pre-nominal counterpart, is highly unusual cross-dialectally and virtually restricted to the Egypto-Sudanic area (Brustad 2000; Vicente 2006). Within the area, minor but potentially significant exceptions in the form of rhetorically/stylistically specified usages and fixed expressions with pre-nominal ordering may be noted for—at least—the dialects of Cairo and the Šukriyya; the implications of these will be considered in Section 4.1. Irrespective of this fact, post-nominal demonstrative order in its present incarnation does appear to present a key point of unity across dialects of the Egypto-Sudanic zone, and a key point of distinction between these and the collective body of Arabic varieties spoken elsewhere.

3.4.4. WH-Movement

Alongside post-nominal demonstrative position, Versteegh notes for “Egyptian Arabic ... as well as ... the related Sudanese dialects” an additional conspicuous syntactic trait: the nonfronting of WH-elements in content questions (Versteegh 2014, p. 209). Such in situ question formation is not typical of Arabic, in which the fronting of interrogative elements, whether accompanied by resumption or gapping, is more usually the unmarked norm (Aoun et al. 2010). Retaining a degree of cautious agnosticism regarding Qift, the descriptive source for which does not provide sentence level examples, in situ question formation is attested across the full set of Egypto-Sudanic varieties sampled. Representative instantiations are provided in (12) and (13), accompanied by a WH-fronted sentence from Lebanese Arabic in (14) for comparison:

12. *ʕawz min-ha ʔe?*
want.MSG from-it what
'What do you want from it?' (B'eri: Woidich 1974, p. 54)
13. *ʃit maʕa minu?*
came.2MSG with who
'Who did you come with?' (Khartoum: Dickins 2007b, p. 562)
14. *ʃu badd-ak tɔ-ʔra b-l-ʔawwal?*
what want-2MSG 2-read in-DEF-first
'What do you want to read first?' (Lebanese: Aoun et al. 2010, p. 140)

In addition to this pattern, Šukriyya departs from the rest of the dialects in containing a parallel set of interrogative pronouns, morphologically distinguished by the lack of an incorporated personal pronoun (see Section 3.2.4), which are not utilized in situ but only in fronted position. Compare the following (with the /n/ of *šin* assimilating to following /b/ in (16)):

15. *bi-d-dūr šinū?*
CNT-2-want what
'What do you want?' (Šukriyya: Reichmuth 1983, p. 117)
16. *šim bi-d-dūr b-u?*
what CNT-2-want with-it
'What do you want with it?' (Šukriyya: Reichmuth 1983, p. 117)

Despite its status as a minor and pragmatically marked variant, the structural properties of this usage have important ramifications for the interpretation of the otherwise regular and distinctive feature of in situ WH-question formation in Egypto-Sudanic dialects. They, and other points noted throughout our review of these varieties' phonological, morphological and syntactic characteristics, will provide a critical qualitative dimension to the global evaluation of linguistic evidence for an Egypto-Sudanic dialect classification based in shared genealogical history. It is to this task we shall turn in the paper's remaining sections.

4. Discussion

4.1. Global Evaluation of Results

Having reviewed the major phonological, pronominal, verbal inflectional and syntactic characteristics of the Arabic dialects of Cairo, Qift, il-Bi'erāt, Khartoum, the Šukriyya, and Nigeria, we will now direct the information adduced toward a linguistic evaluation of existing proposals of an Egypto-Sudanic dialect classification, as has been repeatedly asserted on the nonlinguistic basis of shared genealogical history uniting the region's Arabic speakers. In the event that such nonlinguistic factors as migration history and common descent prove viable grounds for the classification and grouping of language varieties used in the region, expectation is that a substantial number of shared linguistic features will arise to characterize the varieties in question. This would justify the prediction of a meaningful degree of dialectological similarity as a consequence of the historical and demographic unity ascribed to their speakers by extra-linguistic lines of research.

This expectation, however, is not substantively met by the linguistic data gathered through the process of this inquiry. Of over fifty phonological, morphological, and syntactic features identified and discussed in the preceding subsections, only seven may be recognized as uniformly present across all Egypto-Sudanic varieties sampled. These are:

- */θ, ð, ð̥/ > /t, d, d/ (Section 3.1.1)
- */aw, ay/ > /ē, ō/ (Section 3.1.1)
- Initial */a/ > /i/ in second person pronouns (Section 3.2.1)
- Proximal demonstrative paradigm on the pattern *dā, dī, dōl, dēl (Section 3.2.2)
- 1.sg/2.m.sg perfect in -t, or further evolution thereof (Section 3.3.1)
- Post-nominal demonstrative order (Section 3.4.3)
- In situ WH-question formation (Section 3.4.4)

To these, we might, for the sake of consideration, generously add six more—those features which proved characteristic of all but one of the surveyed dialects, and whose incidence may thus have been proved broader in a different sampling. These are:

- */q/ > /g/ [all but Cairo] (Section 3.1.1)
- Distinction of */a, i, u/ [all but Nigerian] (Section 3.1.2)
- /i, u/ > Ø / VC_CV [-stress] [all but Nigerian] (Section 3.1.2)
- 3.f.sg perfect in -at [all but Cairo] (Section 3.3.1)
- Continuous aspect marker from *bi- [all but B'ēri] (Section 3.3.2)
- Genitive exponent from *bitāʕ [all but Nigerian] (Section 3.4.2)

The question, then, stands: Are these features sufficient to corroborate the existence of a linguistically significant Egypto-Sudanic dialect classification, proceeding from a common dialectal input carried by those historical communities who introduced Arabic first to Egypt, then to the Sudanic area via subsequent migration?

Though no conventionalized, objective threshold exists by which to make such a determination, the evidence in the Egypto-Sudanic case is not compelling—neither in terms of its quantity nor, critically, its quality. Of the thirteen isoglossic features identified as uniform or near-uniform across the six varieties examined, two—3.f.sg -at, and distinction of */a, i, u/—are clear retentions from a common Old Arabic inheritance, not innovations distinctive of further dialectal diversification. While thus not contradicting a narrative of dialectal relatedness due to shared migration history, neither do they positively support one: rather, they simply reflect the fact that dialects of the Egypto-Sudanic area have remain largely unimpacted by the mergers of */a, i/ emanating from the west of the modern Arabic-speaking world and */i, u/ associated with its north and east, as well as the change -at > -it typical of a number of Eastern Mediterranean varieties. None of these facts are surprising, and do nothing to indicate a shared developmental history of Arabic varieties in the region—simply a shared, central geography.

Of the remaining features which may be considered genuinely innovative, some are so ubiquitous across modern Arabic as to hold little meaningful value in establishing an identifiable Egypto-Sudanic dialect classification based in shared demographic heritage. Among these are the monophthongization of */ay, aw/, retained as diphthongs only in scattered relict zones; the use of a 1.sg/2.m.sg perfect suffix -t (< *tu), typical of virtually all modern Arabic varieties save those of the northern Fertile Crescent and parts of Yemen; and the change of initial */a/ > /i/ in the second person independent pronouns, identifiable in the vast majority of dialects outside the Arabian Peninsula (and many within it). These traits do not serve to differentiate dialects of the Egypto-Sudanic area from their immediate geographic neighbors in eastern Libya, the Hijaz or the Sinai (Owens 1984; Schreiber 1970; de Jong 2000), nor from the bulk of modern Arabic more broadly. A further number of features are not quite so universal in attestation, but still spread far beyond the bounds of the Egypto-Sudanic region. Fortition of interdental fricatives to corresponding stops, though not typical of the Egypto-Sudanic varieties' closest orbit of northern neighbors in eastern Libya or the Sinai (Owens 1984; de Jong 2000), is shared with the majority of varieties (both "sedentary" and some traditionally "Bedouin") of the remainder of North Africa and the

Levant, as well as urban Hijazi speech across the Red Sea (Schreiber 1970).⁴ Elision of /i, u/ (but not /a/) in unstressed, open-syllable environment is well known outside the region and is present in the Egypto-Sudanic varieties' easterly dialectal neighbors in the Sinai and Mecca; the same is true for the voicing of */q/ > /g/, which is commonplace westward into Libya as well (de Jong 2000; Schreiber 1970; Owens 1984). Reflexes of the verb-modifying prefix *bi- extend beyond the Egypto-Sudanic zone's eastern edges into the urban Hijaz and the Sinai (Schreiber 1970; de Jong 2000), and further into Arabia and the Levant. Though absent from eastern Libya, their presence resumes in that country's west (Owens 1984). These features, then—while of obvious descriptive relevance—do not much contribute toward the definition of a classificatory unit which interprets the Egypto-Sudanic varieties as a discretely identifiable group, distinguished from other, neighboring dialects by the products of a separate developmental history.

The original thirteen features which might have been invoked in this regard, then, have fallen to four: a proximal demonstrative paradigm on the pattern *dā, dī, dōl, dēl, unmarked and obligatory post-nominal demonstrative order, in situ WH-question formation, and use of the genitive exponent *bitāf. These traits, held in common across all or nearly all members of the sampled group, are both innovative and, largely, distinctive—not generally encountered beyond these dialects' immediate environs, neither are they typical even of closely neighboring varieties. A similar demonstrative paradigm is reported for Mecca alongside more common variants with an initial *hā- element, and *bitāf is variably attested in some dialects of the Sinai, but neither trait dominates in either region (Schreiber 1970; de Jong 2000). Both Meccan and eastern Libyan Arabic allow in situ WH-question and post-nominal demonstrative orders, but these are not unmarked or obligatory to the degree identified among the Egypto-Sudanic dialects considered here (Schreiber 1970; Owens 1984). From a synchronic descriptive standpoint, then, these four isoglosses stand as strong candidates to delineate linguistically meaningful boundaries between dialects of the Egypto-Sudanic area and adjacent Arabic varieties.

Such does not automatically, however, render these four features supportive of an Egypto-Sudanic dialect classification of the form so often proposed, predicated on the shared genealogical history of the Egyptian and Sudanic Arabic speech communities. Under such a framework, the claim advanced is that the migration of Arabic speakers from Egypt to the Sudanic region from the early Middle Ages onward carried to the latter a linguistic input characterized by recognizable dialectological features which may be observed to meaningfully describe and unite Arabic varieties of the Egypto-Sudanic zone to this day. There are clear reasons to doubt, however, that three of the four diagnostic features remaining to us represent the products of such a history. The *bitāf-type genitive exponents, for example, may be of reasonable antiquity—possibly attested as early as the eleventh century (Lentin 2018)—yet at the same time show every indication of representing a (Lower) Egyptianism only much later adopted by Arabic speakers of Upper Egypt and the Sudan. In the present sample, reflexes of *bitāf exist below Qift only in variation with other, heterogeneous genitive exponents, and are consistently identified by researchers and speakers alike as carrying urban and Egyptian sociolinguistic valuation (for empirical investigation of this sociolinguistic dimension, see Miller and Abu-Manga 1992; Miller 2005). These facts, combined with the relative novelty of *bitāf forms noted by Hillelson (1935) and their absence from Nigerian, would support a scenario of spread accompanying the colonial expansion and consolidation of Cairene political influence throughout the region under the Ottoman/Khedival and Anglo-Egyptian state apparatuses (ca. 1820–onward), rather than as part of an original linguistic input carried southward during the first waves of Arabization several centuries earlier.

Certain data likewise complicate the identification of two further syntactic features, post-nominal demonstrative order and in situ WH-question formation, as having arrived to Sudanic territory as part of a founding in-migration of Arabic speakers from Egypt. While post-nominal demonstrative ordering is normative throughout the Egypto-Sudanic region today (as the sampled dialects attest), this is known to not always have been the case. Doss

has demonstrated that pre-nominal demonstrative ordering in Egypt long existed as a historical alternative alongside the presently familiar post-nominal, and was “alive and productive” (Doss 1979, p. 356) in direct historical attestations dating as late as the seventeenth and eighteenth centuries; the pre-nominal structure, in fact, still exists in modern Cairene in a number formulaic usages and fixed expressions, including the grammaticalized *dilwa?ti* ‘now’ (< *di l-wa?t ‘this time’). Though lacking a pre-modern textual record to provide comparable direct evidence, similar synchronic clues (e.g., Šukriyya and earlier Khartoum *daḥīn* ‘now’ < *dal-hīn ‘this time’) indicate that exclusively post-nominal demonstrative order has likewise not always been uniform in the Sudanic area (Reichmuth 1983, pp. 122–26). In this light, the present-day regime of obligatory post-nominal demonstrative ordering becomes a far less viable candidate to have been imported to the Sudanic area from Egypt as part of the latter region’s initial Arabicization—not only because it does not appear to always have existed in Sudanic Arabic varieties, but also because it would not seem to have been so established in Egyptian varieties of the relevant era to begin with.

Direct historical attestation of WH-question formation is unfortunately less forthcoming, but internal reconstruction of the multimorphemic Sudanic interrogative pronouns *šinu and *minu may prove similarly revelatory. In contrast to their Egyptian counterparts of the types *ē(h) (< *ēš) and *mīn, these forms incorporate a reflex of a personal pronoun, which in some varieties still inflects to demonstrate agreement with the interrogated noun phrase. This difference is a critical one, in that it points to a structural dissimilarity in the diachronic source constructions that have given rise to the respective sets of interrogatives. Namely, the presence of the incorporated pronoun in the Sudanic varieties indicates the (historical) presence of a syntactic transformation in WH-questions, by which the non-interrogative element undergoes movement and is resumed by a third person pronoun in its deep-structure position. The following alternation of interrogatives with/without incorporated pronouns in the dialect of the Šukriyya is instructive:

- 17. *min ḥaddas-ak?*
 who told.3MSG-you
 ‘Who told you?’ (Šukriyya: Reichmuth 1983, p. 116)
- 18. *al-ḥaddas-ak minū?*
 REL-told.3MSG-you who
 ‘Who told you?’ (Šukriyya: Reichmuth 1983, p. 116)

The pronoun-incorporating structure in (18) would, presumably, have originally had its roots in a more complex, cleft-like structure on the order of (19), which has subsequently been subject to syntactic reanalysis/rebracketing:

- 19. **[al-ḥaddas-ak]_i min [hū]_i?*
 REL-told.3MSG-you who he
 ‘He that told you, who is he?’

While sentences like (15), above, make it demonstrably clear that pronoun-incorporating interrogatives in present-day Sudanic varieties do not (or do not necessarily) carry a synchronic clausal interpretation of this type, the diachronic implication of this developmental pathway should not be overlooked. While questions formed in the manner of *ra?y-ak ē?* and *ra?y-ak šinu?* ‘What’s your opinion?’ (Cairo and Khartoum, own knowledge) may both be validly described synchronically as displaying in situ formation, the latter presupposes an earlier cleft structure (**[ra?y-ak]_i šin [hū]_i* ‘Your opinion, what is it?’), which in turn presupposes the existence of a once-productive, WH-fronted, pronounless *šin* (cf. older Sudanese *šin gōl-ak* ‘What do you say [lit. What’s your saying]?’; Hillelson 1935, p. 62). The former does not, and the congruous modern products are thus assigned to two demonstrably incongruous developmental paths.

In the cases of WH-questions and demonstrative order, then, we must heed Pat-El’s warning that “syntactic reconstruction based on cognate patterns may conflate genuine inherited syntactic material with cases of parallel development” (Pat-El 2020, p. 332)—

or, we may add, cases of contact-induced convergence. Either or both of these syntactic patterns may have emerged in dialects of the Egypto-Sudanic area independently, or either or both may be the products of mutually influenced development through centuries of intra-regional contacts. In light of the historical and internally reconstructed data, however, neither appear to have been imported intact from Egypt to the greater Sudan with the onset of Arab settlement.

In terms of common Egypto-Sudanic features identified by this investigation which do in fact support such a narrative, we are subsequently left with a single linguistic trait: a proximal demonstrative paradigm on the model *dā, dī, dōl, dēl. This commonality is a genuinely striking one—being both innovative and distinctive—and demonstrative pronouns are undoubtedly a substantial feature of relevance to any serious attempt at Arabic dialect classification (see Magidow 2013, 2016). Yet, most would agree that they do not, in isolation, provide a viable solitary basis for the formulation of such groupings. This remaining commonality is thereby rendered less proof positive of classificatory relationship and more enigmatic isogloss to be marked for future investigation in light of broader Arabic demonstrative typologies. The traditional Egypto-Sudanic classification of the Arabic dialectology literature, predicated on the nonlinguistic genealogical relatedness and shared migration history of the region's Arabic-speaking communities, is thus left roundly unsupported following focused linguistic review.

4.2. Whence from Here? An Excursus in Historical Glottometry

Rejection of the traditionally formulated, genealogy-based Egypto-Sudanic dialect classification at a macro-level does not, however, refute or diminish the multifarious and noteworthy dialectal commonalities linking and cross-cutting smaller subsets of Arabic varieties spoken in this region, in varying combinations. These isoglosses, and the linguistic relationships they identify, are real and significant, and merit further study and elaboration—more than can be accomplished in a single contribution, by a single researcher, or, perhaps, via a single perspective on the information at hand. In cases like the present one, in which a long-standing hypothesis has been determined to lack fit, a fresh view on existing data is often as essential, and as conducive to progress, as the gathering of new. Here, one such opportunity (among many) comes in the form of “Historical Glottometry,” a novel approach to linguistic subgrouping recently elaborated by François and Kalyan (François 2014; Kalyan and François 2018).

Historical Glottometry was developed by its creators for application in scenarios in many ways analogous to the Egypto-Sudanic case described heretofore, in which the potential for “tree-like” relationships between once-unitary dialectal entities and “wave-like” patterns of convergence between previously more distinctive groups both loom large, and need both be considered in any comprehensive interpretation of the data. The method accomplishes this by integrating the key dialectological notion of the isogloss with the comparative method's focus on the common innovation, and labors to produce a diachronically interpretable measure of the relative strengths of multiple potential classificatory units revealed by analysis of a given dataset. Such an approach has been called for previously in the study of Arabic dialects (for a forcefully argued articulation, see Magidow 2017), and Historical Glottometry in particular has fruitfully filled this role in the examination of Boni dialect linkages (Elias 2019) and the Sogeram language family (Daniels et al. 2019), among others. I offer a preliminary application of the method here not as a route to a definitive classificatory model, but instead as an exploratory exercise into new views which may inform future analysis of the Egypto-Sudanic data, failing the identification of a meaningful macro-level relationship based in shared migration history. For example, review of the isoglosses presented in Section 3 offered numerous examples of two-way divisions separating northern dialects of the Egypto-Sudanic area from southern, but the precise positioning of isoglosses within this general pattern was observed to frequently shift on the basis of individual features, and to display a number of variable exponences. Can

a technique like Historical Glottometry offer additional, informative perspective which might lead to clarity in the comprehension and description of cases like these?

The tradition of quantitative dialectometry of which Historical Glottometry is a part is not alien to the Arabic dialectological tradition (see Behnstedt and Woidich 2005, pp. 106–35, for discussion), yet similarly has not been widely embraced by the field’s practitioners—for a host of valid critiques. As an analytical tool, Historical Glottometry joins these approaches in the effort to produce a linguistically meaningful yet condensed mathematical summation of data researchers “already know” (Daniels et al. 2019, p. 124) but which is copious and complex enough to defy ready intra-set comparability without transformation. Historical Glottometry accomplishes this via the production of two related values, each attending to a different aspect of linguistic classification generally agreed to hold significance in the field: “cohesiveness,” a measure of the proportion of relevant isoglosses held in common by the members of a potential classificatory unit, and “subgroupiness,” a measure of the number of isoglosses unique to the members of a proposed grouping. For a fully elaborated discussion of these measures’ conception and justifications, see Kalyan and François (2018, pp. 68–71); to summarize, cohesiveness is calculated as the number of innovative isoglosses shared by all members of a proposed grouping divided by the total number of isoglosses attested by any member of the group, thus taking into account both the quantity of isoglosses supporting a group and those conflicting with it; subgroupiness is derived by multiplying a grouping’s cohesiveness value by the number of exclusively shared isoglosses unique to the members of that group, thereby recognizing the importance of distinctiveness to most models of dialect classification while weighting the value of such features to reflect their position in broader dialectological context.

To apply this approach and calculate cohesiveness and subgroupiness scores for the array of dialect linkages attested by the Egypto-Sudanic data, I have accumulated the combined set of isoglosses considered in Section 3, focusing on those features which are clearly identifiable as innovative which are attested in a minimum of two varieties, and determined their presence/absence in each of the six dialects sampled. This tabulation of 342 values (6 dialects × 54 isoglosses) is included in Appendix A. I then calculated cohesiveness and subgroupiness scores for each of the subgroupings attested in the collected data, summarized in Figure 1, below. Cohesiveness scores are shaded in black, subgroupiness scores in white. Acronyms identify the composition of each classificatory group supported in the data by at least one exclusively shared feature (e.g., CBQK is a group consisting of the dialects of Cairo, Qift, il-Bi-eṛāt and Khartoum; KSN those of Khartoum, the Šukriyya and Nigeria, etc.).

The first and most evident take-away from the Historical Glottometry analysis of the Egypto-Sudanic dialects is that two potential classificatory units stand out as particularly strong and “subgroupy”: these are CQB and KSN—in other words, the three dialects of the Egyptian area taken as a group, and the three of the Sudanic. Not only do these respective sets of varieties share a meaningful proportion of their total features, but they also display a high number of exclusively shared features not identifiable outside the confines of the grouping (9 for each group). This geographical polarization of the dialect region, divided into groups representing the three northernmost and the three southernmost varieties of the sample, is replicated in the four-way groupings that emerge, which, with one (weaker) exception, consist of all three members of CQB or KSN in addition to one member of the other triad—the substantial diminution of both cohesiveness and subgroupiness incurred via such additions, though, reinforces the interpretation of the Egyptian/Sudanic split as a primary faultline in the data, rather than a single stage in a more gradual fading between northern and southern features. Indeed, turning to pairwise relationships, we similarly see that, excepting two linkages involving Nigerian, all other two-dialect groupings attested are internal to the CQB or KSN headings. Despite high cohesion, these are on the whole substantially weaker than either of the three-way groupings in terms of subgroupiness. Even the most significant pairing, KS, emerges as notably less strong than its superordinate KSN. These are key indications that the pan-Egyptian and pan-Sudanic dialect entities

KSN and CQB represent are not illusory extracts of a gradated continuum, nor secondary linkages of core plus orbit, but rather demonstrable, classificatorily significant units across which multiple distinctive, innovative features obtain. Historical Glottometry, then, has offered incisive, actionable insight to be further pursued in reshaping understandings of what dialect classifications may succeed the macro-level Egypto-Sudanic hypothesis: a scenario under which an Egyptian and a Sudanic group, though sharing a few broad characteristics and more numerous partially cross-cutting trends, stand out as robustly and independently definable in the absence of overarching linkage.

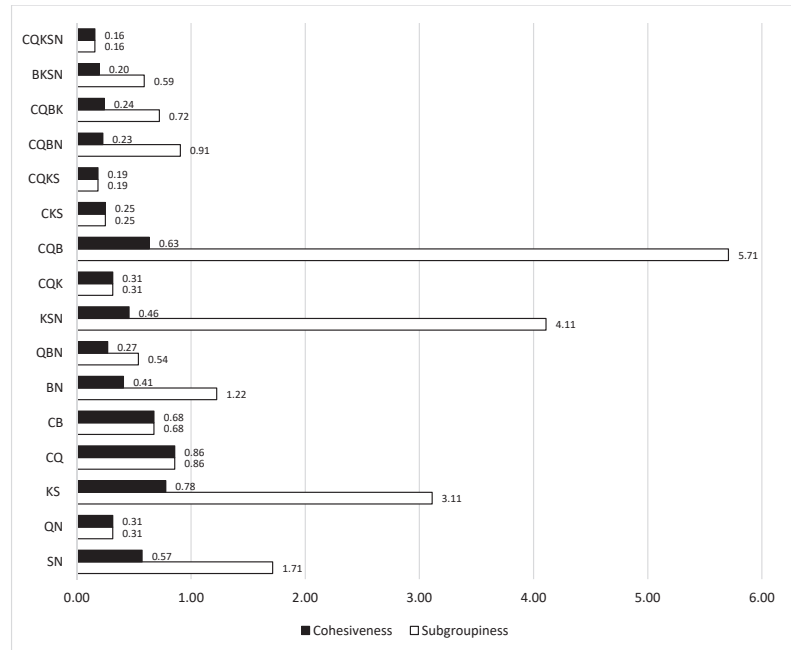


Figure 1. Historical Glottometry Scores of Attested Egypto-Sudanic Dialect Subgroupings.

The exception to this pattern is, as mentioned, Nigerian, which the four-way grouping CQBN shows to pattern more closely to the body of Egyptian varieties than does any other Sudanic dialect sampled; the still stronger pairwise grouping BN shows this affinity to exist more precisely with the dialects of Upper Egypt, particularly those represented by B̄ēri. This finding is notable in light of the well-described linguistic and demographic linkages between Upper Egypt and the Western Sudanic area detailed by Owens (1993b, 2003), including isoglosses beyond those considered here and a set of thoroughly sketched population movements from north to south occurring most prominently in the years leading up to 1500. The significance is thus twofold, serving as: (a) corroboration (admittedly circumstantial) of Historical Glottometry’s compatibility with otherwise-derived understandings of the region’s linguistic interrelationships, and (b) a reminder that the impact of migration events and shared genealogical history is not to be ignored in the interpretation of linguistic classificatory relationships. This last, then, underlines the urgency of the question of how a linguistically meaningful Egypto-Sudanic classification at large, girded by similar nonlinguistic factors, could fail to emerge in our broader analysis?

5. Conclusions

Given that population movements and shared genealogical histories of speech communities can and do influence dialect development in meaningful ways, it is understandable that these factors were utilized as proxy for genuine linguistic data in initial postulations of a classificatory affinity between Egyptian and Sudanic varieties of Arabic. These were generated at a time when such data were not forthcoming, and much of both areas remained dialectological *terra incognita* to Western Arabist scholarship. The window of usefulness for such stand-ins, though, is past. Existing descriptive works treating dialects of the Egypto-Sudanic region, though limited, are shown here to be sufficient to transition beyond this stage to engage in genuine linguistic evaluation of at least a subset of the varieties in question: their similarities, their differences, and their interrelationships. Yet, until this point, such has not been attempted in more than cursory fashion. Instead, once-preliminary assumptions based on nonlinguistic details were carried forward as received linguistic interpretation—fed by confirmation bias to the casual dialectological observer in the form of salient shared retentions, participation of dialects in broad areal trends, and instances of convergence via matter- and pattern-based borrowing or, conceivably, parallel development.

The present inquiry has demonstrated that, when faced with concerted linguistic investigation, little meaningful support can be found for the proposal that contemporary dialects of the Egyptian and Sudanic zones together constitute a viable classificatory grouping that reflects a common linguistic input carried by founding migrations of Arabic-speaking populations from the former region to the latter. The study is not without its limitations—its relatively narrow sampling and inattention to lexical variables, to start—but regardless has advanced a fairly unambiguous conclusion: that the historical demographic and genealogical ties seen to bind the area's Arabic-speaking communities in human relation to one another do not similarly define the relationships of those communities' dialects. Instead, these appear to pattern in discrete Egyptian and Sudanic blocs without significant superordinate connection, as occasionally disrupted by point-specific linkages and recent convergences contravening their general independence.

How is this contradiction between two dimensions of connectedness, the demographic-historical and the linguistic, to be reconciled? The first response of many will, perhaps, be to question the veracity of one set of understandings or the other. The linguistic findings of this study are, of course, not beyond reproach, and room similarly exists to interrogate historical conceptions of the Arabicization of Egypt and the Sudan from initial Muslim conquests (Booth 2013) to consolidation under the early Caliphate (Power 2012) to southward migrations of the medieval period (Spaulding 2000). But prior to—or, perhaps, in conjunction with—such, I would call for a pause. As dialectologists, we should not miss the opportunity to reflect on the assumptions and theoretical stances that have led us to such a conflicting position, and to ask whether the more fruitful questioning is that of the data or that of the frames through which we are wont to interpret it.

Much remains unknown about today's Arabic dialects' collective linguistic past, and much of that unknown is undoubtedly relevant to the sound comprehension and interpretation of their dialectological present. We must not, however, allow pursuit of those unknowns to become a preoccupation that unduly limits our imagination of what dialect classification strives to describe, or how the linguistic reality it represents enters into being. The amount to be learned from painstaking and revelatory excavation of the dialectal foundation laid by the earliest and subsequent waves of Arab migration and expansion is enormous—but it will never constitute a complete account. Arabic's arrival and establishment beyond its pre-Islamic environs via the physical movement of peoples is an obvious, massive watershed; all the same, myopic focus on the legacy of this era risks an artificial confidence that “by the 10th century [or perhaps, in the Egypto-Sudanic case, the fourteenth] dialectal areas were already shaped” (Abboud-Haggar 2006, p. 620).

The linguistic traces of past movements and demographic linkages are often long-lasting and significant—but they are not guaranteed to be present, and nor are they, when present, indelible. Contemporary sociolinguistic scholarship (Trudgill 1986; Al-Wer 2007) has repeatedly shown language use in the wake of demographic upheaval to be highly variable and diffuse, often so much so as to defy stable dialectological description. Similar states have been demonstrated for the Caribbean Englishes of Le Page and Tabouret-Keller (1985) and, at greater time-depth, in the case of Indo-European and Proto-Greek (Garrett 2006), to the extent that the *prima materia* of future dialect formations is reduced to classificatory nondistinctness. Moreover, as much as a dialect linkage is a product of its input, it is in equal or greater proportion an emergent entity which manifests over time, the earlier connectivities and commonalities shaped by its social and interactional past fully prone to being remolded and over-written—or occasionally, as we may be witnessing in the instances of syntactic convergence covered above, created anew as speakers’ present dictates. As Behnstedt and Woidich remind their colleagues following discussion of the development of the Egyptian dialect area, “[in] the historical evaluation of Arabic dialect phenomena, one cannot always assume that a feature was introduced from the original home of the speakers and implanted somewhere. One should also entertain the possibility that a given feature is the result of dialect mixing and dialect contact which eventually led to new dialects and new dialect areas” (Behnstedt and Woidich 2018, p. 95). My hope is that the present investigation of Egyptian and Sudanic Arabics serves to answer and emphasize this timely and pressing call.

Funding: This material is based in part upon research supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1110007. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in the study body and Appendix A.

Acknowledgments: I am grateful to the organizers and attendees of the 11th Conference of the Association Internationale de Dialectologie Arabe for comments on an early stage of this work, as well as to Rama Hamarneh for valuable feedback on the study in its current form.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

Table A1. Values for Historical Glottometry Analysis of Arabic Dialects of the Egypto-Sudanic Region.

Innovation ¹	Cairo	Qift	B̄ēri	Khartoum	Šukriyya	Nigeria
*/g/ > /ǰ/	–	–	+	+	+	+
*/g/ > /ǧʕ/	–	+	–	–	–	+
*/ð/ > /d/	–	–	–	+	+	+
*/θ, ð, ð̣/ > /t, d, ḍ/	+	+	+	+	+	+
*/aw, ay/ > /ē, ō/	+	+	+	+	+	+
/a/ > /e/ / _#	–	+	+	–	–	+
/i, u/ > Ø / VC_CV [-str.]	+	+	+	+	+	–
Ṫ > V / [-str.]	+	+	+	–	–	–
*nihnā > *ihnā	+	+	+	–	–	–
*huwa, hiya > *hū, hī	–	–	+	+	+	+
*huwa, hiya > *huwwa, hiyya	+	+	+	–	–	–
*hum > *humma	+	+	+	–	–	+
*hum > *hun	–	–	–	+	+	–

Table A1. Cont.

Innovation ¹	Cairo	Qift	B ̄eri	Khartoum	Šukriyya	Nigeria
*hinna > *hin	-	-	-	+	+	-
loss of f.pl agr.	+	+	-	+	-	-
*-hu > *-a	-	-	+	-	-	+
*-hu > *-u	+	-	-	+	+	-
*-ki > *-ik	+	+	-	+	+	-
*/h/ > /h ~ Ø/ in 3OBJ PROS	-	-	-	+	+	+
*/a/ > /i/ in 2SBJ PROS	+	+	+	+	+	+
*/nt/ > /tt/ in 2SBJ PROS	-	-	-	+	+	-
leveled */ð/ in DEMs	+	+	+	+	+	+
V-alternating DEM gender	+	+	+	+	+	+
*dō/ēlāk distal pl. DEMs	-	-	-	+	+	+
*dū/ēk distal pl. DEMs	+	+	+	+	-	-
DEM < *DEM + PRO	+	+	+	-	-	-
c.pl *dōl	+	+	-	-	-	-
REL *illī	+	+	+	-	-	-
REL *al-	-	-	-	+	+	+
*mīn ‘who?’	+	+	+	-	-	-
*minu ‘who?’	-	-	-	+	+	+
*ēh ‘what?’	+	+	+	-	-	-
*šīnu ‘what?’	-	-	-	+	+	+
*anhu ‘which?’	+	+	+	-	-	+
*yātu ‘which?’	-	-	-	+	+	+
1.sg PRF *tu > *-t	+	+	+	+	+	+
1.sg PRF *t > *i-(t)	-	-	-	-	+	+
3.m.pl *-u > *-o/-aw	-	+	+	-	-	+
3.m.pl *-ū- > *-ō-	-	-	+	+	+	+
IPRF *Ci-	+	+	+	-	-	+
IPRF *Ca-	-	-	-	+	+	+
IPRF *a-, n- > *n-, n-...-u	-	-	+	-	-	+
IMP *i-	+	+	+	-	-	-
IMP *a-	-	-	-	+	+	-
PASS *t-	+	+	+	+	-	-
PASS *n-	-	-	-	-	+	+
CNT *bi-	+	+	-	+	+	+
CNT *šammāl	+	-	+	-	-	-
FUT *rāh	+	+	+	+	-	-
verbal NEG *mā ...-š	+	+	+	-	-	-
distinct nominal NEG	+	+	+	-	-	+
GEN *bitāš	+	+	+	+	+	-
GEN *hagg	-	-	-	+	+	-
GEN *hūl	-	-	-	-	+	+
GEN *hana	-	-	+	-	-	+
NOUN DEM order	+	+	+	+	+	+
in situ WH-questions	+	+	+	+	+	+

¹ As necessary, innovations have been reformulated from their in-text descriptions to match Historical Glottometry’s sole focus on innovations rather than retentions; only those innovations attested in 2+ varieties are listed, and an innovation is considered present in a given dialect even if its occurrence there is variable. Innovations are presented in the order they are discussed in the article text.

Notes

- ¹ Though see [Brustad \(2000\)](#), for a broad-based comparative analysis.
- ² Development of a fronted realization [j] is likely to be quite old, potentially described by Sibawayh as early as the eighth century (cf. [Zaborski 2007](#)); this remains, however, innovative relative to the realization *[g] reconstructable with reference to several modern Egyptian and Peninsular dialects, as well as Semitic more broadly.
- ³ For further discussion of these and other exponents and their diachronic sources, see Eksell [Eksell Harning \(1980\)](#) and [Leddy-Cecere \(2018\)](#).

- ⁴ Sporadic */ð/ (> */ð/?) > /d/ in Sudanic varieties, but not Egyptian ones, also indicates that more general loss of interdental in those dialects likely post-dates arrival of their speakers to Sudanic territory.

References

- Abboud-Haggar, Soha. 2006. Dialects: Genesis. In *Encyclopedia of Arabic Language and Linguistics*, vol. I. Edited by Kees Versteegh. Leiden and Boston: Brill, pp. 613–22.
- Al-Wer, Enam. 2007. The Formation of the Dialect of Amman: From Chaos to Order. In *Arabic in the City: Issues in Dialect Contact and Language Variation*. Edited by Catherine Miller, Enam Al-Wer, Dominique Caubet and Janet Watson. London and New York: Routledge, pp. 55–76.
- Aoun, Joseph, Elabbas Benmamoun, and Lina Choueiri. 2010. *The Syntax of Arabic*. Cambridge: Cambridge University Press.
- Bar-Asher, Elitzur Avraham. 2008. The Imperative Forms of Proto-Semitic and a New Perspective on Barth's Law. *Journal of the American Oriental Society* 28: 233–55.
- Behnstedt, Peter, and Manfred Woidich. 1985–1999. *Die ägyptisch-arabischen Dialekte*, vol. I–V. Wiesbaden: Reichert.
- Behnstedt, Peter, and Manfred Woidich. 2005. *Arabische Dialektgeographie: Eine Einführung*. Leiden and Boston: Brill.
- Behnstedt, Peter, and Manfred Woidich. 2018. The Formation of the Egyptian Arabic Dialect Area. In *Arabic Historical Dialectology: Linguistic and Sociolinguistic Approaches*. Edited by Clive Holes. Oxford: Oxford University Press, pp. 64–95.
- Bergman, Elizabeth M. 2002. *Spoken Sudanese Arabic: Grammar, Dialogues, and Glossary*. Springfield: Dunwoody Press.
- Bloch, Ariel A. 1967. The Vowels of the Imperfect Preformatives in the Old Dialects of Arabic. *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 117: 22–29.
- Booth, Phil. 2013. The Muslim Conquest of Egypt Reconsidered. In *Constructing the Seventh Century*. Edited by Constantin Zuckerman. Paris: Association des Amis du Centre d'Histoire et Civilisation de Byzance, pp. 639–70.
- Britain, David J. 2016. Sedentarism and Nomadism in the Sociolinguistics of Dialect. In *Sociolinguistics: Theoretical Debates*. Edited by Nikolas Coupland. Cambridge: Cambridge University Press, pp. 217–41.
- Brustad, Kristen E. 2000. *The Syntax of Spoken Arabic: A Comparative Study of Moroccan, Egyptian, Syrian, and Kuwaiti Dialects*. Washington, DC: Georgetown University Press.
- Chambers, J. K., and Peter Trudgill. 2004. *Dialectology*, 2nd ed. Cambridge: Cambridge University Press.
- Daniels, Don, Danielle Barth, and Wolfgang Barth. 2019. Subgrouping the Sogeram Languages: A Critical Appraisal of Historical Glottometry. *Journal of Historical Linguistics* 9: 92–127. [CrossRef]
- Dickins, James. 2007b. Khartoum Arabic. In *Encyclopedia of Arabic Language and Linguistics*, vol. II. Edited by Kees Versteegh. Leiden and Boston: Brill, pp. 559–71.
- Dickins, James. 2007a. *Sudanese Arabic: Phonematics and Syllable Structure*. Wiesbaden: Harrassowitz.
- Dickins, James. 2009. Relative Clauses in Sudanese Arabic. *Journal of Semitic Studies* 54: 537–73. [CrossRef]
- Dickins, James. 2010. Basic Sentence Structure in Sudanese Arabic. *Journal of Semitic Studies* 55: 237–61. [CrossRef]
- Dickins, James. 2011. Dialects of Egypt and Sudan. In *The Semitic Languages: An International Handbook*. Edited by Stefan Weninger, Geoffrey Khan, Michael P. Streck and Janet C. E. Watson. Berlin and Boston: De Gruyter Mouton, pp. 935–53.
- Doss, Madiha. 1979. The Position of the Demonstrative *da, di* in Egyptian Arabic: A Diachronic Inquiry. *Annales Islamologiques* 15: 349–57.
- Eberhard, David M., Gary F. Simons, and Charles D. Fennig, eds. 2021. *Ethnologue: Languages of the World*, 24th ed. Dallas: SIL International. Available online: <http://www.ethnologue.com> (accessed on 24 May 2021).
- Eksell Harning, Kerstin. 1980. *The Analytic Genitive in the Modern Arabic Dialects*. Göteborg: Acta Universitatis Gothoburgensis.
- Elias, Alexander. 2019. Visualizing the Boni Dialects with Historical Glottometry. *Journal of Historical Linguistics* 9: 70–91. [CrossRef]
- Fischer, Wolfdieter, and Otto Jastrow. 1980. *Handbuch der arabischen Dialekte*. Wiesbaden: Harrassowitz.
- François, Alexandre. 2014. Trees, Waves and Linkages: Models of Language Diversification. In *The Routledge Handbook of Historical Linguistics*. Edited by Claire Bowerman and Bethwyn Evans. London and New York: Routledge, pp. 161–89.
- Garrett, Andrew. 2006. Convergence in the Formation of Indo-European Subgroups: Phylogeny and Chronology. In *Phylogenetic Methods and the Pre-History of Languages*. Edited by Peter Forster and Colin Renfrew. Cambridge: McDonald Institute for Archaeological Research, pp. 139–51.
- Hillelson, Sigmar. 1935. *Sudan Arabic Texts: With Translation and Glossary*. Cambridge: Cambridge University Press.
- Holt, Peter M., and Martin W. Daly. 2011. *A History of the Sudan: From the Coming of Islam to the Present Day*, 6th ed. London and New York: Routledge.
- Ingham, Bruce. 1971. Some Characteristics of Meccan Speech. *Bulletin of the School of Oriental and African Studies, University of London* 34: 273–97. [CrossRef]
- de Jong, Rudolf. 2000. *A Grammar of the Bedouin Dialects of the North Sinai Littoral: Bridging the Linguistic Gap between the Eastern and Western Arab World*. Leiden, Boston and Köln: Brill.
- Jongman, Allard, Wendy Herd, Mohammad Al-Masri, Joan Sereno, and Sonja Combet. 2011. Acoustics and Perception of Emphasis in Urban Jordanian Arabic. *Journal of Phonetics* 39: 85–95. [CrossRef]
- Kalyan, Siva, and Alexandre François. 2018. Freeing the Comparative Method from the Tree Model: A Framework for Historical Glottometry. *Senri Ethnological Studies* 98: 59–89.
- Kaye, Alan S. 1976. *Chadian and Sudanese Arabic in the Light of Comparative Arabic Dialectology*. The Hague: Mouton.

- Kaye, Alan S., and Judith Rosenhouse. 1997. Arabic Dialects and Maltese. In *The Semitic Languages*. Edited by Robert Hetzron. London and New York: Routledge, pp. 263–311.
- Khalafallah, Abdelghany A. 1969. *A Descriptive Grammar of Sa'idi Colloquial Arabic*. The Hague and Paris: Mouton.
- Le Page, Robert Brock, and Andr e Tabouret-Keller. 1985. *Acts of Identity: Creole-Based Approaches to Language and Ethnicity*. Cambridge: Cambridge University Press.
- Leddy-Cecere, Thomas. 2018. Contact-Induced Grammaticalization as an Impetus for Arabic Dialect Development. Doctoral dissertation, University of Texas at Austin, Austin, TX, USA.
- Leddy-Cecere, Thomas. 2020. Contact-Induced Grammaticalization between Arabic Dialects. In *Arabic and Contact-Induced Change*. Edited by Christopher Lucas and Stefano Manfredi. Berlin: Language Science Press, pp. 603–23.
- Lentin, J r me. 2018. The Levant. In *Arabic Historical Dialectology: Linguistic and Sociolinguistic Approaches*. Edited by Clive Holes. Oxford: Oxford University Press, pp. 170–205.
- Levin, Aryeh. 2007.  mala. In *Encyclopedia of Arabic Language and Linguistics, vol. II*. Edited by Kees Versteegh. Leiden and Boston: Brill, pp. 311–15.
- MacMichael, Harold Alfred. 1922. *A History of the Arabs in the Sudan: And Some Account of the People Who Preceded Them and of the Tribes Inhabiting D rfur, vol. I*. Cambridge: Cambridge University Press.
- Magidow, Alexander. 2013. Towards a Sociohistorical Reconstruction of Pre-Islamic Arabic Dialect Diversity. Doctoral dissertation, University of Texas at Austin, Austin, TX, USA.
- Magidow, Alexander. 2016. Diachronic Dialect Classification with Demonstratives. *Al-Arabiyya* 49: 91–115.
- Magidow, Alexander. 2017. Linguistic History and the History of Arabic: A Speech Communities Approach. In *Arabic in Context: Celebrating 400 Years of Arabic at Leiden University*. Edited by Ahmad Al-Jallad. Leiden and Boston: Brill, pp. 405–40.
- Manfredi, Stefano. 2014. Demonstratives in a Bedouin Arabic Dialect of Western Sudan. *Folia Orientalia* 51: 27–50.
- Manfredi, Stefano. 2018. The Semantics of Modals in Kordofanian Baggara Arabic. In *Afroasiatic: Data and Perspectives*. Edited by Mauro Tosco. Amsterdam and Philadelphia: John Benjamins, pp. 131–49.
- Miller, Catherine. 2005. Between Accommodation and Resistance: Upper Egyptian Migrants in Cairo. *Linguistics* 43–45: 903–56.
- Miller, Catherine, and Al-Amin Abu-Manga. 1992. *Language Change and National Integration: Rural Migrants in Khartoum*. Readings and Khartoum: Garnett-Khartoum University Press.
- Nishio, Tetsuo. 1995. Characteristics of the Arabic Dialect of Qift (Upper Egypt). *Journal of African and Asian Studies* 48–49: 173–219.
- Ouhalla, Jamal. 2008. Negation. In *Encyclopedia of Arabic Language and Linguistics, vol. III*. Edited by Kees Versteegh. Leiden and Boston: Brill, pp. 355–60.
- Owens, Jonathan. 1984. *A Short Reference Grammar of Eastern Libyan Arabic*. Wiesbaden: Harrassowitz.
- Owens, Jonathan. 1993a. *A Grammar of Nigerian Arabic*. Wiesbaden: Harrassowitz.
- Owens, Jonathan. 1993b. Nigerian Arabic in Comparative Perspective. *Sprache und Geschichte in Afrika* 14: 85–175.
- Owens, Jonathan. 2003. Arabic Dialect History and Historical Linguistic Mythology. *Journal of the American Oriental Society* 123: 715–40. [[CrossRef](#)]
- Owens, Jonathan. 2006. *A Linguistic History of Arabic*. Oxford: Oxford University Press.
- Owens, Jonathan, and Jidda Hassan. 2009. West-Sudanic Arabic. In *Encyclopedia of Arabic Language and Linguistics, vol. IV*. Edited by Kees Versteegh. Leiden and Boston: Brill, pp. 708–18.
- Heikki Palva. 2006. Dialects: Classification. In *Encyclopedia of Arabic Language and Linguistics, vol. I*. Edited by Kees Versteegh. Leiden and Boston: Brill, pp. 604–13.
- Pat-El, Na'ama. 2017. Digging Up Archaic Features: "Neo-Arabic" and Comparative Semitic in the Quest for Proto Arabic. In *Arabic in Context: Celebrating 400 Years of Arabic at Leiden University*. Edited by Ahmad Al-Jallad. Leiden and Boston: Brill, pp. 441–75.
- Pat-El, Na'ama. 2020. On Shared Structural Innovations: The Diachrony of Adverbial Subordination in Semitic. In *Reconstructing Syntax*. Edited by J hanna Bar dal, Spike Gildea and Eugenio R. Lujan. Leiden and Boston: Brill, pp. 314–35.
- Power, Timothy. 2012. *The Red Sea from Byzantium to the Caliphate: AD 500–1000*. Cairo and New York: The American University in Cairo Press.
- Proch zka, Stephan. 2014. Feminine and Masculine Plural Pronouns in Modern Arabic Dialects. In *From Tur Abdin to Hadramawt: Semitic Studies; Festschrift in Honour of Bo Isaksson on the Occasion of His Retirement*. Edited by Tal Davidovich, Ablahad Lahdo and Torkel Lindquist. Wiesbaden: Harrassowitz, pp. 129–748.
- Reichmuth, Stefan. 1983. *Der arabische Dialekt der Sukriyya im Ostsudan*. Hildesheim: Olms.
- Schreiber, Giselher. 1970. Der arabische Dialekt von Mekka: Abri  der Grammatik mit Texten und Glossar. Doctoral dissertation, Westfalische Wilhelms-Universitt zu M nster, M nster, Germany.
- Spaulding, Jay. 2000. The Chronology of Sudanese Arabic Genealogical Tradition. *History in Africa* 27: 325–37. [[CrossRef](#)]
- Taine-Cheikh, Catherine. 2000. Deux macro-discriminants de la dialectologie arabe (la r alisation du q f et des interdentes). *Mat riaux arabes et sudarabiques (GELLAS)* 9: 11–51.
- Taine-Cheikh, Catherine. 2007.  ass niyya Arabic. In *Encyclopedia of Arabic Language and Linguistics, vol. II*. Edited by Kees Versteegh. Leiden and Boston: Brill, pp. 240–50.
- Trudgill, Peter. 1986. *Dialects in Contact*. New York: Basil Blackwell.
- Versteegh, Kees. 2014. *The Arabic Language*, 2nd ed. Edinburgh: Edinburgh University Press.

- Vicente, Ángeles. 2006. Demonstratives. In *Encyclopedia of Arabic Language and Linguistics*, vol. I. Edited by Kees Versteegh. Leiden and Boston: Brill, pp. 569–73.
- Vicente, Ángeles. 2009. Relative Pronoun (Arabic Dialects). In *Encyclopedia of Arabic Language and Linguistics*, vol. IV. Edited by Kees Versteegh. Leiden and Boston: Brill, pp. 70–72.
- Wilmsen, David. 2014. *Arabic Indefinites, Interrogatives, and Negators: A Linguistic History of Western Dialects*. Oxford: Oxford University Press.
- Woidich, Manfred. 1974. Ein arabischer Bauerndialekt aus dem südlichen Oberägypten. *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 124: 42–58.
- Woidich, Manfred. 2006a. B'eri Arabic. In *Encyclopedia of Arabic Language and Linguistics*, vol. I. Edited by Kees Versteegh. Leiden and Boston: Brill, pp. 299–308.
- Woidich, Manfred. 2006b. *Das Kairenisch-Arabische: Eine Grammatik*. Wiesbaden: Harrassowitz.
- Zaborski, Andrzej. 2007. Jim. In *Encyclopedia of Arabic Language and Linguistics*, vol. II. Edited by Kees Versteegh. Leiden and Boston: Brill, pp. 494–96.
- Zeroual, Chakir. 2006. Affrication. In *Encyclopedia of Arabic Language and Linguistics*, vol. I. Edited by Kees Versteegh. Leiden and Boston: Brill, pp. 26–28.

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
www.mdpi.com

Languages Editorial Office
E-mail: languages@mdpi.com
www.mdpi.com/journal/languages



MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland

Tel: +41 61 683 77 34

www.mdpi.com



ISBN 978-3-0365-6140-0