The Dissertation Committee for Thomas Alexander Leddy-Cecere Certifies that this is the approved version of the following dissertation:

Contact-Induced Grammaticalization as an Impetus for Arabic Dialect Development

Committee:

Kristen Brustad, Co-Supervisor

Danny Law, Co-Supervisor

John Huehnergard

Patience Epps

Contact-Induced Grammaticalization as an Impetus for Arabic Dialect Development

by

Thomas Alexander Leddy-Cecere

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2018

Dedication

To NS, where this all started, and RH, with whom it will move forward.

Acknowledgements

Before UT, my schools were always small, but my education never was. I am forever indebted to the Newton School of Strafford, Vermont, where I learned what it means to be a writer, a scientist and a teacher, and to The Sharon Academy, which left me with a directive to *seguir con tus lenguas* – I hope the following pages will not disappoint.

My gratitude goes also to the dedicated faculty in Linguistics and Arabic at Dartmouth College, who both encouraged me to take this road and supplied me with a powerful education to speed me on my way: to Drs. James Stanford, Timothy Pulju, and Diana Abouali, a special thank you.

I found an intellectual family at UT before ever setting foot in Austin; my warm appreciation, (now) Drs. Alex Magidow and Emilie Zuniga, for making me feel welcome on a very early conference morning in San Antonio, and for years to come.

The leader, soul, and limitless inspiration of that family is and forever will be Dr. Kristen Brustad. The encouragement, care and intellectual refinement she has contributed to this dissertation have been indispensable, and nothing more or less than she has given every idea I have brought her over our years of working together, large or small. From a name on the front of a dog-eared book to a trusted adviser, mentor, collaborator and friend, I owe you a debt, *Ustāza*, that I can only repay by serving my own students as skillfully and as tirelessly as you have served me; I may never get there, but I will try with all that I have, and all that you have taught me.

Dr. Danny Law has changed my idea of what it means to be a linguist, and his stamp on this work and my work to come will be clear to any who have benefitted from his scholarship. His insistence on thoughtfulness and rigor, in equal measure, represents

a standard to which he has taught me – through example and kind guidance – to proudly adhere. A thousand thanks for your untiring support of this project, from minute one.

Dr. John Huehnergard has brought to this dissertation an erudition and professional grace to which I will always aspire, and Dr. Pattie Epps a globalness of scope which has shown me how to be consistent but flexible, focused but not myopic. I hope that this work reflects the qualities and input of all of these supervisors and committee members who have given to the project so generously, and that it represent the first step of many to follow in the paths each has blazed.

Beyond the committee itself, my learning at UT would not have been half of what it was without the wisdom and encouragement of Drs. Mahmoud Al-Batal, Barbara Bullock, Jacqueline Toribio, and Anthony Woodbury, consummate educators all. My gratitude, as well, to a cohort of peers who have pushed me to the highest heights and lifted me out of the lowest lows: Mike Turner, Dr. Jason Schroepfer, Dr. Phillip Stokes, Ryan Fan, Dr. Brendan Regan, and others, both within my specialty and without. If what starts here changes the world, I'm honored to have started with you.

I am grateful, also, for the support of my family: to my sister, who is smarter than me; to my father, who recognized and supported my love of all things language long before I ever did; and to my mother, who shows me every day what it means to be both educated and educator, and who has made my dreams her own.

Last and most, I thank Rama Hamarneh. This achievement is ours together, in every way.

This material is based upon work 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.

Contact-Induced Grammaticalization as an Impetus for Arabic Dialect Development

Thomas Alexander Leddy-Cecere, Ph.D.

The University of Texas at Austin, 2018

Supervisors: Kristen Brustad, Danny Law

This dissertation proposes contact-induced grammaticalization as an account for the widespread occurrence of functionally analogous but etymologically distinct grammatical innovations across modern Arabic dialects. Similarities in functional and semantic details of these grammatical items argue for interrelated development, while diversity in form rules out an origin in common inheritance or matter-based borrowing. The dissertation proposes that these developments are products of the diffusion of grammaticalization pathways between neighboring dialects by means of replication. This hypothesis is evaluated using a sample of attested realizations of three relevant classes of developments (future tense markers, temporal adverbs meaning 'now', and genitive exponents) drawn from eighty-one modern Arabic varieties, examined by means of a three-part heuristic which assesses 1) the status of individual innovations as examples of grammaticalization, 2) the multiple replication of attested grammaticalization pathways, and 3) the geographical distribution of modern reflexes as indicating a history of areal diffusion. The results demonstrate substantial evidence for the role of contact-induced grammaticalization in all three sets of developments examined, and the dissertation concludes by discussing the significance of these findings for the study of Arabic diachrony and contact-induced grammaticalization theory more broadly.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Overview

Studies in historical Arabic dialectology have long been stymied by what scholars in the field label the "pluriform development" of common features, whereby "a general trend … has occurred in all Arabic dialects [as against the Old Arabic type attested by the classical language], and an individual translation of this trend in each area" (Versteegh 2001: 108; see also Abboud-Haggar 2006). Commonly discussed examples of this phenomenon include the developments of novel future tense markers, continuous aspect markers, and genitive exponents across the great majority of contemporary Arabic varieties: collectively, each set of developments presents a striking typological and functional uniformity while simultaneously reflecting numerous distinct etymological sources. This diversity of form has generally been viewed as precluding an origin of the otherwise similar items in shared linguistic inheritance; diffusion of the items via dialect contact has likewise been written off due to the belief (not uncommon among practitioners of traditional dialectology) that "typically dialect contact leads to the borrowing of another dialect's markers, not to the borrowing of a structure which is then filled independently" (Versteegh 2001: 108). As such, alternative proposals to explain these seemingly unrelated yet analogous developments – ranging from linguistic drift to substrate influence to creolization processes – have long been circulated among Arabic dialectologists and historical linguists, though none have yet found widespread acceptance.

To date, however, analysis of these phenomena has focused little attention on the abstract evolutionary trajectories of the items in question, nor on the general geographic incidence of those patterns: upon initial examination, many of the seemingly distinct

linguistic items appear to have arisen via grammaticalization processes tracing semantically parallel paths, and moreover these parallel paths are often distributed in a manner which reflects a high degree of spatial contiguity. In this dissertation, I propose that these two facts, taken together, are indicative of the results of the process which Heine and Kuteva (2003; 2005) label contact-induced replica grammaticalization, in which "a grammaticalization process … is transferred from the model (M) to the replica language (R)," without corresponding transfer of an actual phonological form (Heine & Kuteva 2003: 539). Though most often discussed in the context of contact between genetically distinct languages, I apply Heine and Kuteva's approach to the investigation of interaction between related linguistic varieties to evaluate the explanatory power of contact-induced grammaticalization (CIG) as a mechanism underlying the profusion of pluriform developments across modern Arabic varieties.

Following a review of existing literature on the Arabic pluriform developments and previous attempts to account for their origin, I proceed to describe the data sources and methodology utilized in the present investigation: most importantly, this consists of establishing a viable heuristic for my evaluation of the potential role of CIG between Arabic dialects in the development of individual pluriform Arabic items. This heuristic is grounded in the central principles of CIG as proposed by its originators, and is additionally informed by conclusions stemming from prominent critiques of the proposal and the insights of related or overlapping theories pertaining to similar phenomena. The subsequent chapters of the dissertation consist of a series of case studies addressing each of three major pluriform developments in modern Arabic: future tense markers, temporal adverbs signifying 'now', and analytic genitive exponents. In each case, individual iterations of a given development are drawn from a geographically comprehensive sample of eighty-one modern Arabic varieties. The etymologies and developmental

trajectories of these individual forms are identified and compared, and the previously mentioned heuristic is applied to assess the suitability of CIG as an account for the collective development of the forms in question. In each case, plausible accounts relying on processes other than CIG, such as "classic" borrowing, calquing and drift, are considered and, if appropriate, preferred.

In the final chapter, the cumulative results of the three case studies are considered together, and – based on the apparent role of CIG as an impetus for the specific pluriform developments examined – generalizations are drawn regarding the broader potential for the phenomenon as a force of diachronic change in the history of the language. Implications of the findings are also considered from a theoretical perspective, inasmuch as they represent a unique empirical investigation of the results of CIG as it could occur between related language varieties, a scenario left largely undiscussed by Heine and Kuteva for methodological reasons but emphasized by thinkers such as Dahl (2001) as central to a global understanding of the process. The chapter and the investigation as a whole conclude with the identification of open questions which remain to be addressed, as well as a discussion of desired directions for future research and expansion on this study's findings.

I now proceed to describe more thoroughly the phenomenon of the pluriform developments as it is understood in the Arabist literature, and to present and discuss previous explorations of this topic. This elaborated view will frame the central questions driving this dissertation and more precisely situate its aims as they connect to the theoretical precepts and analytical methods outlined in further sections.

1.2 PLURIFORM DEVELOPMENTS IN ARABIC

As mentioned above, a perennial question in the historical dialectology of Arabic revolves around what have been labeled the language's "pluriform developments." According to Versteegh, a pluriform development may be identified when "a general trend … has occurred in all Arabic dialects, with an individual translation of this trend in each area" (2001: 108). By trend, Versteegh intends a specific class of structural innovation shared across the body of dialects; by individual translations, he refers to the set of distinct, seemingly unique phonological forms for which no single proto-form or etymon can be reconstructed to account for all products of said common innovation. An oft-cited example of pluriform development is the case of future tense markers. Though Pre-Islamic Arabic inscriptions have not to date revealed an explicit future tense marker (Macdonald 2004; Al-Jallad 2015) and no modern Arabic varieties present extant reflexes of the Classical Arabic future marker *sawfa* ~ *sa-*, the vast majority of dialects have innovated novel future tense markers of diverse etymological origins. The following represent a small but illustrative sample of this diversity (examples drawn from Versteegh 2001; Brustad 2000):

Dialect	Future Tense Marker
Moroccan	*ɣa-*[1]
Maltese	*seyyer-*
Egyptian	*ħa-*
Iraqi	*raħ-*
Syrian	*raħ-, b(i)-*
Kuwaiti	*b(i)-*
Yemeni	*ʃā-*

Table 1: Examples of Arabic future tense markers

Given the diversity of forms represented above, any practitioner of historical linguistics will recognize the difficulty in positing a viable proto-element from which all of these modern reflexes derive. This begs a question: despite the lack of a reconstructable proto-form, is there a means by which to explain the uniform inclination all of these dialects show to develop an innovative marker of future tense?

A number of alternative proposals have been circulated by those wishing to address questions of this type and account for the evolution of pluriform items in the modern Arabic dialects in a unified (or at least integrated) manner. More often than not, these have taken the form of observations made as part of general descriptions of

[1] In this dissertation, I utilize a regularized system of transcription intended to 1) allow for immediate comparability of data from disparate sources and 2) be readily accessible to both the Arabist/Semitist and the general linguist. The resulting transcription is necessarily quite broad, and should not be taken as indicative of fine phonetic detail. Forms have been adapted in a manner appropriate to phonetic and phonological details provided in their original source materials. Major departures from the International Phonetic Alphabet include <y> for [j] and <j> for [ʤ]. In some cases, hyphenation indicating word-internal morphological structure and bound/free status has also been modified from original sources in order to facilitate comprehension of the data by those not familiar with Arabic in its varied forms. Abbreviations in glosses follow standard Leipzig glossing conventions.

diachronic change in Arabic which treat a wide array of features, for example Rabin (1955). On occasion, however, authors have afforded such questions specific attention in monograph form, as in the case of Versteegh (1984) and Wilmsen (2014), and in so doing have shown a willingness to draw on cross-linguistically applicable theoretical principles to underpin their analyses. This dissertation builds on the latter tradition, adopting a comprehensive scope in terms of data collection and analysis while maintaining a focused theoretical approach to the pluriform developments, grounded in contemporary trends in linguistic thought.

Perhaps the most established opinion arising from prior literature regarding the Arabic pluriform developments is that they represent the independently actuated outcomes of linguistic drift, pre-conditioned by specificities inherent in the similar structural and typological profiles of the various modern dialects. While such a perspective is not alien to modern linguistic inquiry (see, for example, Enfield's (2003) description of "typological poise"), analysts adopting such a position in the Arabic context have tended to espouse versions of the theory significantly less well aligned with contemporary understandings in linguistics at large, appealing instead to broad Sapirian notions of "linguistic stock" and inevitable development effected over deep time scales in order to maintain systemic balances. Moreover, most such authors (Rabin 1955; Corriente 1971-1972; Abboud-Haggar 2006) have chosen to situate their arguments in what may be considered an excessively wide scope by describing outcomes in Arabic varieties as reflective of "Semitic Drift," attempting a unitary description of developments in Arabic varieties as reflecting tendencies inherent to the Semitic language family as whole, or a major constituent part thereof. As formulated by Abboud-Haggar in discussing the genesis of modern Arabic varieties:

> Pre-Islamic colloquial varieties of Arabic, as Semitic languages – its linguistic stock (Sapir 1949: 153) – pertain to the West Semitic Group, most probably the South Semitic subgroup … They were exposed to the group's natural drift – when language moves with time in a current of its own making (Sapir 1949: 150-151) – and variations were assimilated by speakers, becoming, with time, part of its structure (Abboud-Haggar 2006: 616).

While for some features considered among the pool of pluriform developments characteristic of modern Arabic varieties an expanded scope beyond the Arabic branch of the Semitic family may prove informative, for many others its relevance is highly questionable – for example, the use of future tense markers such as those described above is not typical of general Semitic, wherein future actions are typically grouped with present tense ones and marked by imperfective aspect morphology (cf. Gensler 2011). While a more focused examination based solely on the structural inheritance of the Arabic node might well improve upon the rather nebulous findings of prior investigations, to date no such studies have been carried out.

Even leaving aside questions of scope, accounts for the Arabic pluriform developments based on drift face perhaps a more immediate critique. The indiscriminate application of theories of linguistic drift has been called into question in recent years for a number of reasons, chief among them the rising support for grammaticalization theory among scholars of historical linguistics. The problem that grammaticalization poses for drift lies in the fact that it provides an alternative explanation for parallel yet seemingly historically distinct developments identified across languages. This explanation does not rely on a system-internal determinism (verging on predestination) inherent to numerous

languages linked by distant genetic affiliation, but rather posits human cognitive universals common to speakers of all languages and capable of effecting analogous trajectories of linguistic change wherever they may be active. While existence of the latter phenomenon does not negate that of the former, it does certainly circumscribe its identification, and care must be taken to distinguish between the two in arenas once considered the province of drift alone. This distinction has been elaborated upon by Law (2014), whose conclusions are drawn upon when relevant throughout this investigation. With regard to the study of Arabic pluriform developments, it suffices to say that previous accounts invoking drift do not differentiate between the possible results of the two phenomena, and this fact weakens their acceptability in a modern theoretical environment.

A second popular approach to the study of morphosyntactic diversification in Arabic is to view novel developments in modern Arabic varieties not as developments at all, but rather as reflections of pre-diasporic diversity in early Arabic which have, with the passage of time, come to be geographically reallocated across the modern Arabophone world; thus, they are viewed as representing (at least as far as modern Arabic dialectology is concerned) the results of linguistic inheritance rather than innovation. Adherents of this perspective, for example Owens (2006), have tended to avoid in-depth examination of pluriform developments proper in favor of simplex elements like verbal agreement inflection, personal pronouns, and case morphology when making their arguments while simultaneously implying that their conclusions are more broadly generalizable. However, some examinations have approached the topic head-on and applied such an analysis to features often considered to fall under the "pluriform" label; these include Wilmsen's (2014) investigation of a related complex of negators,

interrogatives and indefinite articles, as well as Magidow's (2013) account for the distribution of historical presentative elements in modern Arabic demonstrative forms.

Though individual cases of continuity between older observable forms of the language and modern Arabic varieties can and should be pursued when evident, the blanket adoption of a "what's new is old" philosophy as a *modus operandi* is not without risk. In the quest to establish continuity between older and younger layers of the language, the former most often sparsely attested, there exists potential to underestimate the occurrence of independent change and to overlook pertinent facts deriving from the comparative or internal reconstruction of a given innovation. These might include evidence for the gradual, *in situ* evolution of features or details of relative chronology which serve to definitively trace both pluriform and "classic" linguistic developments to the post-diaspora period. This perspective is thus diametrically opposed to the ideas of drift just discussed: in insisting that because an innovation is known to have occurred once it is unparsimonious to believe it has occurred twice, proposals of this school have been met with skepticism from scholars representing a wide array of theoretical backgrounds.

A third, more radical analysis of the pluriform developments has been championed by Versteegh (1984), who turns to universal correlates of language restructuring to explain the constant appearance of semantically and structurally parallel, yet formally independent, linguistic developments. Noting that many of the pluriform features share similarities with proposed universal products of creolization scenarios, Versteegh references the sociolinguistic setting of the early Islamic conquests to theorize that the initial waves of Arab expansion resulted in a series of Arabic-lexified pidgins-then-creoles, stemming from situations of large scale, untutored adult language acquisition. Following nativization via adoption by subsequent generations born in the

newly Arabized areas, Versteegh proposes that these varieties partially decreolized over time via exposure to Classical Arabic and dialects spoken by isolated, unaffected Bedouin groups. The results would be the Arabic dialects as we know them today, and any apparently inexplicable parallel developments they collectively display, notable for breaking typologically with earlier known forms of the language, would be reflective of universal aspects of their prior, creole-like character.

Versteegh's proposal has the advantage of motivating parallel structural evolution across dialects while obviating the logical necessity of a singular, original innovation, and does so in a far more concrete, cross-linguistically verifiable manner than accounts based on drift or the novel redistribution of old features. However, thirty years on, the proposition has been largely abandoned, even to the point of the author himself dramatically softening his position on the issue (cf. Versteegh 2014). Critiques have been leveled based on the nature of the Arabic data, pointing out that the specific pluriform features Versteegh seeks to explain as creole characteristics in fact represent only a portion of the set of common, parallel developments noted to exist cross-dialectally, thus reducing the explanatory power of the model; further, detractors have highlighted the broadly non-creoloid character of the Arabic dialects when taken in synchronic, cross-linguistic perspective. Perhaps ironically in light of the latter argument, additional doubt is shed on Versteegh's creolization proposal by increasingly widespread assertions in the language contact literature that call into question the very existence of a definable set of universally identifiable creole characteristics as a unique, empirically verifiable phenomenon (e.g., DeGraff 2005).

Across this multifaceted and decades-long debate of the origin of pluriform developments in Arabic, it is notable that convergence and contact between modern dialects has been roundly rejected as a potential explanation for the phenomenon. The

reason for this is the widely-held view among traditional Arabic dialectologists (not alone among many of their more conservative compatriots in other subfields and in contact linguistics more broadly – cf. King 2000) that dialect contact necessarily consists of the borrowing of linguistic "matter," in the meaning of discrete, traceable phonological material. Typical of such a conviction is the following assertion by Versteegh: "The difference in realisation [of pluriform developments] precludes an explanation in terms of later convergence, because typically dialect contact leads to the borrowing of another dialect's markers, not to the borrowing of a structure which is then filled independently" (Versteegh 2001: 108). It is my position that, in light both of developments in the study of dialect contact and language contact more broadly and of the known interconnectivity of the Arab world's social history (Hourani 1991), such a dismissal in not warranted. Instead, I assert that dialect contact need not always result in the transfer of linguistic matter but can instead rest solely on the diffusion of "pattern" alone, and that the identification of such a process in this case provides unprecedented explanatory power as a mechanism underlying the pluriform developments of the modern Arabic dialects.

More specifically, I invoke the notion of contact-induced grammaticalization as formulated by Heine and Kuteva (2003; 2005) as an account with particular promise. This theory extends beyond a straightforward conceptualization of structural convergence to bridge the gap between externally and internally induced language change, and to elucidate the role of contact in influencing the evolutionary trajectories of specific linguistic items over time. Its application in the present case is recommended by a generally overlooked characteristic of the pluriform items in Arabic, relating to the nature of their respective diachronic sources. As previously mentioned, it is extremely difficult to identify a single – or even a few – reconstructable proto-source(s) for the plethora of forms representing a given development. However, the task becomes suddenly much

simpler when one focuses not on specific etyma but on more general etymologies, in the sense of developmental tracks or, as is often the case, grammaticalization pathways. Returning, for example, to the set of future tense markers listed above, the six distinct etyma represented by the seven dialects are in fact reducible to two major grammaticalization pathways, representing respectively the cross-linguistically common deallative (GO) and devolitive (WANT) type futures known from the worldwide grammaticalization literature (Bybee, Perkins & Pagliuca 1994):

Dialect	**Future Tense Marker**	**Etymology**	**Geography**
Moroccan	*ɣa-*	<**ɣādī* 'going'	North/West
Maltese	*seyyer-*	< **sāyir* 'going'	North/West
Egyptian	*ħa-*	< **rāyiħ* 'going'	North/West
Iraqi	*raħ-*	< **rāyiħ* 'going'	North/West
Syrian	*raħ-,*	< **rāyiħ* 'going'	Overlapping
	b(i)-	< **biddu* 'wants'	territory
Kuwaiti	*b(i)-*	< **yibbī* 'wants'	South/East
Yemeni	*ʃā-*	< **yiʃā* 'wants'	South/East

Table 2: Examples of Arabic future tense markers (expanded)

Still more intriguingly, this apparent uniformity in grammaticalization paths is mirrored by a high degree of cohesion in their geographic distribution. In simple terms, the GO-futures are attested by the sample cluster to the north and west of the Arabic-speaking world, while the WANT-futures are found to the south and east, with Syria representing overlapping territory. These two facts, taken together, are indicative of a scenario in

which a select number of grammaticalization pathways have been repeatedly shared and replicated via processes of areal diffusion: in other words, they are strikingly consistent with the anticipated results of contact-induced grammaticalization, as described by Heine and Kuteva (2003; 2005), Dahl (2001), and others.

The identification of CIG as the mechanism underlying the modern distribution of pluriform developments entails several advantages over the previous accounts described above. As against theories of drift, it grounds the innovations observed in system-external motivations linked to the human cognitive faculty, and in so doing avoids the often circular formulations of causality which have troubled linguistic analysts since Sapir's day. It also avoids the latter's restricted explanatory scope, which is necessarily limited to developments which are otherwise attested by language varieties related to Arabic in varying degrees and only widens at the peril of its own coherence. Recommending CIG over explanations based on the redistribution of pre-existing dialect features is the fact that it recognizes the possibility, in fact the virtual certainty, of novel, *in situ* development in Arabic varieties during the thirteen centuries following the initial expansion of the language outside its ancestral area, and allows for the occurrence of analogous innovations without insisting that they represent the results of misinterpreted linguistic conservatism. Though Versteegh's creolization hypothesis shares with CIG these benefits over the first two approaches, in the end it falters in convincingly connecting data to theory from both Arabist and creolist viewpoints: the relevant set of creole features (itself constantly being narrowed) is not sufficient in either quantity or quality to explain the collection of innovative developments Versteegh sets out to examine. The less strictly defined outcomes of CIG thus offer an advantage in describing the evolution of the full set of Arabic pluriform features, as it is not limited *a priori* to a circumscribed list of possible products.

Another consideration that recommends CIG over all three of the aforementioned theoretical perspectives is its capacity to account for the consistent areal patterning of the pluriform developments' evolutionary pathways. This is an aspect of the pluriform development question which has not been addressed in previous examinations of the topic, perhaps because of the problems it poses for the theoretical approaches these investigations have proffered. If the pluriform elements are to be understood as the results of drift, reshuffled common inheritance or creole universals, there is no immediate reason why they should display any degree of geographic organization in their distribution. As intrinsic characteristics of the Arabic linguistic package, forces of drift and common inheritance would in theory have equal potential to affect all modern varieties of the language, regardless of location, and any geographical distinctions in the progress of creolization processes would need to be motivated by socio-cultural discrepancies between regions, for which the necessary historical evidence is wanting. An account based in CIG, however, provides a clear rationale for the occurrence of coherent areal groupings, as will be explained in the full description of the phenomenon found in the following sections.

1.3 THEORETICAL BASES

In the subsections included under this heading, I describe three aspects of diachronic linguistic theory which feed most directly into the present investigation's conceptualization and execution. First, I offer a brief introduction to the theory of grammaticalization at its present state of research, and then proceed to discuss some current understandings of language and dialect contact studies as they pertain to the topic and approach of this dissertation. Thirdly, I turn to proposals of contact-induced

grammaticalization specifically, to be taken in light of the two broader discussions referenced in framing the analysis to follow.

1.3.1 Grammaticalization

Most linguists would agree that it is possible to synchronically classify the majority of linguistic forms along a cline from "more lexical" to "more grammatical," in a manner roughly consistent with the following progression as conceived by Hopper and Traugott (2003):

CONTENT WORD > GRAMMATICAL WORD > CLITIC > INFLECTIONAL AFFIX

Historical linguists would add to this synchronic observation the diachronic reflection that it is common to observe a single etymological item advancing through the successive stages of this cline as it develops as part of a linguistic system over time. In fact, the sheer frequency of examples indicating such a trajectory of evolution in the world's languages has led to the identification of this type of development as a universally applicable theoretical phenomenon known as grammaticalization. The following definition of grammaticalization provided by Hopper and Traugott is representative of several currently referenced in the field, which – though differing in emphasis and points of detail – are broadly aligned in central principle:

> [Grammaticalization is] the change whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions and, once grammaticalized, continue to develop new grammatical functions (Hopper & Traugott 2003: 18).

Though useful for purposes of general definition, the above, largely intuitive formulation of the phenomenon and the cline which it follows must be further deconstructed if their implications are to be operationalized as part of a rigorous analysis. Andersen (2008) summarizes the issue succinctly when he observes that the grammaticalization cline as classically articulated conflates numerous discrete dimensions of language change in presenting them as unified steps in a chain: the shift from lexical to grammatical word is one of semantic content, while that from word to clitic to affix involves morphosyntax and any associated loss of phonological material is best understood as a phonological development. Beginning at the early stages of grammaticalization research, more complex approaches to the description of the phenomenon were proposed based on the concurrent evaluation of multiple parameters. Lehmann (1985; 2015), for example, identifies integrity, paradigmaticity, paradigmatic variability, scope, bondedness, and syntagmatic variability as relevant criteria for assessment; these are to be understood as manifestations of the parameters of weight, cohesion and variability varying along the two opposed axes of paradigmaticity and syntagmaticity. Along similar lines, Andersen (2008) prefers to understand grammaticalization as consisting of a four-way composite of changes to content, content syntax, expression, and expression syntax. Others, including Hopper (1991) and Heine (2007), opt instead to define analogous parameters in terms of diachronic processes, thus rendering them more directly relatable to the modes of diachronic analysis which underlie the bulk of investigations in grammaticalization research.

In the context of the present dissertation, I prefer such a process-oriented approach over parameter-based descriptions. This is for precisely the methodological interest just noted: as historical linguistic techniques are designed not for the description of static characteristics but rather the inference of dynamic change between multiple

stages of evolution, a focus on processes avoids a potentially calamitous co-identification of synchrony and diachrony. Of the process-based approaches proffered by the grammaticalization literature, Heine's (2007) account, resting on the four interrelated processes of desemanticization, extension, decategorialization, and erosion, is perhaps the most successful in synthesizing advances in understanding achieved by multiple strains of grammaticalization research, and is thus granted special attention below.

The intent of the following discussion is to provide a guided introduction to several interlocking and complex phenomena commonly discussed under the "grammaticalization" label, and is not an attempt to bind those phenomena by a set of restrictive or categorical definitions. While a degree of reductivism is to some extent inherent in any theoretical distillation, care must be taken to recognize the processes and products discussed here as both complicated and multi-stepped, and to not equate the observation of salient characteristics with definitional reification.

To ground the following excursus in an illustrative example, I will consider each of the four processes of desemanticization, extension, decategorialization and erosion in relation to one of the rare examples of grammaticalization in Arabic for which a reasonably complete chain of development is historically documented: that of the Egyptian (Cairene) Arabic future tense marker *ħa-*. Sporadically documented specimens of sixteenth and seventeenth century Egyptian Arabic (Kahle & Jacob 1930; Davies 1981) evidence occasional uses of the active participle form *rāyiħ* 'going' with a following imperfective verb to indicate a future tense value, presumably with a source in the use of motion verbs (participial and otherwise) with a subordinated finite verb to indicate a purposive adjunct to the action, a construction still active in the dialect today (Woidich 2006a). Descriptive linguistic sources of Egyptian Arabic dating to the turn of the twentieth century, such as Vollers' (1895) sketch grammar and the phrasebook

composed by Elias and Elias (1981; compiled 1899), show a regular future marker *raħ* preposed to the imperfective verb, which in turn comes to be replaced by the modern Egyptian future prefix *ħa-* (Abdel-Massih et al. 2009). This series of changes, leading from 1) the lexical word *rāyiħ* 'going' through 2) a version of the same entailing an abstract, grammatical meaning of future tense to 3) a reduced and increasingly obligatory form *raħ* used with purely grammatical value and finally 4) a bound morphological marker *ħa-*, represents a prototypical example of a grammaticalization chain, displaying simultaneous evidence of the four processes described by Heine taking place across multiple stages of evolution.

Desemanticization involves the loss of concrete lexical ("content") meaning and the corresponding rise in abstract grammatical function associated with the use of a given item in particular contexts. This often represents the first observable stage of grammaticalizing change, and, as its name suggests, primarily concerns the semantic content of the item rather than its incidence, form, or syntactic behavior. Indeed, evidence of desemanticization is the chief factor leading to the identification of some of the earliest examples of grammaticalized *rāyiħ*, as in the following sixteenth century attestation:

1) 16th century Egyptian Arabic (Kahle & Jacob 1930: 35)
   rāyiħ a- rūħ li- n- nasʿrānī
   go.PTCP 1SG-go.IPFV to-DEF-Christian
   'I'm going to go to the Christian.'

The use here of *rāyiħ* to modify the lexical verb *rāħ* 'go', from which it is itself derived, clearly demonstrates a weakening of the item's original lexical content, as the reading of

the phrasing would otherwise be unmotivatedly redundant. Instead, *rāyiḥ* is better understood as lending a particular grammatical value to the verb, in this case one of imminent futurity (the speaker goes on to ask the friends he is speaking with to relay a message to his brother, who is due to arrive soon following his imminent departure).

Desemanticization is closely coupled with extension, namely the novel use of a grammaticalized item in pragmatic usage contexts where it was not previously employed. While desemanticization consists of a change to semantics or function, extension is defined as a change in incidence. The use of the form in new contexts is inextricable from its acquisition of novel meaning, as it is the latter which prompts the former to occur: as the semantics of the item become less specified and more abstract, its use becomes applicable to a wider variety of contexts/constructions. An example of this tandem occurrence of desemanticization and extension lies in the generalization of grammaticalized *rāyiḥ* from imminent future meanings/contexts to general future meanings/contexts. Davies (1981) observes that sixteenth/seventeenth century *rāyiḥ* only appears to indicate imminent future actions, as in (1) above, or the following (glossed as IMMF):

2) 17th century Egyptian Arabic (Davies 1981: 241)
ʔanā rāyiḥ a- ɣannī ʕalē-h
SBJ.1SG IMMF 1SG-sing.IPFV on -GEN.3MSG
'I'm going to sing about it.' (and proceeds to sing)

In cases where an explicit time adverbial or other context indicates that the action will not immediately follow but will instead occur at a point of more distant future, a bare imperfect form (i.e., without *rāyiḥ*) is used in these texts:

3) 17th century Egyptian Arabic (Davies 2016: 94)
bukra *ti-ʃūf* *mā* *a-* *ʕmal*
tomorrow 2- see.IPFV REL 1SG-do.IPFV
'Tomorrow you'll see what I'm doing.'[2]

In the turn of the century texts, however, as in modern usage, use of the later forms *raħ* and *ħa-* in the presence of a non-immediate temporal adverb is fully acceptable, and – at least in modern Egyptian – obligatory for a reading of time-specified future (Davies 1981). This change, exemplified below, signals a clear extension of acceptable usage contexts for the grammaticalized form (glossed as FUT) to include more general future contexts:

4) 19th century Egyptian Arabic (Elias & Elias 1981: 157)
rāħ *yi-gī* *bukra*
FUT 3- come.IPFV tomorrow
'He'll come tomorrow.'

5) Present-day Egyptian Arabic (Abdel-Massih et al. 2009: 268)
ħa- *yi-rgaʕ* *bukra*
FUT-3- return.IPFV tomorrow
'He'll return tomorrow.'

A third essential process of grammaticalization, decategorialization, describes the changes by which a grammaticalized item comes to lose the characteristic

[2] The remainder of the passage makes clear that *bukra* 'tomorrow' here is to be interpreted literally, and not as the metaphorical use 'some day'.

morphosyntactic properties of its source's original word class, for example word order freedom or agreement inflection. Though the sparsely documented, oldest stratum of examples for grammaticalized *rāyiḥ* attests only instances governed by singular masculine subjects, it is likely that it continued to follow the same agreement patterns as the active participle that served as its source and maintained inflected forms *rāyiḥ* (MSG), *rāyḥa* (FSG), *rāyḥīn* (CPL), in light of the fact that such a pattern still applies to the reduced form recorded three centuries later by Vollers (1895): *raḥ* (MSG), *raḥa* (FSG), *raḥīn* (CPL). Already by the end of the nineteenth century, however, Vollers notes that such agreement marking is not obligatory, and that invariant *raḥ* is often employed with feminine and plural subjects. The latter pattern presages modern Egyptian usage, in which *ḥa-* displays no trace of a former adjectival agreement pattern. This development is indicative of the item's gradual withdrawal from its former word class and loss of corresponding morphosyntactic properties, a clear case of decategorialization associated with grammaticalization.

The fourth process considered by Heine and others to be a key component of grammaticalization is erosion, referring to the gradual reduction and lenition of phonological form beyond what is accounted for by regular sound change. A wide range of incremental and sporadic changes serve to link original *rāyiḥ* to (almost maximally) eroded *ḥa-*. The sixteenth/seventeenth century examples (Kahle & Jacob 1930; Davies 1981) attest *rāyiḥ* (also interpretable as *rāyḥ*, given the orthography); the turn of the century sources show a variety of forms, with *raḥ* the most commonly used but slightly less eroded *rāḥ* evidenced by Elias and Elias (1981) and more eroded *ḥa-* noted as an alternative realization by Vollers (1895). Modern descriptions such as Abdel-Massih et al. (2009) identify both *ḥa-* and an even further lenited variant, *ha-*. These developments, none of which are attributable to regular sound change, may be arranged in the following

progression: *rāyiħ* > *rāħ* > *raħ* > *ħa-* > *ha-*. Cross-linguistically, similar examples of phonetic erosion are commonly observed alongside the processes of desemanticization, extension, and decategorialization described above in cases of change leading to diachronic advance along the grammaticalization cline.

The ubiquity of this composite set of changes across the languages of the world, combined with the comparative dearth of examples demonstrating the opposite progression, has led theories of grammaticalization to be strongly linked to notions of unidirectionality, in other words that change along the cline occurs only in the direction of more lexical to more grammatical and not vice versa. While most authors admit the existence of counterexamples and no longer insist on an absolute formulation of this hypothesis, the recognition of a strong unidirectional tendency remains integral to the understanding of grammaticalization on both empirical and theoretical grounds; for many, as is seen been below, it is what warrants the identification of grammaticalization as distinct from other acknowledged types of diachronic grammatical change.

Hopper and Traugott (2003) join a number of prominent grammaticalization specialists in striking what might be described as a conciliatory position in their promotion of reanalysis, a well-understood standby of "classic" diachronic investigation, as the primary mechanism by which grammaticalization is actuated. In this dissertation I adopt a somewhat stronger stance, and join Haspelmath (1998) and others in asserting that reanalysis, though doubtless a central means through which individual grammaticalization processes occur, is not successful in accounting for the empirically verifiable unidirectional tendency described just above. Logically speaking, reanalysis, defined as a change to the underlying structural description adduced for an existing linguistic string, ought to equally facilitate movement between any two links of the observed chain of grammaticalization regardless of direction: the same abductive

inferences which lead to a grammatical word being reanalyzed as a clitic or a clitic as an affix equally allow the reanalysis of the affix as a clitic, or the clitic as a grammatical word. And indeed such reanalyses do occur as the rare but much touted examples of "degrammaticalization," those exceptions which demonstrate that bidirectional change is in fact possible and which make recourse to an additional motivating force (or forces) to explain the observed unidirectional tendency of grammaticalization all the more pressing. The strongest candidate put forward for such a force is the set of universal cognitive and communicative principles common to the human mental faculty, inclusive of but extending well beyond the realm of language. Though a complete description of this set is far from possible (or warranted) in the present context, individual principles and capabilities which have been referenced in relation to grammaticalization processes include the concrete to abstract directionality of metaphor (Claudi & Heine 1986), the attribution of pragmatic intensity to novelty (Lehmann 2015), and tendencies toward automation and habituation through repetition (Bybee 2003). These cognitive universals would interact with mechanisms like Hopper and Traugott's reanalysis to nudge the results of grammatical change in a particular direction along the grammaticalization cline; additionally, they have the further benefit of providing an account for the pervasive occurrence of grammaticalization worldwide, in a manner that does not appear to be linguistically or culturally constrained.

The observations made here on the nature of grammaticalization may be summarized in defining the phenomenon as follows: Grammaticalization is the change or series of changes through which a given linguistic item shifts from lexical to grammatical or grammatical to more grammatical in nature. It is characterized by the four interrelated processes of desemanticization, extension, decategorialization and erosion. While actuated through processes known to other domains of linguistic change, it is not

coextensive with them, but is set apart by its direct relation to human cognitive and communicative universals. This definition is intended in all subsequent reference to grammaticalization in the present investigation, unless otherwise specified.

1.3.2 Language Contact and Dialect Contact

Theoretical discussions of language and dialect contact are broad and far-ranging, extending well beyond the purview of the present investigation. I therefore confine myself here to describing aspects of language/dialect contact studies most directly relevant to the questions considered by this dissertation, namely those relating to the phenomenon of structural borrowing/convergence. More precisely, I intend the areal diffusion of linguistic "pattern" – underlying syntactic and semantic constructs – as opposed to that of linguistic "matter" consisting of discrete elements of lexicon or morphology composed of actual strings of phonological material (terminology following Matras & Sakel 2007). In the following paragraphs, perspectives on such borrowing phenomena are presented both from scholars of contact between distinct languages and those of interaction between closely related linguistic varieties.

In the field of language contact studies, Gumperz and Wilson's (1971) demonstration of the diffusion of syntactic and semantic structures across the languages of Kupwar opened contemporary conversation on the topic of pattern borrowing. Previously viewed by many as a highly constrained, exceptional occurrence (e.g. Meillet 1912; Weinreich 1953), the phenomenon was further "mainstreamed" by its inclusion in Thomason and Kauffman's seminal (1988) formulation of language contact processes, and is at present largely accepted as a potential outcome of language contact. However, serious debate remains regarding the mechanism(s) by which the diffusion of structural

features across language occurs. The account proposed by Winford (2003) is by and large representative of current views on the topic and focuses heavily on the role of bilingual individuals in actuating structural transfer from one language to another. Without arriving at a definitive answer, Winford centers his debate around the dichotomy formulated by van Coetsem (1988; 2000), exploring the question of whether structural diffusion is best viewed as the product of direct borrowing (reflecting, in van Coetsem's terms, recipient language agentivity) or something more akin to substrate influence (rooted in source language agentivity). In many respects, the question is effectively one of awareness. The first mechanism would entail speakers of one language, having learned an additional language, more or less consciously replicating structures of the latter while conversing in their original idiom; these replications are subsequently available to be picked up and propagated throughout the remainder of the speech community by sociolinguistic means, just like any other linguistic innovation. The second mechanism would consist of speakers of one language unwittingly carrying structural elements of their native tongue over to an additional language they have learned, which, should these speakers carry sufficient weight in numbers or social prominence, have the potential to become conventionalized and disseminate through sociolinguistic means to monolingual speakers.

A potential flaw in the applicability of van Coetsem's and Winford's recipient language/source language dichotomy is its logical insistence on assigning speakers' linguistic "allegiance" to one and only one of the interacting languages involved (further conflating within this the assumption that that language will necessarily be the speaker's dominant one). This is problematic in that it leaves no room for any established fact of individual bilingualism, namely the balanced, simultaneous bilingual: under Winford's scheme, any such individual must be arbitrarily affiliated with one and only one of the

interacting languages in question, even if such a determination cannot be objectively made on purely linguistic grounds. Both Matras (2009) and Ross (2007) offer accounts which overlap to a great extent with Winford's but do not rely on a binary dominance judgment, and thereby remain consistent with current understandings of the forms in which bilingualism can manifest on an individual level as well as a societal one. Matras proposes that the diffusion of linguistic structure takes place via a means broadly similar to that described by Winford's first alternative, that of conscious transfer of linguistic pattern, in this case identifying the creative impulses of proficient bilinguals as the primary motivators of change. Critically, structural convergence in Matras' account is not driven by any subconscious cognitive need but rather the active desire for communicative innovation. Ross takes an opposing perspective, echoing Winford's second alternative in asserting that structural diffusion – though enacted by a similar population of highly proficient bilinguals – is instead the product of effort-reducing processes operating below speakers' level of awareness. Both authors convincingly argue their positions, and each is able to reference several detailed examples in which their respective accounts indeed appear preferable to any obvious alternative.

In light of this state of debate, it is relevant to note that there is no theoretical necessity that the two schools of interpretation described above be mutually exclusive, or that the occurrence of conscious borrowing in some instances negate the existence of subconscious interference in others (or vice versa). In Winford's words, "[in] order to find explanations for contact-induced changes in structure, we need to consider specific cases of contact and examine their social settings and the dynamics of language use by the groups involved" (2003: 64). In other words, the analyst of pattern diffusion must keep her/his ear to the ground and follow the social and linguistic evidence where it leads

in each case, whether that conclusion be one of active creativity à la Matras or passive linguistic interference as predicted by Ross.

Studies of structural diffusion in the context of contact between closely related dialects are far less developed than those stemming from the rich debate in the language contact literature cited above. Historically, the diffusion of structural features has largely been of secondary interest to dialectologists and sociolinguists, whose main interest has lain in the description of lexical and phonological variation; even when the field's collective attentions have extended to morphosyntactic phenomena, emphasis has generally been placed on the documentation of variable surface forms of structural features rather than treating underlying syntactic or semantic frameworks as isoglosses themselves. In addition to simple scholarly inertia, the reasons for this state of affairs are likely twofold. First, there stands the simple fact that pattern-based variation across closely related varieties is often minimal, at least when compared to the vast differences one encounters when operating cross-linguistically, and leaves lexical and phonetic/phonological features to operate as the primary points of sociolinguistic and taxonomic distinction (Wolfram & Schilling-Estes 1998; Siegel 2010). Secondly, the study of purely structural features can prove methodologically thorny for both the traditional dialectologist and the variationist sociolinguist. For the former, so often diachronically oriented, the lack of overt phonological material renders the application of comparative reconstruction questionable at best and undercuts the utility of the structural feature to her/his broader documentary undertaking. For the latter, the Labovian principles of accountability and equivalence are significantly more complicated to satisfy when dealing with structural material as opposed to lexical or phonological variants, and extensive qualitative work is often required in order to circumscribe the variable context

which is so essential to any meaningful interpretation of variationist findings (see Davydova 2014 for further discussion).

Thus deprived of a substantial supporting literature, the limited number of scholars of dialect contact who venture into the structural realm often find themselves forced to rely on theoretical understandings drawn from contact linguistics rather than their own discipline proper. Hence we find statements such as the following by Szmrecsanyi, who argues "that by and large what is true for language contact must also be true for dialect contact, and hence there is no good reason why morphosyntax should not diffuse geographically" (2014: 102). This may not be so straightforward a recourse as it seems, however, when it comes to explicating specific instances of pattern sharing, for as we have just seen the field of language contact studies is far from unified in agreeing on the mechanisms underlying the sharing of structural content. If we are to accept a hypothesis of continuity between dialect contact and language contact in this area, then, of which type of language contact are we speaking? Of the two major proposals described above, split on the issue of awareness of change, the fact that clear sociolinguistic patterning can be shown in numerous cases of structural diffusion through dialect contact scenarios (e.g., Britain 2007) would seem to indicate some degree of conscious, intentional borrowing, such that the adoption of specific features is able to be modulated/manipulated by speakers along sociological lines. In rough terms, this would be in line with the equivalent language contact process described by Matras (2009). We find less evidence for cases of structural dialect contact operating as interference below the level of awareness to pattern with Ross' (2007) proposal, but the concept has been discussed in cases of full-fledged second dialect acquisition – in other words acquisition of a complete new code rather than a single novel feature – in a manner analogous to substrate influence (Siegel 2010).

This lack of clear consensus (as well data upon which to build consensus) highlights the role of the present investigation in moving to clarify the aspects of the questions presented above. Should the Arabic pluriform data prove consistent with a theory of contact-induced structural change which favors one of the two schools of interpretation debated for language and dialect contact, this would be important evidence relating to debates in both of these fields as well as reflect on their precise relation to one another.

1.3.3 Contact-Induced Grammaticalization

As previously stated, in this dissertation I approach the Arabic pluriform data with reference to Heine and Kuteva's (2003; 2005) model of contact-induced grammaticalization (henceforth CIG). This model, synthesizing the two modes of inquiry described above, represents the phenomenon by which "a grammaticalization process … is transferred from the model (M) to the replica language (R)," without corresponding transfer of an actual phonological form (2003: 539). As paraphrased and clarified by Law, this occurs when one language, "the 'replica language,' develops a feature observed in another language, the 'model' language, but goes through a path of universal development using resources internal to the replica language" (2014: 151). Specifically, I intend here what Heine and Kuteva label contact-induced replica grammaticalization, in which the grammaticalization pathways utilized in the two languages are congruent throughout the trajectory of development, and do not merely converge at the end result of those pathways. Such an effect is proposed to be actuated according to the following model (Heine & Kuteva 2003: 539):

1. Speakers of language R notice that in language M there is a grammatical category M_x.
2. They develop an equivalent category, R_x, using material available in their own language (R).
3. To this end, they replicate a grammaticalization process they assume to have taken place in language M, using an analogical formula of the kind $[M_y > M_x] = [R_y > R_x]$.
4. They grammaticalize R_y to R_x.

This proposal of contact-promulgated grammaticalization trajectories shared across linguistic varieties is presaged by Bisang's (1998) observation of the potential synergy between grammaticalization, which he – as per Bybee et al. (1994) – views primarily as a construction-based process, and structural convergence of the type labeled by Ross (2006; others) as metatypy. The possibility for the areal diffusion of grammaticalization processes is also recognized and described by Dahl in the form of what he identifies as "gram families," consisting of groups of "grams [grammaticalized items] with related functions and diachronic sources that show up in genetically and/or geographically related groups of languages, in other words, what can be assumed to be the result of one process of diffusion" (2001: 1469).

In their more elaborated formulation of CIG, Heine and Kuteva (2005) draw heavily on Dahl's theorizations, though they diverge from him in two critical ways. First, they are significantly more conservative than Dahl in identifying examples of the phenomenon, insisting on corroborating evidence of language contact rather than inductively inferring its occurrence given genetic relatedness or proximity. Second, they do not necessarily attempt to link multiple replications of the same pathway into a single

process of diffusion but instead prefer to treat them as individual instances of contact between two participating languages.

Heine and Kuteva's model is largely formulated in the context of contact between genetically distinct languages. Dahl, on the other hand, sees the occurrence of CIG in cases of contact between related language varieties as generating the bulk of evidence for the phenomenon, stating that "in the majority of all such cases [of areally diffused grams], the languages involved are more or less closely related" (2001: 1469). Heine and Kuteva are generally more wary of such identifications; critically, though, their reasons for adopting this position are methodological rather than theoretical. They actively choose to rely on the principle of genetic patterning as "an empirically well-founded tool for identifying cases of contact-induced linguistic transfer" (2005: 33-34), essentially betting that examples of CIG occurring between otherwise unrelated languages are relatively easier to spot and rigorously demonstrate. Regarding the overall occurrence of the phenomenon, however, they state that "genetic relationship is entirely irrelevant" (2005: 184), and that CIG may occur between related languages just as it does between unrelated ones. They are, however, more careful than Dahl to set apart cases attributable to inheritance of any stage of the grammaticalization chain from a common ancestor, which could lead to a superficially similar result not in fact dependent on any degree of contact. Along the same lines, Law (2013) reminds us that when dealing with closely related languages the possibility of drift or typological poise precipitating parallel development rises dramatically in likelihood; thus, the analyst must be stringent in linking proposed cases of CIG to cross-linguistically attested paths and parameters of grammaticalization and not to the local idiosyncrasies of a given language family or subgroup.

A prominent alternative account to the contact-induced grammaticalization proposal has been put forward by Ross (2007), who prefers to reformulate the hypothesis as one applicable to more general processes of bilingual calquing which he views as the necessary precursor to the more dramatic phenomenon of linguistic convergence he identifies as metatypy. Under Ross' conceptualization, rather than speakers of Language M identifying the startpoint and endpoint of a grammaticalization process in Language R and inferring an evolutionary trajectory linking the two, they instead comprehend the full functional range of an existing construction M_x and consequently expand/narrow/adapt the functional range of a corresponding construction R_x to more closely match that of the former. According to the specific new functions to which R_x has been adapted, it may secondarily undergo processes parallel to those which Heine and Kuteva attribute to grammaticalization but which Ross views as the natural result of increases in frequency and automization stemming from the adoption of a new function or function(s).

While Ross' proposal is not by necessity antithetical to the proposal put forward by Heine and Kuteva, it does differ from it in one critical respect: namely, his conviction that "one cannot reasonably argue" for the active role of speakers in processes of pattern borrowing, and that contact-induced grammatical changes are instead "largely driven by effort reducing practices of which speakers are only marginally aware" (Ross 2007: 135). While such assertions of the primacy of psycholinguistic factors are not unusual in discussions of bilingualism and contact between unrelated languages, they are opposed by views of language change espoused by a number of prominent theorists in the subfields of dialect contact and grammaticalization theory. Though the accessibility of structural features for manipulation and borrowing has been observed to vary along multiple axes (for background, see Silverstein 1981), the role of socio-pragmatically motivated borrowing practices – at times demonstrating a significant degree of awareness

of linguistic structure – has long been acknowledged in the variationist and broader dialectological literature (Labov 1972; Chambers & Trudgill 1998). This understanding has itself served as the catalyst for the acceptance of similar precepts among scholars of grammaticalization. Lehmann (1985) identifies the very source of grammaticalization in synchronically variable linguistic behavior, and recognizes the same potential for intentional, innovative creativity present in that context. For Lehmann, such a creative drive is the primary impetus for language-internal grammatical change:

> To the degree that language activity is truly creative, it is no exaggeration to say that languages change because speakers want to change them … they do not want to express themselves the same way they did yesterday, and in particular not the way that somebody else did yesterday (Lehmann 1985: 315).

Building from Lehmann, I assert that in scenarios of language or dialect contact the innovating speaker may very well wish to express herself/himself the same way somebody else did yesterday, if the means of expression used are novel to a distinct speech community with which the speaker finds her/himself interacting today. That abstract understanding of another language or dialect's structures may serve as fodder for innovative, creative expression is not a new proposal (see Matras 2011). Critically, because this occurs via an active impulse to replicate rather than a passive acquiescence to burdens of processing load, it is logical that it would take place through the same mechanisms of abductive reasoning, methaphor extension, and automization as other modes of socio-pragmatically driven linguistic innovation, i.e. grammaticalization as it is broadly known. Thus, any abrupt dismissal of contact-induced grammaticalization as beyond the active capabilities of speakers is not warranted, and this type of creative,

social and interactive language change remains a valid proposal to account for the specific outcomes of linguistic development observed in the present study.

In the following section, I transition from discussion of CIG as a diachronic process to that of CIG as a synchronically apparent outcome, the perspective upon which the ensuing investigation is based. Specifically, I propose an operationalization of the above theories of CIG which will assist me in evaluating its explanatory power as a mechanism behind the Arabic pluriform developments; my intent is that the results of this evaluation may then be placed alongside those of other existing case studies in an attempt to assess the broader suitability of the model. In addition to this operationalization, which takes the form of a three-part heuristic, I also discuss data sources and other methodological topics as they relate to the theoretical points described in the previous subsections.

1.4 METHODS

This dissertation combines synchronic dialect geography with techniques of diachronic reconstruction to examine the utility of CIG as a lens for interpreting the pluriform developments in modern Arabic. For each of three functional items falling under the "pluriform" label, I reconstruct and assess the various grammaticalization pathways evidenced by contemporary forms across a large sample of Arabic dialects. I then map and compare the geographic incidence of these grammaticalization pathways to assess the likelihood of dialect contact as the mechanism behind their propagation and modern distribution.

I draw my data from a sample consisting of eighty-one specific Arabic varieties. Dialects included in the sample have been selected primarily based on their physical

locations, in order to provide sample points constituting as comprehensive a geographic coverage of the modern Arabic-speaking world as possible for the analysis. The genetic structure and internal classification of the Arabic dialect family remains in many respects opaque (cf. Kaye & Rosenhouse 1997; Versteegh 2014), and as such is not a central component of this study's design; as a result of the geographic comprehensiveness just noted, however, all major proposed dialect subgroupings are guaranteed representation in the sample. The only set of Arabic dialects actively excluded from the analysis are a) those not geographically contiguous with the greater Arabophone area, consisting primarily of the scattered Central Asian Arabic Sprachinseln of Uzbekistan, Afghanistan, and Khorasan, and b) those chronologically asynchronous with the remainder of the study's data, such as the existing attestations of the medieval Andalusi Arabic dialect bundle.

The three pluriform developments whose realizations are to be examined in detail are 1) future tense markers, 2) temporal adverbials meaning 'now', and 3) analytic genitive exponents. Of these, the future markers and genitive exponents have been selected because they are among the most commonly cited examples of Arabic pluriform developments in the existing literature (cf. Versteegh 2001). All three have been the subjects of diachronic investigations of more or less global scope. Perhaps most thoroughly, Eksell Harning (1980) presents a near exhaustive accounting of forms and function of the genitive exponent in modern Arabic dialects. Her intent in the work is more properly descriptive than theorizing, and as such she does not speculate as to a unified developmental mechanism; however, the depth and breadth of her documentation and reconstruction are virtually unmatched among comparable diachronic studies of specific Arabic grammatical categories. Leading among recent examinations of the future tense markers is Stewart's concise but insightful (1998) article sketching the range

of plausible grammaticalization pathways attested by forms observed cross-dialectally, as well as working to situate a novel analysis of the Classical Arabic future tense morphemes in relation to this broader context. The remaining pluriform element examined in this study, the temporal adverb 'now', has been chosen from among the remaining variables due to its reflexes' typically transparent etymologies (thus both easing and strengthening the interpretation of likely grammaticalization pathways) and for the ready availability of data, representing as it does so basic a linguistic operator as to be an essential component of even the most concise grammatical sketches or lexical collections.

As the initial step of analysis in each of the three cases, proposed etymological sources are identified for all reflexes of a given feature represented in the sample. These etymologies are determined through the application of the comparative method, followed by reference to existing Arabic lexical resources. First, putatively cognate reflexes are grouped and a plausible source form (or set of possible source forms) generated through standard techniques of linguistic reconstruction. This operation, though in many respects straightforward, is at times complicated by the occurrence of irregular phonological and morphological changes – perhaps to be anticipated in a dataset constructed on the basis of its resemblance to grammaticalization data (see §1.3.1, above). In reconciling these irregularities, fruitful use is at times made of any apparent occurrence of grammaticalization chains linking together more and less reduced reflexes of a given item within the context of a single dialect to help establish details of an earlier etymological form, as described on a case-by-case basis in the following chapters. Once a viable reconstructed proto-form has been generated, it is compared to formally similar and semantically linkable etyma attested in Arabic lexical sources in search of a plausible match. While a defensible etymology for a given set of cognate items may still be

proposed in the absence of such a direct match, it shall be seen to be possible in nearly all cases considered below, providing substantive corroboration of conclusions reached via the application of comparative reconstructive techniques alone; in cases where a direct lexical match is not identified, related etyma are sought at the more theoretically removed level of the triconsonantal root. Contemporary lexical sources consulted include dictionary-length treatments of modern dialect lexica available for several major regional varieties (e.g., Hinds and Badawi (1986) for Egyptian; Woodhead and Beene (1967) for Iraqi) as well as the geographically comprehensive *Wortatlas der arabischen Dialekte* presented by Behnsteht and Woidich (2012; 2014). Proposed etymologies are provided temporal and genealogical depth through further reference to the Classical Arabic lexicographical tradition, whenever possible in the form of the oldest known Arabic lexical collection, the *Kitāb al-ʿAyn* of al-Farāhīdī (d. 791). In cases where proposed source etyma are attested in such works, they may be understood as reconstructable at least to the nearest common ancestor of Classical Arabic and the modern dialects in which they occur. In cases where etymologies have been proposed by early authors, they are considered and often adopted, but in all cases independently verified via the method described here.

Once viable etymologies have been identified for all reflexes, I commence evaluation of the results as consistent with a history of development through CIG. This evaluation rests on analysis via a heuristic I have developed to address three questions, each posed based on the discussion of CIG theory included above, with particular reference to the conclusions of Heine and Kuteva (2003; 2005) and Law (2013). This heuristic is not intended to define CIG as a linguistic phenomenon, but rather to assist in the identification of its results as they present in a large-scale dialectological dataset like that utilized in the present investigation. As a function of its design, it is by intention

overly rigid in its requirements: failure to meet the heuristic's conditions does not categorically eliminate a given innovation as a possible product of CIG, but simply indicates that such a determination has not been made to the level of rigor deemed necessary for an unambiguous confirmation of the CIG hypothesis in this context. Instead, success in meeting said conditions identifies those cases in which clear examples of more or less prototypical grammaticalization share sufficient semantic and functional similarity and geographic proximity to render CIG both a plausible and a parsimonious account for their modern occurrence. The heuristic is as follows, and details of its constituent elements are discussed below:

1. Are the proposed grammaticalization pathways identified in the data attested cross-linguistically, and thus potentially representative of universal processes? Are these proposed paths consistent with described general principles of grammaticalization?
2. Are pathways represented by more than one distinct realization, here defined as consisting of distinct but synonymous etyma?
3. If more than one realization is identified for a given pathway, is the distribution of realizations consistent with diffusion via dialect contact?

The first condition relates to the need, emphasized by Law, to confirm any proposed instances of CIG as bona fide examples of grammaticalization. This status is evaluated with reference to the four major component processes discussed as typical of grammaticalization earlier in this section, namely desemanticization, extension, decategorialization and erosion; all proposed examples of grammaticalization addressed in this investigation are examined for evidence of these phenomena. This requires

examination of an individual item's semantic value, pragmatic usage contexts, morphosyntactic behavior and phonetic form in comparison to its diachronic source. Taken as additional, corroborating evidence of grammaticalization is the common recurrence of any proposed grammaticalization pathways in the cross-linguistic descriptive literature. While such recurrence is by no means conclusive evidence of grammaticalization status in and of itself, it does bear directly on the theorized universal aspect of grammaticalization as rooted in the general human cognitive faculty, and common attestation worldwide thus supports an identification of grammaticalization based on the presence of the component processes described just above. Entailed in acceptance of this criterion, however, is the cautionary observation that the likelihood of coincidental occurrence of a given pathway in two varieties under study rises with increased cross-linguistic frequency, a fact which renders the heuristic's additional limiting conditions all the more pertinent.

The role of the heuristic's second condition is to distinguish the potential results of CIG from those of other historical linguistic processes. To illustrate in an example, the GO-based futures deriving from **māʃī* 'going' found in Sousse, Tunis, and Djidjelli along the North African coast may well have evolved as the result of successive processes of CIG affecting the three dialects in turn, but they could just as likely be the products of shared inheritance, a single shared innovation, or classic ("matter"-based) dialect borrowing. The GO-future form attested in neighboring Algiers, though – based on the distinct but synonymous **rāyiħ* 'going' – is not linkable to the first three by any of these latter processes, though it might be via CIG. The choice to restrict identifications of CIG to those cases involving distinct realizations thus potentially underestimates the incidence of the phenomenon, ignoring as it does the possibility of three **māʃī*-based grammaticalizations like Djidjelli *māʃī*, Tunis *bāʃ*, and Sousse *māʃ* having been replicated

separately. In some cases, detailed examination of relevant sound changes and relative progression along grammaticalization clines might prove elucidatory; however, the consistent application of such nuanced analysis is not feasible given the scope of the present investigation, and as such I proceed with the perhaps overly conservative standard described here.

In identifying distinct but analogous realizations of a given grammaticalization pathway, I address an aspect of CIG not thoroughly discussed in the existing literature (though highlighted in the previously referenced treatment by Ross (2007)), namely the degree of functional correspondence between replica and model constructions. A major strength in the formulation of CIG is precisely its ability to account for imperfect matches in function in this area as individual effects stemming from each item's independent, language-internal grammaticalization process (see Bybee, Perkins & Pagliuca 1994). In this regard, the CIG framework could be seen to predict the occurrence of such functional variation, whereas in other models of structural transfer, such as bilingual calquing, it must be explained away as a secondary phenomenon. Still, it holds that any items allegedly related via processes of CIG must be characterized by a certain degree of functional similarity if one is to propose the modeling of one based on the other. Thus, for my part, any linked developments I propose below are contingent on viable synchronic or diachronic correspondence of functional detail between individual tokens considered.

The heuristic's third question addresses the areal cohesion/contiguity of a given grammaticalization path's occurrence, which here, following Heine and Kuteva (2005) as well as studies of linguistic areality more generally, is crucial in judging the possible spread of the feature through contact. Though dialect contact has been observed to occur in discontinuous fashion (for example, features "hopping" from one urban area to another

while passing over intervening rural territory – see Trudgill (1986) for discussion), establishing such potential cases requires a degree of sociohistorical scrutiny not possible on a broad scale in this project, and I will therefore strike a cautious stance once again by taking geographic adjacency as a necessary condition for claiming the occurrence of contact-based diffusion.

In plotting the geographic distribution of forms for each item, I adopt standard conventions of isogloss mapping as practiced in the dialectological and language contact literature; for general sources, see Hudson (1999) and Britain (2014). Isoglosses drawn on each map are used to circumscribe the geographic occurrence of a particular realization of the specific item in question. Though at the end of the day such a practice remains very much an interpretive one, overall I endeavor to remain "close" to the data in my delineations. In cases when this is not possible, such as in areas where consistent coverage is lacking in the geographic sample – the Libyan desert, for example, or the interior of the Arabian Peninsula – the boundaries shown are necessarily less certain. When relevant, isoglosses which must be drawn in less described regions such as these will be informed by more general characterizations present in the Arabic dialectological literature. For example, it is generally accepted that the complex of dialects known by the name Ḥassānīyah extends fairly homogeneously across Mauritania to Mali (Taine-Cheikh 2006); thus, while reasonably complete descriptions are only available for Nuakchott on the Atlantic coast and Timbuktu on the far interior end of this zone, the previously observed uniformity will lead me to posit continuity across the intervening territory, in the absence of any specific information to the contrary. Colors of isoglosses are used to represent membership in a particular group of parallel grammaticalization pathways (e.g., NOW < THIS + TIME), while different styles of line within each color – solid, dotted, dashed, etc. – represent the individual, distinct realizations which make up

these groups (e.g., ‘Awāmra *halwagit*, Aleppo *hallaq* < **hā* ‘this’ + **l-waqt* ‘time’ as against Aden *daħīn* < **ðā* ‘this’ + **l-ħīn* ‘time’). Groups marked by color are thus roughly equivalent to Dahl’s gram families, and the realizations marked by color and line style akin to individual grams. The available sample dialects for each item will be represented as points on the map, and any overlap in isoglosses surrounding these points will indicate that multiple forms are attested in the data for the same relevant location. After mapping the sample data for a given pluriform development of the three described, I review the results in relation to the adjacency/contiguity criterion just discussed.

This three-part heuristic for evaluating the suitability of CIG as an account for the data collected here is applied separately for each of the three pluriform items examined. Heine and Kuteva (2005) list two additional prerequisites for identifying examples of the phenomenon, which I will discuss here, as relates to all three items at once. The first of these is the requirement that contact between speakers of the varieties in question be known to have taken place, or at least represent a viable historical possibility. In the case of the Arabic-speaking world, that such contact has occurred and continues to occur between speakers of neighboring varieties is evidenced by the nature of the modern Arabic dialects as a gradated continuum (Versteegh 2014). Further, the social history of the region has for centuries been one of large, well-integrated political and economic systems, with constant and extensive interaction of individuals and groups across the Arabic-speaking territories and beyond (Hourani 1991; Robinson 1996). Additionally, the reader will note that natural boundaries such as the Libyan Desert and the Red Sea have not been treated by default as inhibiting social contact, as historical and anthropological studies have repeatedly shown this not to be the case (Burr & Collins 2008; Power 2012). In short, this study assumes that if two Arabic dialects border one another then contact between them is both possible and plausible.

Secondly and finally, Heine and Kuteva insist that proposed cases of CIG not involve attestation of all or part of the relevant grammaticalization pathway in a common ancestor to the two interacting languages, as in such a case the shared pathway between the two varieties may be attributable to common inheritance rather replication via contact. Frustratingly, no viable common ancestor to the modern Arabic dialects has yet been identified or adequately reconstructed; on the generally accepted unsuitability of Classical Arabic as a hypothesized progenitor to the modern varieties, see Owens (2006), Abboud-Haggar (2006), *inter alia*. As such, for the purposes of the present study I must proceed with this condition unaddressed. As per Dahl, however, I do not see this fact as overly damaging to the final analysis, and perhaps in an indirect way the pursuant results will contribute toward one day filling this gap in knowledge in the field at large.

My choice of data sources for this project has been informed by other recent inquiries of similar scope (e.g., Magidow 2013). I rely primarily on descriptive grammars, both in the form of grammar sketches and full-fledged reference grammars. The core of my eighty-one point sample is provided by the thirty-two point-specific descriptions included in the *Encyclopedia of Arabic Language and Linguistics* (Versteegh Ed. 2006), the unified format and balanced geographic spread of which adapt quite readily to this project. In addition to descriptive grammars, dialect atlases are also utilized when available – in relation to this work I have consulted atlases compiled for Egypt (Behnstedt & Woidich 1985), Yemen (Behnstedt 2016), and the greater Syrian area (Behnstedt 1997). As a third set of resources, annotated dialectological texts have been published for several Arabic speech communities not yet the subjects of more complete descriptive works; when these permit the inclusion of sample points not otherwise represented on the study's maps, they are drawn upon for primary data. In the

specific case of Aswan, gaps in information available in published sources have been filled via personal communication with experts in local dialect forms of the region.

Besides such mundane concerns as availability, three major interests affect my selection of specific source materials. The first of these is perhaps the most straightforward: does a given reference include the necessary data to inform analysis of as many as possible of the three pluriform items under investigation? In this respect, the descriptive grammar is the most dependable of the three major resource types listed above – all items selected represent basic grammatical operators likely to be included in such a work, even if it is merely a cursory sketch. The items selected are somewhat less typical of the content of dialect atlases, though still present to a degree. Perhaps least reliable of all as a source for data on the specific set of items examined is the annotated dialectological text, as the presence/absence of a given feature is thoroughly dependent on the content of the text itself representing an appropriate context for its occurrence.

The second concern is for geographic specificity of the data a source provides – as reliable dialectological mapping is a key pillar of this study, the ability to link a given data point to a particular location is necessary. In this respect, the dialect atlas is king for obvious reasons, as this identification will already have been made in exacting detail by the original author. Annotated texts and descriptive grammars, however, are frequently less precise with regard to the geographic origin of the data they comprise: texts vary greatly concerning the amount of ethnographic data provided to describe participating speakers, and grammars are by definition abstractions of linguistic patterns found across a more loosely or more tightly defined community of practice. In the Arabic context more particularly, a secondary issue arises from the fact that many grammars and text compilations are not explicitly linked to a given location but rather to a social unit assumed to be of dialectological significance, such as a tribe or religious sect. In such

cases, secondary sources have been consulted in order to ascertain the most likely geographic implications of the linguistic data provided.

Thirdly, sources must be evaluated in light of the degree of detail they provide in their presentation of the functional specificities of the pluriform items in question. Subsumed within the categories under investigation are myriad variations of values – a future tense marker is not a future tense marker is not a future tense marker – and these details are relevant to a proper interpretation of each item's individual development and the subsequent identification of cross-dialectal parallels, thus affecting the evaluation of data vis-à-vis the heuristic as described above. Despite their geographic precision, dialect atlases are notoriously problematic in that they typically provide little to no context or analysis to describe the specific functional aspects of the features whose physical distribution they present. In an ideal scenario, it might be hoped that a descriptive grammar would serve significantly better, its stated purpose being to describe and exemplify just such particularities of usage. In reality, however, Arabic dialect grammars often fall short of this goal in their tendency to adopt a uniformitarian, "cookie cutter" approach to documenting dialectal variation (i.e., a future tense marker *is* a future tense marker *is* a future marker, just in different surface guises). Diversity of form and function within given varieties is likewise often left unanalyzed, variant forms presented side by side with little to no discussion of semantic, pragmatic, or social distinctions in their usage. Thus, in this respect, it is perhaps annotated text sources which provide the richest detail for my inquiries, as multiple examples of contextualized usage enable the secondary researcher to arrive at and support analyses of his or her own which are lacking in other genres of published materials.

A final issue I wish to bring to light in the discussion of data sources is one not generally discussed in the broader Arabist literature, but which I believe to have risen to

increased relevance in recent decades – that of chronology. In embarking on a cross-dialectal study of a scale similar to that of the present dissertation, not only are one's Arabic data sources likely to encompass a large geographic area but also potentially a span of well over a century in publication dates. Thus, the apparently synchronic "snapshot" portrayed by an isogloss map based on such data may well comprise information representing upwards of three of four distinct generations of speakers, a sufficient time depth to permit appreciable change in the distribution of dialectal features to occur (Trudgill 1986). While such a state of affairs is likely unavoidable at the current state of research in Arabic, at least without considerably thinning the available geographic representation so crucial to projects like this one, it is important to hold this concern under consideration when interpreting the accumulated results of the research.

At this point, I proceed to apply the methods described over the course of the preceding subsection in evaluating each of the three pluriform developments in turn, assessing them by means of the established heuristic for evidence of their evolution through processes of CIG, as based on their contemporary linguistic characteristics and geographical distribution. The first case considered is that sketched briefly as part of this chapter's introductory pages, the modern Arabic future tense markers.

Chapter 2: Future Tense Markers

2.1 Introduction

In the following chapter, I evaluate the forms and distribution of modern Arabic future tense markers[3] for evidence of the occurrence of CIG according to the heuristic described in §1.4. Among the most commonly cited examples of the Arabic pluriform developments, the historical development of the future tense markers has previously been examined by Stewart (1998) in a grammaticalization framework, though without reference to the dimension of contact or other historical relations between the distinct etyma proposed. The etymologies and paths of development identified here are largely consistent with those identified by Stewart, though extend to include a larger sample; in cases of disagreement, the route of diachronic analysis is presented in sufficient detail to justify the departure. Rubin (2005) also briefly considers the development of future tense markers in his broader account of grammaticalization processes across the Semitic family, though his coverage of forms is far from comprehensive and some of the grammaticalization processes he proposes suffer from inaccurate identification of initial etymology.

In the following sections, I first present the future tense markers of the current sample on the basis of etymological origin, describing in detail the phonological and morphosyntactic changes leading from each etymon to its suite of attested modern forms. The focus of this review will be a synopsis of raw dialectological data presented in the sample, without significant theorization beyond the processes of etymologization described in §1.4, working to link contemporary reflexes to proposed lexical sources.

[3] I use the term "marker" here as a generic cover for the morphosyntactically diverse set of items in the sample used to indicate future tense value, ranging from inflected verbs to radically reduced clitics. I also acknowledge that in the Arabic context the categories of tense and aspect are often closely interwoven (see Brustad 2000), and that "tense" is in many cases an oversimplification of a complex combination of values.

Once these basic etymologies are established, I proceed to assess the resulting data in accordance with the steps of the described heuristic for evidence of CIG. In this stage I will consider first the validity of pathways identified as defensible examples of grammaticalization, followed by the search for multiply attested pathways and the analysis of their geographic incidence as indicative of areal diffusion. The chapter concludes with a discussion of the results of these analyses and a cumulative evaluation of the role of CIG in the development of the future tense markers.

2.2 Attested Forms by Etymology

I provide below a comprehensive account of the source etyma for future tense markers found in the current sample. Accompanying each etymon is an identification of its lexical source and a brief description of its geographic occurrence (to be considered in greater detail in subsequent sections). This is followed by a full recounting of the phonological and morphosyntactic changes attested by each etymon's modern reflexes, as well as mention of any remaining noteworthy phenomena.

**rāyiḥ*

Algiers *rāḥ* [+][4, 5]	Khuzestan *raḥ*	ʻAwāmrah *laḥ ~ ḥa-*
Baghdad *raḥ*	Malta *ḥa-*	Hit *raḥ*
J-Baghdad *ɣaḥ*	Mecca *rāḥ ~ ḥa-*	Basra *rāyiḥ* [+]
Beirut *raḥ*	Negev *rāyiḥ*	Cherchell *rāyiḥ* [+]
Bʻērāt *rāḥ ~ raḥa ~ ḥa-*	Sinai *rāḥ ~ raḥ ~ ḥa- ~ ha-*	Djidjelli *rāyiḥ*
Cairo *ḥa- ~ ha-*	Tripoli *ḥā-*	Benghazi *ḥa-*
Damascus *raḥ ~ laḥ ~ raḥa ~ laḥa ~ ḥa-*	Tozeur *ḥa-*	ʻAbābdah *rāḥ ~ ḥa-*
Jerusalem *rāyiḥ ~ rāḥ ~ ḥā-*	Kadugli *ha-*	Dellys *rayəḥ* [+] *~ ḥa-*
Amman *rāyiḥ ~ rāḥ ~ ḥa-*	Aswan *ḥa-*	Jisr az-Zarqa *rāyiḥ*
Khartoum *ḥa-*	Kharga *ha-*	

Table 3: Future Tense Markers from **rāyiḥ*

Future tense markers deriving from **rāyiḥ* ‘going’ are found across a broad east-west swath of the Arabic-speaking world, extending from south and central Mesopotamia through much of the Levant and the upper Nile Valley and stretching a thin arm across North Africa to cover significant Algerian territory in the west. The form **rāyiḥ* is

[4] From this point forward, whenever a form is provided in the format <Location *form*> (e.g., Algiers *rāḥ*) without further citation information, its source is that indicated as the primary data source for the relevant location in the study’s appendix. Names of localities are provided in accepted English language spellings, while those of tribes/ethnic groups and those locations without a customary English representation are transliterated using Library of Congress romanization conventions. The prefix <J-> to a location (as J-Fez), indicates that the variety cited is one ascribed to a Jewish minority population at the given location; similarly, B-Kadugli indicates forms attributed to the city’s Baggara community specifically.

[5] In this and all following tables, the symbol [+] indicates that the form indicated displays the adjectival or verbal agreement morphology characteristic of its etymological source. It should be noted that the total number of such inflecting forms may be underrepresented here, as a number of references fail to provide explicit characterizations on this point.

originally an active participle of the verb **rāḥ* ‘go’, carrying in its lexical form a progressive or expectant state reading of ‘going’. The verb is known from the earliest period of Classical Arabic in a meaning of ‘to journey’ (whether coming or going), appearing as such in the eighth century *Kitāb al-ʿAyn* (al-Farāhīdī), the inaugural lexicon of the classical period. In the modern day, it has semantically bleached to become the generic term for ‘go’ in the vast majority of contemporary Arabic varieties from Egypt eastward, and attestation in Algeria and the western Sudan hint also at a former westerly extension of this range (Behnstedt & Woidich 2014: 14, 16). As a future marker, some operators of this set are recorded as expressing a specific value of immediate future (Algiers *rāḥ*) or future intent (Jerusalem *rāḥ*) while the majority are associated with a meaning of general futurity.

In a number of dialects, reflexes of this form maintain the full phonological and morphological integrity of their source, for example Basra *rāyiḥ*, which displays the expected adjectival gender/number agreement with its subject. Others show loss of agreement inflection accompanied by phonetic erosion and of loss of morphosyntactic autonomy. It is common for such morphologically invariant forms to simplify the internal sequence /āyi/ > /ā/ as in Mecca *rāḥ*, or to further shorten it to /a/ as in Beirut *raḥ*. Though irregular in all cases, a corollary to this change may be seen in the sporadic monophthongization following syncope in lexical forms of this item in some dialects, perhaps following a lowering influence from the following pharyngeal: Aswan *rāyiḥ* ‘going (MSG)’ vs. *rāḥa* (< **rāyḥa*) ‘going (FSG)’. Some varieties also show sporadic deformation of the initial /r/ > /l/, as in Damascus *raḥ* ~ *laḥ*, and in one case a regular change of /r/ > /ɣ/ (J-Baghdad *ɣaḥ*). A number also insert a final /a/, as demonstrated by Bʿērī *raḥa*. This can be understood either as a frozen instance of the feminine agreement marker *-a* (and thus more properly deriving from an earlier **rāyḥa*) or as a case of

spontaneous anaptyxis triggered by the pharyngeal /ħ/, for which there is phonotactic precedent in a number of dialects (though not the full set for which the /a/-final forms are attested). In either case, this development would also seem to be the source of the vowel in the frequently occurring, highly reduced forms resembling Cairo *ħa-*. Though the dialectological literature shows a linguistically unmotivated tendency to treat such forms as recent, fashionable Cairene imports across virtually the entire non-Egyptian portion of their range (e.g., Souag 2005; Dickins 2006), little to no tangible synchronic or diachronic evidence has been cited to support this position and the wide, contiguous distribution of the forms in question (stretching from Amman *ħa-* to Dellys *ħa-*) does not indicate a path of development qualitatively different from those of the majority of isoglosses considered in this study. Among these reduced forms, lenition of initial /ħ/ > /h/ is not uncommon (Kharga *ha-*), and in one case a long vowel /ā/ is attested (Tripoli *ħā-*).

**ɣādī*

Casablanca *ɣādi ~ ɣa-*	Goulimine *ɣadi*	Tetuan *ɣadi*
Saoura *ɣādi*	J-Fez *ɣadi*	Marrakech *ɣādi ~ ɣa-*

Table 4: Future Tense Markers from * *ɣādī*

Future tense markers based on **ɣādī* are common throughout Morocco and adjacent areas of Algeria. Formally an active participle, in its lexical sense **ɣādī* is a verb of motion with a meaning of 'going'; the participial form is most common, and finite forms of the verb **ɣadā* are only rarely attested in contemporary sources (Heath 2002). The lexical verb is known from Classical Arabic sources with the meaning 'go,

set off in the morning' (al-Farāhīdī, *Kitāb al-'Ayn*), and has bleached to a general meaning of 'go' in dialects of modern day Morocco and Algeria as well as varieties spoken in scattered locations throughout the Arabian peninsula (Behnstedt & Woidich 2014: 14, 18). These two pockets of usage at opposite ends of Arabic's modern geographic expanse could plausibly be viewed as the product of common innovation at a very early stage or as two distinct semantic developments; the question, though, does not bear directly on the history of the future tense markers considered here, which are restricted to western North Africa.

All forms deriving from **ɣādī* in the present sample are inflectionally invariable in their use as future tense markers, though Heath (2002) describes Moroccan dialects for which **ɣādī* as a tense marker shows the same adjectival agreement pattern it does in its participial use. Alongside unreduced reflexes, some varieties contain shortened, cliticized forms which display diminished morphosyntactic autonomy and have lost their entire second syllable to phonetic erosion: Casablanca *ɣādi* ~ *ɣa-*. It may in reality be that such reduced forms are more frequent than is represented here, as the source consulted for the dialects of Tetouan, Goulimine and the Jews of Fez (Heath 2002) simply states that variable reduction to **ɣa-* is common throughout the Moroccan area but does not provide point-specific detail for these three locations. Differences in vowel length between, for example, Goulimine *ɣadi* and Saoura *ɣādi* are attributable to regular sound change (or, potentially, differing phonemic analyses between researchers of North African dialects).

**māʃī*

Tunis *bāʃ ~ biʃ*	Djidjelli *māʃi*	Anjra *māʃ*
J-Tunis *māʃ*	J-Fez *masi ~ mas*	Mateur *bāʃ*
Sousse *māʃ ~ bāʃ*	Tetouan *maʃ*	

Table 5: Future Tense Markers from **māʃī*

Future tense markers deriving from **māʃī* 'going' are found in two distinct geographic pockets, one centered on north-central Morocco and the other encompassing the majority of Tunisia and adjacent areas of eastern Algeria. In its lexical form, **māʃī* is an active participle of the verb **maʃā* 'go'. The source verb **maʃā* is common to virtually all dialects considered in the sample and to Classical Arabic (al-Farāhīdī, *Kitāb al-ʿAyn*) with a meaning of 'walk'; the verb has extended its meaning to 'go' in a large contiguous region encompassing North Africa east of Egypt and the greater Sudanic area, as well as additional attestations in eastern Oman (Behnstedt & Woidich 2014: 14, 16).

None of the **māʃī*-based forms in the present sample maintain participial agreement marking in their use as indicators of future tense. In terms of phonological form, relatively unreduced reflexes are encountered (e.g., Djidjelli *māʃi*) but more commonly the final /ī/ is lost, resulting in forms such as Anjra *māʃ*, Jewish Tunis *māʃ*. Relatively common in the Tunisian area is the sporadic deformation of initial /m/ > /b/, as observed in Mateur *bāʃ* and in variable pairs like Sousse *māʃ ~ bāʃ*. In one case, variable vowel reduction is reported (Tunis *bāʃ ~ biʃ*), and in the Judeo-Arabic dialect of Fez /ʃ/ becomes /s/ via regular sound change: J-Fez *masi ~ mas*. Differences in vowel length should be viewed either as the product of regular change or as transcriptional choices on the part of the descriptions' original authors.

**sāyir*

Malta *sɛyyɛr* [+] ~ *sɛr-* ~ *sɛ-*		

Table 6: Future Tense Markers from **sāyir*

Alone in the sample, Maltese attests a future marker deriving from **sāyir*. Lexically, this represents an active participle of **sār*, an archaic verb meaning 'go' which survives in Maltese primarily in suppletive forms in the inflectional paradigm of unrelated modern *mār* 'go' (including the active participle). The verb is also known in minority and suppletive variants recorded at various locations in North Africa and is still in productive use in the southern portion of the Arabian Peninsula (Behnstedt & Woidich 2014: 14, 18), seemingly reconstructable in its lexical form to a common ancestor of the these varieties and Classical Arabic, in which it is attested from an early period (al-Farāhīdī, *Kitāb al-'Ayn*).

Malta *sɛyyɛr* is the regular participial derivation of this root, and in unreduced form maintains adjectival agreement patterning even as a future tense marker: *sɛyyɛr* (MSG)/*sɛyra* (FSG)/*sɛyrin* (PL). Phonetically eroded and cliticized forms *sɛr-* and *sɛ-* are also in common use, which have lost syntactic autonomy and do not host agreement morphology.

**yabɣā ~ yabɣī*

Bahrain *b-*	Misīrīyah *bi-*	Benghazi *yibbi* [+] ~ *yib-* ~ *ibi-*
Khartoum *bi-*	Kabābīsh *bi-*	Abha *b-*
Kuwait *b-*	Rubāṭāb *be-*	Kordofan *b-*
Sudayr *yabi* [+] ~ *ab-* ~ *b-*	Ḥarb *yabɣa* [+] ~ *yaba* [+] ~ *ba-*	Dhofar *bā-*
Tripoli *bə-* ~ *b-*	Shukrīyah *bi-*	Nigeria *b-*
Hadhramaut *bā-*	Abéché *b-*	Āl Wahībah *bi-*
Kadugli *bi-*	Aden *bā-*	Al-Khaburah *b-*
Fezzan *bī* ~ *b-*	Abu Dhabi *b-*	

Table 7: Future Tense Markers from **yabɣā ~ yabɣī*

Forms deriving from the imperfective verb **yabɣā ~ yabɣī* serve as future tense markers across a large portion of the Arabic-speaking world, stretching from the Arabian Peninsula in the east across the Red Sea to the greater Sudanic area and then northward through modern Libya. In their lexical forms, these etymologically cognate verbs carry a meaning of 'want', current throughout the Arabian Peninsula and adjacent areas of the Syrian desert as well as Libya and parts of Tunisia (Behnstedt & Woidich 2014: 498, 500-502). This area coincides with the distribution of modern future tense markers from this source, with the exception of those attested in the broader Sudanese area; the medial position of the Sudan, though, between areas of western Arabia and Libya – where lexical **yabɣā ~ yabɣī* is prevalent – points to the possibility of an earlier contiguous distribution which was later disrupted, perhaps by the appearance in the Sudan of the innovative

**ydawr* volitives described in more detail below. Historically, cognate *yabɣī* is attested with a meaning of 'desire, seek' from early in the Classical Arabic period (al-Farāhīdī *Kitāb al-ʿAyn*), and the modern lexical forms are etymologically related to, but morphologically distinct from, North African **bɣā* 'want', which is considered separately below.

It should be noted that the variations in source forms noted here are largely the products of morphological rather than phonological change: dialects differ widely in their selection of /a/ or /i/ as the vowel of the imperfective prefix, and the distinction between final /ā/ and final /ī/ reflects a difference in inflectional class. Many forms also evidence a sporadic phonological change of the sequence /bɣ/ > /bb/ (> /b/), which has occurred already in the lexical form of the item. As a future tense marker, the initial syllable is most often lost; many Arabic varieties attest a nearly maximally reduced realization, such as Abu Dhabi *b-*, while others retain evidence of original final vowel quality, as in Khartoum *bi-*, Dhofar *bā-*. Dialects are attested which display multiple stages of this development simultaneously active as future tense markers, ranging from fully inflected verbal forms showing little or no phonetic erosion to increasingly reduced and invariant reflexes of this type, as witnessed in Ḥarb *yabɣā ~ yabā ~ ba-*, Benghazi *yibbī ~ yib- ~ ibi-*. In the cases of neighboring Nigeria *b-* and Abéché *b-*, the semantics of the form would seem to have bleached considerably, and it is now a marker of general imperfective aspect and survives only as an allomorph preceding vowel-initial agreement prefixes (the pre-consonantal realization being *Ø-*).

**biddu ~ widdu*

Cilicia *baddu* [+] ~ *baddi-* ~ *bad-*	Negev *widdih* [+] ~ *d-*	Aleppo *bəddo* [+] ~ *bdo* [+] ~ *b-*
Damascus *bəddo* [+] ~ *b-*	Sinai *widdih* [+] ~ *biddu* [+]	Bdūl *b-*
Jerusalem *biddo* [+] ~ *b-*	Soukhne *b-*	Jebel Ansariye *baddo* [+] ~ *bado* [+] ~ *b-*
Amman *biddo* [+]	Khawaytnah *bəddu* [+]	Jisr az-Zarqa *b-*

Table 8: Future Tense Markers from **biddu ~ widdu*

Future tense markers arising from **biddu ~ widdu* are identifiable throughout the broader Levantine area. These arise from a prepositional construction which in a number of dialects has acquired pseudoverbal morphosyntactic properties (see Brustad 2000) and come to represent the primary means to indicate volition and desire, reflecting an original **(bi-)widd-u* 'it is (in/by) his will'. In its lexical form, this construction is the generic expression for 'want' throughout the Levant and adjacent Northwest Arabia, with a few isolated cognates attested also in the greater Yemeni area and the Persian Gulf region (Behnstedt & Woidich 2014: 504-505); there is also historical documentation of its use in the Egyptian area (Vollers 1895). The reconstructable time depth of the source construction extends at least to the Middle Ages, as the collocation *bi-widd* is recorded with a meaning of 'wish, want' in the Classical Arabic lexicon *Lisān al-ʿArab* (Ibn Manẓūr) in the thirteenth century, citing earlier sources; it is not present in the eighth century *Kitāb al-ʿAyn* (al-Farāhīdī), though the single lexeme *widd* is found with a meaning of 'love, desire'.

Though some future tense uses of the Northwest Arabian area may be traced to an original **widdu* 'want' (e.g., Negev *widdih*), the majority of reflexes clearly derive from **biddu*. These forms at times reflect local variations in the quality of the internal vowel common to both lexical and grammatical uses of the item; these may be products of regular sound change, as /i/ > /ə/ in Damascus *bəddo*, or spontaneous developments like /i/ > /a/ in Cilicia *baddo* ~ *baddi-* ~ *bad-*. Changes to final vowels typically represent regular developments in the form of clitic personal pronouns, though they may be the result of epenthesis as syntactic autonomy is lost (as in Cilicia *baddi-*, above). Phonetic erosion is common, with loss of material not ascribable to regular sound change. This may occur regressively from the end of the item, as demonstrated by the Cilician forms just seen, but more commonly begins with the collapse of the vowel and coda of the initial syllable: Aleppo *bəddo* ~ *bdo* ~ *b-*, Jebel Ansariye *baddo* ~ *bado* ~ *b-*. At their most reduced, such forms are often superficially indistinguishable from the products of **yabɣā* ~ *yabɣī* discussed above; the presence of several fully attested chains of development for both families of items, however, prompts the identification of separate origins, the dividing line between the two bordering zones presumed (for want of further data) to follow the coinciding lexical divide in terms for 'want'. In one case deriving from **widdu*, Negev *widdih* ~ *d-*, the irregular loss of phonological material takes a progressive trajectory, affecting the initial segments of the word along with its suffixed inflectional morphology.

It is of note that in some varieties, grammaticalizations of **biddu* ~ *widdu* operate alongside other markers of future tense to designate a more specified value: Damascus *bəddo* ~ *b-*, for example, is reported to denote a modal value of possible or planned future, as opposed to the **rāyiḥ*-derived forms *raḥ* ~ *ḥa-* which indicate a higher

degree of certainty or expectation. In other dialects, reflexes of **biddu ~ widdu* would appear to have further bleached and extended to a value of more general futurity.

**yiʃā*

Sana'a *ʃa-*		

Table 9: Future Tense Markers from **yiʃā*

A number of Yemeni dialects contain future tense markers deriving from **yiʃā*, including one represented in the present sample: Sana'a *ʃa-*. In its lexical form, the imperfective verb **yiʃā* has a meaning of 'want'. Some researchers have preferred to link *ʃa-* and related forms to a source in the verb **yiʃtī*, also meaning 'want', but the quality of the final vowel is strong evidence against this identification. Classical Arabic attests a cognate verb *yaʃāʔ* in a meaning of 'will, wish' (al-Farāhīdī, *Kitāb al-ʿAyn*), but modern **yiʃā* in a general meaning of 'want' is confined to a narrow coastal strip of western Yemen and far southwestern Saudi Arabia, proximate to but not including modern Sana'a, where reflexes of **yiʃtī* presently dominate in lexical usage (Behnstedt & Woidich 2014: 504-506).

In Sana'a *ʃa-*, the opening syllable of the source form is lost and its final /ā/ is shortened, neither of which are attributable to regular sound change. It is notable that Sana'a *ʃa-* is used only with the first person singular verb, while forms deriving from **ʕād* (described below) are used to mark future tense for other numbers and persons. Such person-based asymmetry in future tense marking is not unusual cross-linguistically (Bybee, Perkins & Pagliuca 1994), and in some Yemeni varieties reflexes of **yiʃā* are used with all persons (Behnstedt 2016). In the case of Sana'a *ʃa-* specifically, though,

this person specificity means that its ultimate source might more properly be identified with **aʃa:* 'I want', though the distinction has little bearing on the present analysis.

**ydawr*

Nouakchott *ydōr* [+]	Mali *ydawr* [+]	

Table 10: Future Tense Markers from **ydawr*

Varieties belonging to the Ḥassānīyah dialect complex of Mauritania and neighboring Mali are recorded as utilizing a form of **ydawr* with a following imperfective verb to denote a value of intentional future. On its own, **ydawr* is a lexical verb meaning 'want'. This lexical meaning of 'want' is encountered only in Saharan and Sub-Saharan Arabic varieties, characteristic of the greater Sudan and the Ḥassāniyyah zone; even more specifically, the use of **ydawr* as a finite verb is typical of Ḥassāniyyah and the Western Sudanic area only, as in the Sudan proper an active participle form **dāyir* is more commonly used (Behnstedt & Woidich 2014: 498, 502). The lexical meaning 'want' is not associated with this verb in Classical Arabic, where it means 'go around (in circles)' (al-Farāhīdī, *Kitāb al-ʿAyn*); this seems to have subsequently semantically extended to 'search, look for' (more common cross-dialectally) and further to 'want' in the Saharan/Sub-Saharan zone alone (Behnstedt & Woidich 2014: 502).

Relatively little in the way of phonological or morphosyntactic change has affected these forms in their usage as future tense markers; they remain fully inflected, and in one case display a regular monophthongization of */aw/ > /o/: Nouakchott *ydōr*.

**bɣā*

Saoura *ba-*	Marrakech *bɣa* [+] ~ *ba-*	

Table 11: Future Tense Markers from **bɣā*

Dialects of southern Morocco and southwestern Algeria occasionally attest forms of **bɣā* expressing a future tense value. In its lexical form, **bɣā* is a syntactically perfective/semantically imperfective verb meaning 'want', and is in fact the perfective form of the same lexeme **yabɣā ~ yabɣī* discussed above. The fact, however, that the local means for expressing 'want' in the region is question is based on the perfective stem argues for a separate identification of these two reflexes for the purposes of etymology, as is supported by the phonological shape and morphological behavior of these items as future tense markers. The use of a formally perfective stem **bɣā* to express lexical 'want' is typical of a large region of western North Africa, inclusive of but (Behnstedt & Woidich 2014: 498, 501-502) extending far beyond the two sample points attesting **bɣā* as a future marker, and would seem by all accounts to be a regionally specific innovation without a precise correlate in the Classical Arabic record. The development may be viewed as consistent with the expanded declarative/performative function of the perfective noted for Moroccan Arabic varieties as compared to other dialects (Brustad 2000).

In one of the two attestations present in the sample, Marrakech *bɣa ~ ba-*, fully inflected verbs forms are in use as markers of future tense alongside a reduced and cliticized reflex which has sporadically deleted medial /ɣ/ and no longer displays verbal agreement inflection. In the remaining example, only the more reduced version is found: Saoura *ba-*.

**ħattā*

Anatolia *tə-* ~ *də-*	Cyprus *tta-* ~ *ta-*	Khawaytnah *ta-*
Mosul *də-*		

Table 12: Future Tense Markers from **ħattā*

Reflexes of **ħattā* are used to indicate future tense in areas of northern Mesopotamia, the coastal Levant, and Oman. Attested cross-dialectally as a preposition 'until', **ħattā* is commonly used with a verbal complement – in these areas and beyond – to create a purposive construction; this usage is attested in Classical Arabic as well, with the purposive reading further reinforced by a following subjunctive:

1) Classical Arabic (Wright 1898: 30)
   sir -tu ʔilā l- kūfat -i ħattā ʔa- dxul -a -hā
   go.PFV-1SG to DEF-Kufa-GEN PURP 1SG-enter.IPFV-SBJV-OBJ.3SGF
   'I journeyed to Kufa that I might enter it.'

Given the evident continuity in preverbal syntax and irrealis semantic value, it is directly from this usage that the function of **ħattā* as a future tense marker most likely evolved.

The forms in the sample are proclitics of which the first syllable has been sporadically lost, commonly accompanied by degemination of */tt/: Cyprus *tta-* ~ *ta-*, Khawaytnah *ta-*. In the case of Mosul *də-* and Anatolia variant *də-*, the remaining */t/ undergoes further lenition in the shape of irregular voicing, and the final vowel is additionally reduced.

*ʕād

Sana'a *ʕā-*	Tozeur *ʕa-*	Aswan *ʕa-*

Table 13: Future Tense Markers from **ʕād*

Future tense markers traditionally identified as based on **ʕād* 'return' are attested in three locations in the cross-dialectal survey: Yemen, Upper Egypt and interior Tunisia. The Classical Arabic cognate *ʕāda* 'return' is known from the earliest recorded stages of the language (al-Farāhīdī, *Kitāb al-ʿAyn*), but in modern varieties 'return' is more commonly represented by reflexes of **rajaʕ*. Contemporary **ʕād* is attested as a lexical verb, though, in areas of western North Africa, Mesopotamia, Yemen and the Persian Gulf (Behnstedt & Woidich 2014: 79-80) and is used as part of a negated verbal construct meaning 'no longer' in both Syria and Kuwait (Brustad 2000). This depth of attestation and far-flung geographic distribution recommend the reconstruction of the verb **ʕād* to a node preceding both Classical Arabic and the modern dialects, thus potentially explaining through shared inheritance its occurrence as a future marker in two varieties (Tozeur and Aswan) where its lexical form is not otherwise known.

The forms found in Egypt and Tunisia are highly reduced – Tozeur *ʕa-*, Aswan *ʕa-* – and thus difficult to ascribe definitively to a specific source. In both of these dialects the markers in question vary with a 'go'-derived future *ħa-* and could thus plausibly represent an erosion of the latter in the form of a sporadic voicing lenition of */ħ/ > /ʕ/ (in Aswan, at least, the two phones are known to alternate in cases of conditioned devoicing). A further viable source for Aswan *ʕa-* may lie in local *ʕāyiz* ~ *ʕāwiz* 'want'. In the lack of further information, it is difficult to make a decisive determination. At least in the case of the Yemeni forms, however, an origin in **ʕād*

seems clear, as reduced forms such as Sana'a *ʕā-* display an allomorph *ʕad-* in prevocalic contexts.

**lāhī*

Nouakchott *lāhi*	Goulimine *lahi ~ la*	Mali *lāhi ~ lā*

Table 14: Future Tense Markers from **lāhī*

In the Ḥassānīyah dialects of Mauritania, Mali and far southern Morocco, the future tense marker derives from **lāhī*; such forms express a value of general futurity and are more commonly attested than the **ydawr*-based forms described above. Original **lāhī* would itself appear to represent an active participle form of the verb **lhā*, uniquely here 'busy oneself' (Taine-Cheikh 2006). Cross-dialectally, and even in Malian varieties, this verb is generally transitive and means 'occupy, distract (s.o.)' (cf. Hinds & Badawi 1986; Woodhead & Beene 1967; Heath 2004); the Classical Arabic record, though, provides precedent for an intransitive reading, with *Kitāb al-ʿAyn* (al-Farāhīdī) defining *lahā* as 'be captivated by, obsessed with' and *Lisān al-ʿArab* providing a definition of 'divert oneself, play' (Ibn Manẓūr). This derivation has been preferred over an alternative formally apparent etymology, that of previously described **rāyiħ*. While precedent does exist for the sporadic transformation of */r/ > /l/ and */ħ/ > /h/ (cf. Damascus *laħ*, Cairo *ha-*), these realizations typically exist in variation with unlenited forms, of which none are present in the available data for any of the three locations discussed here. Additionally, and perhaps more significantly, the spontaneous generation of final /i/ implied by such an account is phonologically unmotivated, save perhaps

through the influence of a following *y*- third person inflectional prefix. For these reasons, a reconstructed source in **lāhī* has been deemed more probable here.

Forms derived from **lāhī* generally maintain their phonological integrity but do not display expected participial agreement marking, as invariant Nouakchott *lāhi*. In two of three varieties in the sample containing reflexes of **lāhī*, the item is reported to undergo variable, irregular deletion of the final syllable, as in Mali *lāhi* ~ *lā*. The reflexes in Goulimine *lahi* ~ *la* are subject to regular changes governing vowel quantity distinction typical of the Moroccan area.

2.3 ANALYSIS

Having presented etymologies for the complete set of future tense markers attested by the Arabic dialect sample, I now turn to analysis of the data by means of the three-step heuristic presented in §1.4. The first subsection below relates to conditions (i) and (ii) of the heuristic, the occurrence of theoretically verifiable grammaticalization processes and the organization of the specific processes into multiply-attested paths, and the second addresses the geographic incidence of those pathways and their constituent developments.

2.3.1 Evaluation of Grammaticalization Status and Multiply Attested Pathways

In this subsection, I determine whether the various derivations presented in §2.2 constitute valid examples of grammaticalization and which if any of them may be grouped together as representatives of the same grammaticalization pathway. This is accomplished with reference to cross-linguistically recognized pathways for the grammaticalization of future markers as identified in the general theoretical literature as

well as to Heine's (2007) four parameters of grammaticalization, as described in detail in §1.3.1. Each proposed pathway attested by the Arabic data is presented below, followed by independent assessments of all individual derivations assignable to the path and displaying recognized characteristics of grammaticalization.

2.3.1.1 Futures from 'go' (FUT < GO)

As has been seen, developments of future tense markers from forms of lexical verbs meaning 'go' are well represented in the Arabic data, comprising four distinct etymological sources. As a grammaticalization path, FUT < GO is among the most commonly identified routes for the evolution of future markers worldwide, and is well established by such broad-based studies as Bybee, Perkins and Pagliuca (1994) and Heine and Kuteva (2002) as a major source of grammaticalizations cross-linguistically. In the present sample, four distinct sets of forms of those described above share an etymological origin in a motion verb meaning 'go': **rāyiḥ, *ɣādī, *māʃī* and **sāyir*.

In support of identifying these particular Arabic developments as verified examples of grammaticalization and representatives of the broader, cross-linguistic GO > FUT grammaticalization path, evidence of the four characteristic processes of desemanticization, extension, decategorialization and diffusion may be observed as follows, summarized in the Table 15. The numerical values listed refer to the proportion of reflexes for each etymon which show evidence of a given process, details of which are discussed further in the following paragraphs.

Source	*Desemanticization?*	*Extension?*	*Decategorialization?*	*Erosion?*
**rāyiħ*	47/47	47/47	44/47	40/47
**ɣādī*	8/8	8/8	8/8	2/8
**māʃī*	11/11	11/11	11/11	9/11
**sāyir*	3/3	3/3	2/3	2/3

Table 15: Occurrence of Diagnostic Grammaticalization Processes, FUT < GO

Desemanticization is observed across all reflexes of **rāyiħ, *ɣādī, *māʃī* and **sāyir* via a distancing from the source semantics in lexical verbs meaning 'go' in the direction of an abstract value of (more or less general) future tense. In the case of a few reflexes of **rāyiħ* like Algiers *rāħ*, this grammatical function is reported to denote a more specified value of immediate futurity and thus the desemanticization of these items may be viewed as lesser in degree than that of those signifying general futurity, but it is still present. In all cases, these semantic developments combine with corresponding extensions of the relevant forms to novel usage contexts not characterized by the linguistic expression of corporeal motion:

2) Marrakech (Sánchez 2014: 182)
ɣādi t- tfəntər
ɣādi 3FSG-get.fat.IPFV
'She'll get fat.'

3) Jisr iz-Zarqa (Belinkov 2014: 70)
miʃ rāyiħ y-sʿīr wala iʃi
NEG *rāyiħ* 3-happen not.a thing
'Nothing will happen.'

Further, decategorialization is observed through the loss of morphological agreement patterning typical of the participial source etyma; such loss of categorial membership is noted for all reflexes of **ɣādī* and **māʃī* attested in the sample, as demonstrated by the lack of expected number inflection in invariant Anjra *māʃ na-rʒaʕ* (FUT 1-return.IPFV) 'I'll return', *māʃ nə-mʃi-w* (FUT 1-go.IPFV-PL) 'we'll go' (vs. lexical *māʃī* 'going (MSG)', *māʃyīn* 'going (PL)') (Vicente 2000: 74, 107). The same is true of the great majority of forms deriving from **rāyiħ* and in the reduced forms of the Maltese **sāyir*-futures, which dominate their unreduced counterparts in terms of usage frequency (Vanhove 1993). In more advanced cases, decategorialization is also evidenced by the loss of morphosyntactic autonomy associated with the transition from word to clitic. Some varieties contain only cliticized reflexes, such as Khartoum *ħa-*, which are highly restricted in syntactic position and bound to the following verb. Phonetic erosion is observed in the majority of reflexes considered here, as seen in Table 15: while phonologically intact variants do exist in each case, most forms encountered in the sample display evidence of sporadic deletion and lenition not accounted for by processes of regular sound change. Such developments are present across the sample in widely differing degrees, as Mecca *rāħ* and Cairo *ha-* (< **rāyiħ*), with more and less eroded forms often in variation with one another within the same dialect: Malta *sɛr-* ~ *sɛ-* (< **sāyir*). Erosion is noticeably less frequent among reflexes of **ɣādī* than those of the other three source etyma, though irregularly reduced forms like Casablanca *ɣa-* are attested.

In light of these findings, it is justified to identify the diachronic products of **rāyiħ, *ɣādī, *māʃī* and **sāyir* as representing the results of grammaticalization and to group them together under the single grammaticalization pathway of FUT < GO, attested by four discrete realizations involving distinct etyma. As such, both conditions (i) and

(ii) of the heuristic are met for these items, which to this point appear viable candidates for identification as the products of CIG.

2.3.1.2 Futures from 'want' (FUT < WANT)

Similarly widespread in the Arabic dialect data are future tense markers deriving from lexical verbs meaning 'want'. Like the FUT < GO path just discussed, FUT < WANT is an extremely common source of grammaticalized futures cross-linguistically, figuring prominently in the general theoretical literature (cf. Bybee, Perkins & Pagliuca 1994; Heine & Kuteva 2002). As has been observed, multiple distinct etyma meaning 'want' are implicated in the evolution of future tense forms provided by the present sample, including **yabɣā ~ yabɣī, *biddu ~ widdu, *yiʃā, *ydawr* and **bɣā*.

Relating to whether these individual Arabic developments may be confidently identified as examples of grammaticalization, processes of desemanticization, extension, decategorialization and erosion may be observed among the relevant reflexes in the patterns summarized in Table 16.

Source	***Desemanticization?***	***Extension?***	***Decategorialization?***	***Erosion?***
**yabɣā ~ …*	31/31	31/31	27/31	27/31
**biddu ~ …*	22/22	22/22	10/22	9/22
**yiʃā*	1/1	1/1	1/1	1/1
**ydawr*	2/2	2/2	0/2	0/2
**bɣā*	3/3	3/3	2/3	2/3

Table 16: Occurrence of Diagnostic Grammaticalization Processes, FUT < WANT

Desemanticization has occurred in all reflexes as their concrete lexical meaning of 'want' has faded and is replaced by a more abstract value of future tense. In some dialects, this new value shares functional space with additional future tense markers of distinct etymological sources to express a more restricted notion of futurity. Damascus *bəddo ~ b-*, Jerusalem *biddo ~ b-* express a modal value of possible or planned future, as in Damascus *bəddo yūʔaʕ* 'it's going to fall down at any time' (Lentin 2006: 553); Bahrain *b-*, Kuwait *b-*, and Benghazi *yibbi ~ yib- ~ ibi-*, meanwhile, are described as implying an immediate or proximate future time frame, for example:

4) Benghazi (Owens 1984: 156)
   is- sukkar yi-bbī yi-kmil
   DEF-sugar 3- *bbī* 3- be.finished.IPFV
   'The sugar is about to finish.'

In the remaining cases, the items' meanings have further generalized to encompass a general future reading, as for Amman *bidd-ī ʔa-rūħ bukra* (*bidd*-1SG 1SG-go.IPFV tomorrow) 'I will go tomorrow' (Al-Wer 2006: 514). Correspondingly, these reflexes have also undergone extension to pragmatic contexts devoid of, or even contrary to, volition: Benghazi *ni-bbī ni-mrəðˁ* (1-*bbī* 1-get.sick.IPFV) 'I'm about to get sick' (Benkato 2014: 75).

In terms of decategorialization and erosion, the lone reflex of **yiʃā* in the sample, Sana'a *ʃa-*, displays both in good measure, having lost both its verbal agreement properties and its entire first syllable of phonological material. Both attested reflexes of **ydawr*, Nouakchott *ydōr* and Mali *ydawr*, on the other hand, retain verbal person and number inflection in their grammatical usages and show no phonological developments

not attributable to regular sound change: Noukchott *ndōrˁ nəmʃî* 'I'm going to go' (Taine-Cheikh 2006: 248). As for the remaining source etyma, the majority of derivations of **yabɣā ~ yabɣī* and **bɣā* and a significant portion of those of **biddu ~ *widdu* show the results of both decategorialization and erosion, the details of which are included in Table 16. In cases like Negev *widdih ~ d-* and Marrakech *bɣa ~ ba-*, dialects utilize morphologically and phonologically unaltered forms in variation with reduced reflexes which have lost syntactic independence, morphological inflection, and significant amounts of phonological material (unaccounted for by regular sound change). Such overlapping areally concurrent realizations constitute examples of what have been labeled in the literature as grammaticalization chains (Heine, Claudi & Hünnemeyer 1991), and occasionally extend beyond a simple unreduced/reduced pairing to maintain multiple stages of increased decategorialization and/or erosion, as in Sudayr *yabi ~ ab- ~ b-*. In other varieties, such as Āl Wahībah *bi-*, the decategorialized and eroded reflexes exist alone, with no direct attestation of an earlier, unreduced form. It is noticeable that at least variable decategorialization and erosion are present in reflexes of these three etyma for all varieties examined in the sample, even if in some dialects these reduced forms would seem to be described as less frequent variants.

Following from this evidence, it is possible to identify the Arabic reflexes of **yabɣā ~ yabɣī*, **biddu ~ *widdu*, **yiʃā*, **ydawr* and **bɣā* as the products of grammaticalization, representing multiple distinct realizations of the cross-linguistically corroborated grammaticalization pathway FUT < WANT. The generally strong case is perhaps the weakest for derivatives of **ydawr*, which display only two of the characteristic processes associated with grammaticalization, but it is crucial to note that these two processes are precisely those predicted to occur earliest in the development of a grammaticalized form (Heine 2007). It can be posited that the grammaticalization of

**ydawr* is thus analogous to the others examined under this heading but simply younger, and ought therefore to be viewed as an instantiation of the same, multiply-attested grammaticalization path. The full set of forms examined in this subsection therefore meet conditions (i) and (ii) of the heuristic and remain in the running as potential products of CIG.

2.3.1.3 Futures from 'come' (FUT < COME)

Together with the verbs 'go' and 'want' already discussed, a third major source of future tense marker grammaticalization in the languages of the world is composed of verbs meaning 'come' (Bybee, Perkins & Pagliuca 1994; Heine & Kuteva 2002). Given their extremely frequent occurrence cross-linguistically, the path of FUT < COME gives the impression of being underrepresented in the Arabic sample, attested only by reflexes of the single etymon **ʕād* 'return', in this case the venitive semantics of the lexical source qualified with an additional notion of repetition or cyclicity. As discussed in §2.2, two of these three reflexes prove difficult to link definitively to an etymological source in **ʕād* and may instead be relatable to 'go'- or 'want'-based future forms, but in the case of Sana'a *ʕā(d)-* at least the determination can be made with a reasonable degree of certainty.

This ambiguity of source construction aside, the identification of constituent processes of grammaticalization in derivations of **ʕād* proceeds unequivocally, as displayed in Table 17.

Source	*Desemanticization?*	*Extension?*	*Decategorialization?*	*Erosion?*
**ʕād*	3/3	3/3	3/3	3/3

Table 17: Occurrence of Diagnostic Grammaticalization Processes, FUT < COME

Desemanticization occurs with the increased distance of these forms' meaning from concrete, lexical 'return' toward abstract, grammatical future time reference. Pragmatic extension follows in close fashion as the resultant changes in meaning allow these forms to be employed in novel contexts not reliant on actual motion or even those contrary to the venitive directionality of original **ʕād*:

5) Tozeur (Saada 1984: 60)
zōz erʒāl ʕa-ir-rowħ -u ɣadwa
two men *ʕa*-3- go.IPFV-PL tomorrow
'Two men will leave tomorrow.'

All three reflexes of **ʕād* drawn from the sample show evidence of decategorialization, as all display signs of cliticization and none carry the agreement morphology characteristic of their perfective source verb, which exists as lexical 'return' in Sana'a (Watson 1993: 53) and semantically shifted to 'become' in Tozeur (Saada 1984: 60) (comprehensive lexical data for Aswan is not available). Phonetic erosion is also demonstrated in all reflexes in the form of sporadic sound change, progressing to a slightly greater extent in Tozeur *ʕa-* and Aswan *ʕa-* than in Sana'a *ʕā*, *ʕad* / _V, where a less eroded form of the item is maintained in specific allomorphic contexts.

In spite of uncertainty surrounding the derivation of specific reflexes, it seems clear that grammaticalized forms of **ʕād* represent the grammaticalization path FUT <

COME in the Arabic dialect data. Unlike the paths FUT < GO and FUT < WANT, however, the FUT < COME pathway is represented by only a single etymological source. As such, it fails to meet condition (ii) of our heuristic, and – while apparently presenting a valid example of grammaticalization – the forms based on **ʕād* will therefore add no supporting evidence to our evaluation of CIG's role in the development of Arabic future tense markers.

2.3.1.4 Futures from purposive constructions (FUT < PURP)

A further source of future tense markers in the Arabic data involves the grammaticalization of a purposive operator, a path not widely discussed in the cross-linguistic literature on grammaticalization. Intriguingly, the reverse trajectory of PURP < FUT is noted by Bybee et al. (1994), and other less direct semantic connections between the two categories are observed by Heine and Kuteva (2002). The primary sticking point in establishing a well-defined FUT < PURP grammaticalization path would seem to rest in the identification of a clear process of desemanticization, as it is difficult to judge on objective terms precisely which function between FUT and PURP is more concrete/abstract than the other. Despite this ambiguity, the ubiquitous occurrence of extension, decategorialization and erosion in the Arabic **ħattā*-based forms seems to recommend their identification as products of a grammaticalization process. The details are as summarized in Table 18.

Source	*Desemanticization?*	*Extension?*	*Decategorialization?*	*Erosion?*
**ħattā*	4/4* (*see discussion)	4/4	4/4	4/4

Table 18: Occurrence of Diagnostic Grammaticalization Processes, FUT < PURP

Reflexes of **ħattā* in all forms demonstrate extension from pragmatic contexts defined by purpose and intent to those involving more general notions of futurity. This likely proceeded in accordance with a semantic evolution from purposive meaning through a value of future intent to a general future reading. The three stages of this proposed progression may be viewed in the following examples of Khawaytnah *ta-*, the value of first of which is virtually indistinguishable from a purposive reading, that of the second expressing proximate future intent, and that of third devoid of agentive intentionality:

6) Khawaytnah (Talay 1999: 183)
mʃõ, ta-rˤ- rˤūħ n- jīb -ha
let's.go, *ta*-1PL-go.IPFV 1PL-get.IPFV-OBJ.3FSG
'Let's go, we'll go get her.'

7) *ta-Ø- ʕmal -lək ɣada*
ta-1SG-make.IPFV-DAT.MSG lunch
'I'll make you lunch.'

8) *yomēn t -ʃūf -u ta-t- xāf mənn-u*
when 2-see.IPFV-OBJ.3MSG *ta*-2-fear.IPFV from -OBJ.3MSG
'When you meet him, you'll be afraid of him.'

While the extension beyond original purposive contexts is clear, as previously noted it is less than obvious whether this semantic change is to be considered desemanticization. However, if this trajectory is indeed that which has been followed, I assert that the latter stage of development from more specified intentional to less specified general future certainly meets the criteria for the identification of desemanticizing change. Thus, I tentatively propose that desemanticization has occurred for all **ħattā* reflexes alongside the more readily recognizable extension, with the caveat that additional revelations into the precise semantic evolutions of these items may alter this judgment.

Clearer evidence exists of decategorialization and erosion. Decategorialization is observed in the shrinking of allowable complementation types for future tense marking reflexes of **ħattā*, which must be strictly verbal as opposed to the purposive's verbal or sentential – e.g. Khawaytnah purposive *ħatta* [*mā y-tʿīq-ōn y-īj-ōn ʕalē-kam*] (PURP [NEG 3-be.able.IPFV-PL 3-come.IPFV-PL to-GEN.2PL]) 'so that they can't come to you' with sentential complement (Talay 1999: 191) – as well as in the general loss of morphosyntactic freedom associated with the cliticization of all reflexes encountered in the sample. Phonetic erosion via irregular sound change is also widespread, affecting all attested derivatives of **ħattā*, including deletion of the initial syllable, varying degrees of lenition of the cluster /tt/, and reduction of the final vowel, resulting at its most advanced in forms like Mosul *də-*, Anatolia *tə- ~ də-*.

In sum, processes of desemanticization, extension, decategorialization, and erosion may be seen to typify the development of Arabic future tense markers of purposive origin; on their own, these facts may be taken as evidence that the Arabic forms therefore represent the products of grammaticalization. However, this identification is complicated by the fact that direct cross-linguistic corollaries are not

readily apparent; this calls into question the status of this particular evolutionary pathway as linked to universal cognitive processes or tendencies, a characteristic discussed as integral to the working definition of grammaticalization laid out in §1.3.1. However this argument may be decided, it appears to be a moot one in this case. Regardless of whether it meets condition (i) of the heuristic as a bona fide example of grammaticalization, similarly to FUT < COME the grammaticalization pathway FUT < PURP is attested by a sole representative etymon in the Arabic data, and thus does not succeed in condition (ii), the requirement for a multiply attested grammaticalization pathway. The reflexes of **ħattā* will therefore not be further considered in the search for evidence of CIG, regardless of grammaticalization status.

2.3.1.5 Futures from 'busy oneself' (FUT < ACTIVITY)

Three Arabic dialects in the sample utilize a future tense marker appearing to derive from a grammaticalized form of a verb of activity, in this case **lāhī* (active participle of **lhā* 'busy oneself'. Such a path of development, here labeled FUT < ACTIVITY, is apparently not frequent enough to be discussed in the global examinations of the grammaticalization of future tense markers consulted, but perhaps finds at least a single counterpart in the grammaticalized usage of Southern American English *fixing to ~ fixin' a ~ fi'na* to express proximate futurity (Wolfram & Schilling-Estes 1998). Notwithstanding this dearth of obvious cross-linguistic correlates, clear evidence of desemanticization, extension, decategorialization and erosion of the Arabic source form is indicated across the three reflexes, as shown in Table 19.

Source	*Desemanticization?*	*Extension?*	*Decategorialization?*	*Erosion?*
**lāhī*	3/3	3/3	3/3	2/3

Table 19: Occurrence of Diagnostic Grammaticalization Processes, FUT < ACTIVITY

Reflexes of **lāhī* show desemanticization as they depart from their concrete meaning of their lexical verb to express an abstract value of future time reference. Accompanying this process is extension of the forms' use to pragmatic contexts beyond those involving actual physical activity:

9) Nouakchott (Taine-Cheikh 2006: 249)
mā -hu lāhi yə-bki
NEG-OBJ.3MSG *lāhi* 3- cry.IPFV
'He isn't going to cry.'

10) Mali (Heath 2003: 32)
ʔʃən hu lli lāhi y-ʒədd ʕlī -k?
what SUBJ.3MSG REL *lāhi* 3-be.convincing.IPFV upon-2SG
'What is going to be convincing to you?'

Decategorialization is evidenced by all reflexes of **lāhī* present in the sample through the absence of number and gender agreement characteristic of active participles in the relevant dialects. Instead, they occur in invariant form regardless of sentential context:

11) Mali (Heath 2003: 20, 44)

lā lāhi t- rīħ -i fī-ha

NEG *lāhi* 2-rest.IPFV-F in-GEN.3FSG

'You will have no rest from it.'

12) *gāl -ū -lu ʕlann -hum lāhi yə-ktl -ū -h*

say.PFV-3PL-DAT.3MSG COMP-OBJ.3PL *lāhi* 3- kill.IPFV-PL-OBJ.3MSG

'They told him they were going to kill him.'

Finally, erosion may be seen in the reduced forms Goulimine *la* and Mali *lā*, the last syllable of which is deleted via irregular sound change.

Taken together, the collective impression of these findings is similar to that discussed in relation to the products of FUT < PURP just above. While the results of the FUT < ACTIVITY pathway display the component processes of grammaticalization in good measure, their acceptance as such is hampered by the dearth of cross-linguistic parallels indicating a universally applicable path of development – though in this case at least one analogous, unrelated development is noted in Southern American English *fixing to ~ fi'na*. The existence of the latter raises the question of the completeness in cross-linguistic grammaticalization research, and whether lack of attestation in the existing literature is truly evidence against the global occurrence of a given pathway or simply an artefact of incomplete documentation. As was previously the case for FUT < PURP, however, the success or failure of FUT < ACTIVITY to meet condition (i) of the heuristic is preempted by its inability to meet condition (ii). Given that **lāhī* is the only Arabic etymon representing the pathway in the sample, the developmental trajectory shows no clear evidence of potential replication and the above reflexes will consequently not figure in the next phase of the analysis, the assessment of geographic incidence.

2.3.2 Evaluation of Geographic Distribution

The previous evaluation of grammaticalization status and grouping has established two grammaticalization pathways represented by multiple, distinct etymological realizations: FUT < GO and FUT < WANT. Together, products of these pathways' constituent etyma account for the great majority of forms encountered in the Arabic data and geographically cover nearly the entire extent of the modern Arabic speaking world. This geographic distribution, together with those other, singly attested grammaticalization pathways described above, may be viewed in Figure 1. In this map, the incidences of specific grammaticalization paths are circumscribed with isogloss lines. These pathways are distinguished from one another by color of line, and individual etymological derivations comprising each path are differentiated by line styles (plain, dashed, dotted, etc.). The values of these different line colors and styles are indicated in the legend included with the figure.

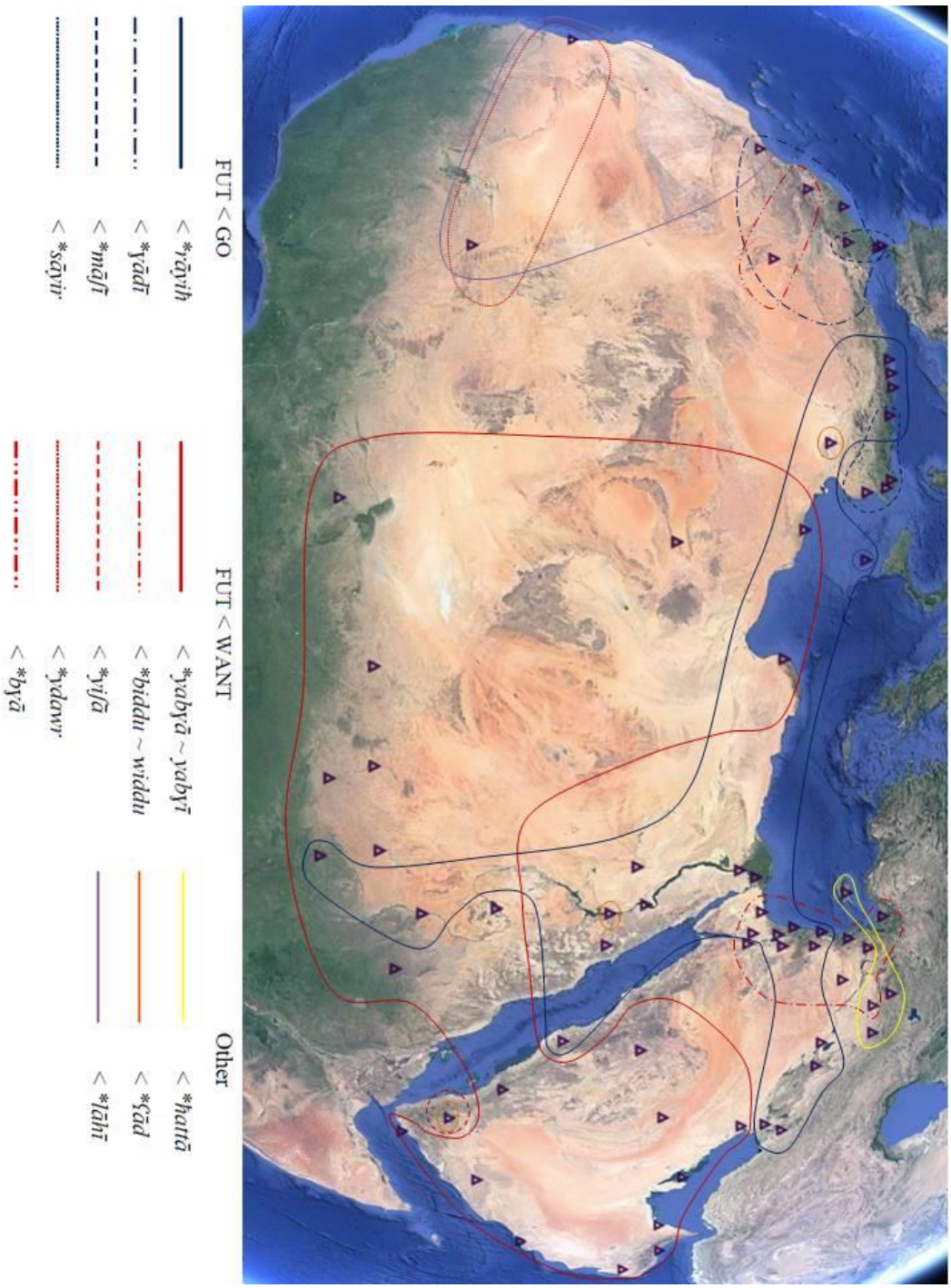

Figure 1: Geographic Distribution of Future Tense Markers (Map data: Google, S.O., NOAA, U. S. Navy, NGA, GEBCO; Image: Landsat/Copernicus)

As seen in Figure 1, the members of the multiply-attested grammaticalization pathways FUT < GO and FUT < WANT are distributed across the map in a largely contiguous manner. Of the source etyma representing the path FUT < GO, forms based on **rāyiħ* show the widest incidence. From an apparent center of heaviest concentration in Lower Egypt and the Levant, **rāyiħ* reflexes extend eastward through southern and central Iraq and southward along the Nile Valley as far as central Sudan and the western littoral of the Arabian Peninsula. To the west, **rāyiħ*-derived futures are attested in a band reaching across modern Libya and Tunisia to Algerian territory, where such forms are again found in abundance. Outside of the Middle East proper (here referring to North Africa and the Sudan), there has been a consistent trend among descriptive Arabic dialectologists (Souag 2005; Dickins 2006; Benkato 2014; among others) to identify these items as recent, relatively unintegrated morphological borrowings from the regionally prestigious Arabic variety of Cairo. While this may in fact be justified in the Sudanese context, wherein **rāyiħ*-based forms are broadly unattested in earlier documentation (cf. Hillelson 1935) and only begin to appear in force following generations of political union with neighboring Egypt, further afield in North Africa the assertion seems to be based solely on the (admittedly notable) similarity in phonological form of the reflexes across much of this region, as *ħa-* or *ħā*, and an unquantified impression that in some cases such forms are in more frequent use among young people. However, the presence of *ħa-* forms in such isolated locations as Malta *ħa-* and Tozeur *ħa-* and the occurrence of alternation with unreduced reflexes on the "bookends" of this zone (Dellys *rayəħ* ~ *ħa-*, ʻAwāmrah *laħ* ~ *ħa-*) would seem to argue against a twentieth century "matter" borrowing out of Egypt and in favor of a much deeper history in the area. It is likely that the adoption of the Egyptian hypothesis has been facilitated by notions of cultural centricity and perceptual dialectology; no similar proposals have been

advanced to explain Damascus *ħa-*, for example, which is generally regarded as an original development.

Bordering on the Algerian arm of the **rāyiħ*-based forms we find additional representative realizations of the FUT < GO path. Immediately to its east are encountered derivatives of **māʃī* in Tunisian territory, and to its west those **ɣādī* in the greater Moroccan area. Adjacent to these in turn are found the remaining blocs of FUT < GO realizations, consisting of the **sāyir*-based Maltese forms and an additional set of **māʃī* reflexes in northern Morocco. The four distinct realizations representing this multiply attested grammaticalization path thereby display complete geographic continuity.

As for the second path, FUT < WANT, the widest spread set of reflexes are those deriving from **yabɣā ~ yabɣī*. These are attested over nearly the entirety of the Arabian Peninsula, and additionally across the Red Sea throughout the greater Sudanic area. We also encounter such forms as we move northward from this region, encompassing modern Libya. Bordering the Arabian branch of these items to the north and west are the reflexes of a second instantiation of the FUT < WANT pathway, those of **biddu ~ widdu*. These forms are found throughout the broader Levant, sharing semantic space with **rāyiħ*-derived 'go' future forms throughout the southern portion of this area. Representing the FUT < WANT pathway alongside these two derivations are reflexes of **yiʃā*, which occur in immediate proximity to **yabɣā ~ yabɣī* forms in the southwest corner of the Arabian Peninsula. In interior North Africa, the next encountered future tense markers moving west from the Libyan examples of **yabɣā ~ yabɣī* include reflexes of **ydawr* in Mali and Mauritania and those of **bɣā* in southern Morocco and southwestern Algeria. Though the scarcity of sample points in this region means that significant "empty" territory exists between the westernmost reflexes of **yabɣā ~ yabɣī* on the one hand and the easternmost ones of **ydawr* and **bɣā* on the other, no intervening negative examples are found to

dissuade us from judging the two sets of forms to be direct neighbors. Moreover, the existence of related dialectological isoglosses spanning this same Saharan zone supports the proposed linguistic connection, and historical and anthropological studies of human movement within and across this area affirm the plausibility of contact between its inhabitants even across formidable terrain and distances (Lydon 2009; Burr & Collins 2008). The geographic distribution of the FUT < WANT developments includes two additional stretches across territory which may plausibly be conceived of as a natural barrier: the crossing of the Red Sea from the Arabian Peninsula to the Sudan and the bridging of the Syrian desert to link the Levant to the Arabian area. In both instances, additional linguistic isoglosses and nonlinguistic studies of social and cultural connectivity demonstrate that the proposition of linguistic continuity in these cases is eminently plausible. Thus, like the members of the FUT < GO grammaticalization pathway, the constituent realizations of FUT < WANT display a fully contiguous spatial contribution.

In light of these findings, both multiply-attested grammaticalization pathways which have passed conditions (i) and (ii) of the heuristic similarly meet condition (iii), the requirement for a broadly contiguous geographic distribution consistent with a history of diffusion via areal contact. As such, both these sets of forms have met the heuristic criteria selected to identify strong possible cases of products of CIG, the significance of which is discussed in the following section.

2.4 Discussion and Conclusions

In this chapter, the modern Arabic future tense markers have been evaluated for evidence of the occurrence of CIG. As an initial step, the forms of these markers were examined in detail and assigned to a smaller set of respective source etyma, of which twelve were identified. Once these etymological relationships were established, the results were subjected to the three conditions of the heuristic described in §1.4. First, the developments leading to the modern reflexes of the various source etyma were evaluated as defensible cases of grammaticalization via the identification of component processes of desemanticization, extension, decategorialization and erosion. In this case, all developments surveyed passed this examination and were recognized as examples of grammaticalization change. These results were then viewed with an eye as to which might be successfully grouped together under the heading of a particular grammaticalization pathway, as a unitary grammaticalization path attested by multiple, etymologically distinct chains of evolution is a logical precursor to the (conservative) identification of CIG. Two such paths were identified – FUT < GO comprising the reflexes of **rāyiħ, *ɣādī, *māʃī* and **sāyir* and FUT < WANT those of **yabɣā ~ yabɣī, *biddu ~ widdu, *yiʃā, *ydawr* and **bɣā* – the constituent members of which might each plausibly represent replications of one another's semantic-functional models.

Once these two multiply-attested paths were identified, the geographical incidence of the etymological derivations which compose them was considered in an attempt to ascertain whether they may have been propagated via processes of areal diffusion. The members of both the FUT < GO and FUT < WANT pathways showed complete geographic contiguity, and thereby successfully meet the third and final condition of the described heuristic.

The collected products of FUT < GO and FUT < WANT grammaticalization pathways may thus be identified as representing the results of repeated processes of CIG. Under this scenario, one "founding" member of each of these two pathways may be hypothesized to have originated independently, a development not surprising in light of the preponderance of these grammaticalization paths cross-linguistically. Subsequently, once the innovation had become established within a given area and the grammaticalization process had begun, speakers of a neighboring dialect or dialects encountered the change, comprehended its etymological source and trajectory, and replicated it in their own Arabic variety utilizing distinct but synonymous etymological resources. While the limited nature of the present survey does not allow for a nuanced analysis of chronology or directionality, presumably these processes repeated and the resulting innovations spread to the point that the products of the FUT < GO and FUT < WANT paths came to dominate the expression of future tense value across the observed body of modern Arabic dialects, with the results of other lines of development (reflexes of **ʕād*, **ħattā* and **lāhī*) relegated to numerically infrequent and geographically peripheral status: together, reflexes of these three unaligned etyma account for only eleven of the one hundred thirty-nine total examined and, with the exceptions of Tozeur *ʕa-* and Aswan *ʕa-*, occur only in the northeast (**ħattā*), southeast (**lāhī*), and southwest (Sana'a *ʕā*) extremes of the Arabic-speaking world. Given the correspondence of this narrative to the raw dialectological data and the theoretical criteria set out by this study's heuristic, processes of CIG may thus be seen to provide significant explanatory power in the historical analysis of the Arabic future tense markers, a systematic account of which has not previously been offered.

Chapter 3: Temporal Adverbs 'Now'

3.1 Introduction

The material considered in this chapter consists of the set of temporal adverbs signifying 'now' attested in the present sample, to be evaluated for evidence of CIG via the same three-part heuristic applied to the future tense markers. Words for 'now' are not as commonly cited in discussion of the Arabic pluriform developments as are the previous two features, but they nonetheless comfortably fit Versteegh's definition of the phenomenon as "a general trend that has occurred in all Arabic dialects, and an individual translation of this trend in each area," as will be seen upon review of the data (Versteegh 2001: 108). Intriguingly, and with special relevance to the present discussion, the analysis of terms for 'now' has figured more prominently in the small body of initial explorations into grammaticalization in the modern Arabic dialects, presented as a test case for grammaticalized adverbs by Watson (2006). Their development has also been partially addressed in Procházka's (2000) etymological investigation of temporal adverb formation. Though the sampling in this prior work has been far from comprehensive and the accompanying argumentation at times lacking in detail, it represents a valuable starting point and serves to flag these items as worthy of more thorough examination.

This chapter takes the form of the previous case study. First, the raw data gathered from the sample will be presented and assigned to appropriate etymologies using the methods described in §1.4. Next, the status of these evolutionary trajectories will be considered as grammaticalization pathways, and the occurrence of any multiply-attested pathways comprising multiple realizations within the sample will be noted. Finally, the third condition of the heuristic will be applied in assessing the geographic distribution of these pathways as consistent or inconsistent with a history of areal spread.

The chapter concludes with a discussion of these combined results and their significance as evidence of CIG in the development of these temporal adverbs.

3.2 Attested Forms by Etymology

Below is a full recounting of lexical adverbs meaning 'now' collected in the Arabic sample. As previously, these forms are organized as reflexes of proposed source etyma; I discuss the original composition and semantics of each etymon in turn, in addition to providing a brief overview of geographic incidence. Following this, I sketch the relevant evolutions in form which lead from these etyma to their individual attested reflexes. Uniquely to this chapter, I also reserve space at the end of this section for the description of secondary developments observable in the data which are best discussed separately from attempts at primary etymologizing.

**ðī l-waqt*

B'ērāt *dilgē, dilgēti*	Aswan *dilwagti*, *dilwakītī*	Nigeria *duggut*
Cairo *dilwaʔti*	'Awāmrah *dilwagt*	
Sinai *dilwagtiy*	'Abābdah *dilwagti*	

Table 20: 'Now' from **ðī l-waqt*

Reflexes of **ðī l-waqt* are encountered across Egypt, and extend south to appear in portions of the West Sudanic area. Etymologically, they consist of a demonstrative element **ðī* 'this' and the lexeme **waqt* 'time' modified by the definite article **l-*, with a compositional meaning of 'this time'. Demonstratives deriving from a simplex **ðā*

(MSG) / *ðī (FSG) appear to represent a fairly deep stratum in Arabic though innovative in relation to earlier Ancient North Arabian (Dadanitic) *ðā* (MSG) / *ðāt* (FSG) and a proposed distinction *ðā (MSG) / *tī (FSG) based on the internal reconstruction of multiple multimorphemic Classical Arabic forms (cf. Magidow 2013). Demonstratives reconstructable to *ðā / *ðī are attested from early in the Classical Arabic period (al-Farāhīdī, *Kitāb al-ʿAyn*) and in the modern era are ubiquitous throughout the Nile Valley and occur variably throughout the Arabian Peninsula and (less frequently) North Africa (Magidow 2013; Vicente 2006). However, the specific use of (at least formally) feminine gendered, preposed *ðī to modify otherwise masculine nouns such as *waqt would seem to be restricted to the Egyptian area, well attested historically but marginal in the present day (Zack 2009). This fact would seem to be the determining factor in limiting the level of the reconstructed periphrastic source, as the substantive *waqt 'time' is represented in all dialects in the sample for which sufficient lexical resources exist to consult and is further attested in Classical Arabic (al-Farāhīdī, *Kitāb al-ʿAyn*). The definite article *al- is considered to be a diagnostic innovation of the Arabic node as a whole, characterizing Classical Arabic, all modern dialects save a select few in northern Yemen, and even some forms of Ancient North Arabian (Versteegh 2001; Huehnergard 2017); as such, it does not contribute to defining a precise level of linguistic reconstruction and in the interest of concision will not be addressed explicitly in any following discussions of this point.

All varieties showing reflexes of *ðī l-waqt in the sample have undergone a regular merger of the interdental series with corresponding alveolar stops, and as such the inherited /ð/ is consistently reflected by /d/; the sole exception in this regard is Sinai *dilwagtiy*, in which */ð/ > /d/ is not regular. The following /ī/ is shortened to /i/ according to general phonological rules governing elision with definite article, and in the

West Sudanic area is realized as /u/ (Nigeria *duggut*). Though the precise conditioning of the latter change remains unclear (perhaps related to the proximity of /w/), widespread interchange of historical */i/ and */u/ is common in the region while */a/ remains distinct (Owens 1993), thus favoring a reconstruction of **ðī l-waqt* over alternative **ðā l-waqt*. Historical */q/ is realized as /ʔ, g/ in accordance with regular sound change: Cairo *dilwaʔti*, Aswan *dilwagti*. The final vowel /i/ of most Egyptian forms is likely the frozen product of a general anaptyctic process CC#C > CCi#C (indeed, in an example from his grammar of Cairene, Woidich (2006) presents a variant form *dilwaʔt*, given in pausal position). In at least one form from the sample, NW Sinai *dilwagtiy*, the result is not consistent with native phonological processes and thus likely represents a wholesale borrowing from a neighboring Egyptian dialect, complete with reinterpretation of the phonological status of the final /i/ (a hypothesis supported by the irregular treatment of */ð/). In the Nigeria *duggut*, cited above, the second /u/ similarly resolves the final cluster though according to local rules: CC# > CiC#, following by a harmonic backing under the influence of the /u/ in the first syllable. The insertion of /ē, ī/ in B‘ērāt *dilgē*, *dilgēti* and Aswan *dilwakītī* is treated §3.2.1, below.

**ðā l-waqt*

Algiers *dork*, *ðork*	M'zab *dˁarwek ~ dˁrūk ~ ðrūk ~ dˁarka*	Marrakech *drūk*, *drūka*
Nouakchott *ðˁark*	Saïda *darwək ~ ðˁorwok*	Mali *ðark*, *ðrayk*
Cherchell *ðərwaq*	Djidjelli *delwoq ~ derwoq*	Dellys *ðurwək ~ ðurk ~ ðˁukk ~ ðˁukka*
Tlemcen *derwaq*	Goulimine *druk*	
Larbaâ *darwak ~ dark ~ darka*	'Abābdah *dalwagti*	

Table 21: 'Now' from **ðā l-waqt*

Forms reconstructable to **ðā l-waqt* occur across the southern Sahara, extending westward as far as Mauritania and north to the Algerian coast, in addition to one possible attestation further east among Upper Egypt's 'Abābdah tribe. They consist of a demonstrative **ðā* 'this' combined with the definite article **l-* and the lexeme **waqt* 'time' to create a construction with a meaning of 'this time'. The relatively wide-spread and deep chronological attestation of the demonstrative form **ðā* (MSG) has just been discussed. The form is typical of the Saharan area considered here and, though not current in coastal North Africa, isolated reflexes in locations such as Malta and Tetouan combine with historically documented forms in the region to point to a once more common incidence (Magidow 2013). As previously described, **waqt* would appear to be reconstructable to the common ancestor of Classical Arabic and the great majority of modern dialects. Thus, the source construction **ðā l-waqt* could theoretically have arisen

quite early, but would need to have been established prior to the apparently recent and still ongoing loss of productive **ðā* demonstrative forms in the North African littoral.

In terms of phonological evolution, ‘Abābdah *dalwagti* parallels the *dilwagti* variant discussed for the same dialect previously, save that the demonstrative element reflects masculine **ðā* rather than feminine **ðī* ; the masculine agreement is more typical of demonstrative constructions used outside Egyptian area, and this variation may thus be characteristic of the overlap between Egyptian and Sudanese isoglosses noted elsewhere in studies of ‘Abābdah dialectology (see de Jong 2002). Further west, development takes a different path but displays a consistent trajectory over a wide area. Inherited */ð/ is afforded either an alveolar or an interdental realization as consistent with each dialect’s regular sound change laws governing interdental fricatives: Cherchell *ðərwaq*, Goulimine *druk*. In some cases, this initial consonant is also pharyngealized, most probably under influence from a neighboring “emphatic” /r/ (Nouakchott *ðˁark*, M’zab *dˁrūk*) or as a shift in the locus of pharyngealization accompanying the loss of said /r/ (Dellys *ðurk* ~ *ðˁukk*). The /r/ in question is itself an irregular product of inherited */l/ which dominates in most forms traceable to this etymology, though at least one location displays variable forms which make the correspondence clear: Djidjelli *delwoq* ~ *derwoq*). The original */w/ is varyingly present or deleted, and at times apparently vocalized to /ū, u/ as in Marrakech *drūk*. Original */q/ is realized as /q, k/: Tlemcen *derwaq* vs. Larbaâ *dark*. Forms displaying /k/ are found in precisely those Algerian and Saharan regions where one expects a change of */q/ > /g/ (including Algiers *dork*, where variation exists between /q/ ~ /g/ in the dialect more generally). For many of the better documented dialects in this area, an idiosyncratic realization of */q/ as /k/ is noted specifically for the lexeme *wakt* ‘time’; this change is most likely due to the influence of the adjacent /t/, whether in the form of fronting directly from */q/ > /k/, as proposed by Grand’Henry (1976), or as an

example of devoicing assimilation following a regular sound change (i.e., */q/ > */g/ > /k/). In either case, the change must have preceded the loss of the final /t/ of these forms, apparently categorical across the greater Algerian/Saharan area. The occasional final /a/ which variably appears in interior Algeria (e.g. Larbaâ *darka*) is of uncertain origin.

The identification of a source for these forms in demonstrative **ðā*, rather than the **ðī* of the Nile Valley and West Sudanic area, is confirmed by the quality of the vowel in the first syllable. Several forms considered here must be viewed as ambiguous with regard to this determination: some show no vowel in this position (e.g., M'zab *dˁrūk*) while those of the Algerian littoral typically contain /ə/ or a phonetic variant thereof (e.g., Cherchell *ðərwaq*), which due to a historical merger of /a/ and /i/ across that region could equally point to either etymon. However, in the Algerian interior and Mauritanian Sahara, several forms occur which distinctly show /a/ in the first syllable: Saïda *darwək*, Mali *ðark*. These dialects have not undergone the aforementioned merger of /a, i/ (Cohen 1970) and are thus best interpreted as faithfully transmitting the vocalism of an original **ðā*. Once this fact is made clear, both geography and the shared suite of irregular phonological developments described previously (*/l/ > /r/, */t/ > Ø, and occasional */q/ > /k/) argue for the grouping of the less conclusive zero- and reduced forms with this latter set as sharing an origin in etymological **ðā*.

**hā l-waqt*

Antiochia *ħallaq*	Damascus *hallaʔ*	ʻAwāmrah *halwagit*
Beirut *hallaʔ*	Jerusalem *halʔēt*	Jebel Ansariye *hallaq*
Cilicia *hallaq*	Amman *halla*	
Cyprus *ʔalok*	Aleppo *hallaq ~ hallaqtēn ~ hallaqne*	

Table 22: ʻNowʼ from **hā l-waqt*

Reflexes of **hā l-waqt* occur across the greater Levant, extending from Cyprus and the Cilician plain in the north to parts of the Nile Delta in the south, and inward from the coast as far as Jordan and central Syria. They are composed of a demonstrative element **hā* modifying a definite **l-waqt* ʻ(the) timeʼ, together delivering an original meaning of ʻthis timeʼ. As previously discussed, **waqt* is reconstructable to at least a common ancestor of Classical Arabic and the modern dialects. Demonstrative **hā* is perhaps more restricted, in use today as a productive demonstrative element in eastern North Africa, the Levant, Mesopotamia, and parts of the Arabian Peninsula (Vicente 2006; Brustad 2000). It has been proposed that this function is an extension of an original role as a presentative particle, which the item continues to play in the remaining modern dialects and in Classical Arabic (Pat-El 2009; Magidow 2013). Uses of **hā* in an apparently demonstrative capacity are attested in the Safaitic inscriptions, though the direct comparability of these to the modern examples is complicated by the lack of an intervening definite article between demonstrative and noun (Al-Jallad 2015). Whatever the ultimate time depth, the regional specificity of the modern demonstrative **hā* is

consistent with the distribution of the **hā l-waqt* reflexes considered here, all of which are encountered in the greater Levantine area.

The initial consonant of these forms generally remains /h/, though in Cyprus *ʔalok* becomes /ʔ/ via regular sound laws and in Antiochia *ħallaq* undergoes a sporadic change to /ħ/. The /a/ of the first syllable undergoes shortening due to assimilation to the definite article. The /w/ of **waqt* is retained in ʻAwāmrah *halwagit*, but is generally deleted with compensatory gemination of the preceding /l/: Beirut *halla?*. Inherited */q/ is reflected by /q, ʔ, g, k/ according to regular local reflexes: Jebel Ansariye *hallaq*, Damascus *halla?*, ʻAwāmra *halwagit*, Cyprus *ʔalok*; in at least one case, Amman *halla*, it is deleted entirely. Of the forms represented in the sample, final /t/ is lost in all cases save ʻAwāmra *halwagit*, though it occasionally resurfaces in in the presence of added augmentative morphology: Aleppo *hallaq ~ hallaqtēn*. Such secondary morphological developments, also present in Jerusalem *halʔēt*, are discussed in §4.2.1, below.

**ðā l-ħīn*

Mecca *daħħīn*	Rubāṭāb *daħīn*	Abha *ðalħīn*
Sana'a *ðalħīn*	Ḥarb *ðaħīn*	Aden *daħīn*
Hadhramaut *ðalħīn ~ ðaħħīn*	Shukrīyah *daħīn*	Kordofan *dahīn*
Kabābīsh *taħīn*	J-Tunis *dəlħīn*	Dhofar *ðalħīn ~ ðaħħīn*

Table 23: ʻNow' from **ðā l-ħīn*

Forms reflecting an etymology **ðā l-ħīn* occur in the western portion of the Arabian Peninsula as well as the adjacent eastern and central portions of the Sudan; in

addition, a sole attestation occurs much further east: J-Tunis *dəlħīn*. These forms consist of a demonstrative element **ðā* combined with the lexeme **ħīn* 'time', modified by the definite article **l-*, rendering a compositional meaning of 'this time'. The deep reconstructability of the demonstrative **ðā* has already been described; its prenominal position in the **ðā l-ħīn* construction is not typical of word order in the modern Sudanese dialects cited in the table above, but the variably prenominal or postnominal position of demonstrative **ðā* in closely related dialects of the Arabian Peninsula provides ample grounds for positing its plausibility at an earlier stage. The substantive **ħīn* 'time' is attested from the earliest phases of Classical Arabic (al-Farāhīdī, *Kitāb al-ʿAyn*) and is current through much of the Arabian Peninsula today (cf. Qafisheh 1997). Outside of this area, it would seem mainly to exist in the context of fixed expressions (e.g., Hinds & Badawi 1986; Harrell 1966). It is not attested as a stand-alone lexeme in the limited lexical documentation available for the contemporary Sudan.

Inherited */ð/ appears according to local reflex of the sound more generally: Abha *ðalħīn*, Šukrīyah *daħīn*; in one case, Kabābīsh *taħīn*, there appears to be a sporadic devoicing to /t/ where /d/ would otherwise be expected. Original */ā/ is consistently shortened (and in Tunis further reduced to /ə/ through regular sound change) due to assimilation with the definite article. The /l/ of the latter is variously retained, irregularly deleted, or deleted with compensatory gemination of the following /ħ/: Sana'a *ðalħīn*, Rubāṭāb *daħīn*, Mecca *daħħīn*. In some dialects, multiple phonological outcomes co-exist simultaneously on this count (e.g., Wadi Hadramawt *ðalħīn* ~ *ðaħħīn*).

**hā l-ħīn*

Bahrain *alħīn*	Ḥarb *halħīn*	Āl Wahībah *halħīn ~ alħīn*
Kuwait *alħīn*	Banī Ṣakhr *halħīn*	Sudayr *halħīn*
Negev *halħīn ~ halħīniy*	Bdūl *halħīn*	Ẓafīr *hālħīn ~ hāħīn*
Sinai *ħalħīn ~ alħīn ~ halħīniy ~ alħīniy ~ halħīnit*	Abu Dhabi *halħīn ~ alħīn*	

Table 24: 'Now' from **hā l-ħīn*

Forms reconstructable to **hā l-ħīn* occur across Arabia, from the Sinai Peninsula in the west to Oman and the shores of the Persian gulf in the east. Their constituent elements are a demonstrative **hā*, the definite article **l-* and the lexeme **ħīn* 'time', which combine to form the meaning 'this time'. As previously discussed, the demonstrative element **hā* is reconstructable to a large, contiguous bloc of dialects encompassing the Levant, Mesopotamia, and northern and eastern portions of the Arabian Peninsula and **ħīn* at least to a common ancestor of the modern Peninsular dialects and Classical Arabic (an overlap of features consistent with the observed incidence of modern **hā l-ħīn* reflexes).

In all forms save Ẓafīr *hālħīn ~ hāħīn*, the inherited /ā/ of **hā* is shortened as a regular product of assimilation with the definite article, and the initial /h/ is most often preserved as such. One exception, Sinai *ħalħīn*, seems to represent an example of anticipatory assimilation at a distance to the subsequent /ħ/ of **ħīn*. In other cases, it is possible to identify the deletion of earlier /h/, as in Kuwait *alħīn*. Care must be taken to distinguish such forms from reflexes of simple **al-ħīn* 'the time', a separate etymon to be discussed below. In dialects such as Kuwaiti, it is possible to make a distinction on the

basis of the realization of the definite article, which is not *al-* in these varieties but *(i)l-*. The anomalous /a/ of forms such as *alḥīn* is thus better analyzed as reflecting earlier **hā*. That the distinction is a valid one is supported by the fact that multiple forms traceable to both **hā l-ḥīn* and **al-ḥīn* may be found side-by-side in a single dialect: Abu Dhabi *halḥīn*, *alḥīn* < **hā l-ḥīn*, *lḥīn* < **al-ḥīn*. In the case of Ẓafīr *hālḥīn* ~ *hāḥīn*, /l/ is variably deleted through an irregular change also noted to affect other high frequency lexical items in the dialect, for example *albārḥa* ~ *abārḥa* 'last night' (Ingham 1982: 252).

Forms identified in the area of the Sinai Peninsula and Negev Desert seem to indicate the presence of a close variation of this etymology, existing alongside more straightforward reflexes of **hā l-ḥīn*. The additional element included at the end of Sinai forms *halḥīniy*, *halḥīnit* and Negev *halḥiniy* would appear to represent regular reflexes of the feminine singular nominal marker **-a/-at*, pointing to an etymological **ḥīna* rather than **ḥīn*. While such a form is not otherwise directly attested, an explanation may lie in the common use of the feminine singular morpheme as a marker of individuation and specificity cross-dialectally (Brustad 2000), here expressing a notion of a particular time or instance. Another interpretation would be to view this as the product of analogy with the reinterpreted final epenthetic */i/ of general Egyptian *dilwaʔti* vs. Sinai *dilwagtiy*, discussed above.

**hā s-sāʕa*

Baghdad *hassa*	Shukrīyah *hassaʕ ~ hassiʕ*	Khuzestan *hassa*
Khartoum *hassaʕ ~ hassi*	Abéché *hassa*	Nigeria *hassa*
Rubāṭāb *hassaʕ*	Hit *hassaʕ*	Mosul *hassaʕ ~ hassaʕta*
Khawaytnah *hassaʕ ~ hassaʕēn*	Basra *hassa*	Jisr az-Zarqa *hassa*
Banī Ṣakhr *hassāʕ ~ hassaʕ*	Kordofan *hassaʔ*	

Table 25: 'Now' from **hā s-sāʕa*

Reflexes of **hā s-sāʕa* occur in two major blocs, the first comprising Mesopotamia and adjacent areas of the Levant and the second covering the greater Sudanic area. They are composed of a demonstrative element **hā* modifying the lexeme **sāʕa* 'time, while, hour'. The clitic **s-* is an allomorph of the definite article **al-*, as conditioned by the initial consonant of **sāʕa*. The compositional meaning of the construction is 'this time'. The substantive **sāʕa* in this meaning would seem to be common to all dialects in the sample for which reliable lexical resources exist, and additionally occurs in Classical Arabic (al-Farāhīdī, *Kitāb al-ʿAyn*). The reconstructability of presentative-turned-demonstrative **hā-* has already been described, though here it is of note that such a form is not current in the contemporary Sudan, where several reported reflexes of **hā s-sāʕa* occur. Some evidence from internal reconstruction, however, indicates the historical use of Sudanese **hā* in demonstrative contexts (Magidow 2013), and given the significant distance between the Sudanese reflexes and cognate forms to the northeast it seems preferable to view them as a self-standing development based on a now obsolete pattern of demonstrative marking (as seen

for pan-Egyptian **ðī l-waqt*, in that case witnessed by a stratum of historical documentation not available for the Sudanese area).

The /ā/ of **hā* is consistently shortened due to phonological rules governing assimilation with the definite article. The /ā/ of **sāʕa* is also most often sporadically shortened to /a/ (e.g., Hit *hassaʕ*), though not in all cases (Banī Ṣakhr *hassāʕ*). Twice in the Sudanic area this vowel undergoes an irregular raising to /i/: Shukrīyah *hassiʕ*, Khartoum *hassi*. The final /a/ of this lexeme is deleted across the sample. The /ʕ/ is at times retained, at times deleted, and in one case becomes /ʔ/ via regular sound change: Khartoum *hassaʕ*, Jisr az-Zarqa *hassa*, Kordofan *hassaʔ*. One exception to the loss of final /a/ might be seen in Mosul *hassaʕta* (attested alongside expected *hassaʕ*), which would appear to reflect an original **hā s-sāʕata* with adverbial accusative inflection in the style of Classical Arabic, though without broader knowledge of adverb formation in the dialect it is impossible to make a definitive determination.

**hā l-ħazza*

Bahrain *halħazza*	Kuwait *halħazza*	Abu Dhabi *halħazza* ~ *alħazza*

Table 26: ‘Now’ from **hā l-ħazza*

Forms reflecting an etymology of **hā l-ħazza* are found in the area of the Persian Gulf. The principal components are demonstrative **hā*, definite article **l-* and the lexeme **ħazza* ‘time, moment’, together constituting a construction meaning ‘this time’. The reconstructability of demonstrative **hā* to a wide swath of northern and eastern Arabic varieties has already been discussed. The noun **ħazza* in the meaning of ‘time,

moment' (an apparent extension from a more widespread, original 'notch, nick') seems to be well documented in the modern day only the Arabian Peninsula, particularly the Persian Gulf region which comprises all three locations included above (Qafisheh 1997). The temporal meaning of the term is not attested in the early Classical Arabic lexicon represented by *Kitāb al-ʿAyn*, but is known at least from Classical Arabic dictionaries dating to the medieval era (Lane 1968).

In terms of phonological development, the /ā/ of **hā* is in all cases shortened due to its combination with the definite article. The initial /h/ is at times maintained, at times irregularly deleted: Bahrain *halħazza*, Abu Dhabi *alħazz*. Such cases of deletion are distinguishable from reflexes of simple **al-ħazza* 'the time' by means of the quality of the vowel in the first syllable, as discussed for forms descending from **hā l-ħīn* above. As for **hā l-ħīn*, products of both etymologies may be present within a single variety: Abu Dhabi *halħazza*, *alħazza* < **hā l-ħazza*, *lħazza* < **al-ħazza*.

**ðī l-ʔawān*

Ḥarb *ðulwān ~ ðuwān*	Banī Ṣakhr *ðilwān*	

Table 27: 'Now' from **ðī l-ʔawān*

Two forms that can be traced to the source construction **ðī l-ʔawān* are found in an area stretching from central Jordan south to the northwest portion of the Arabian peninsula. They are most likely composed of a feminine demonstrative **ðī* and the lexeme **ʔawān* 'time' (modified by definite article **l-*) though viable alternatives to both these etyma exist, namely masculine demonstrative **ðā* and the related lexeme **ʔān* 'time'. Each will be discussed below, and the preferred analysis of **ðī l-ʔawān* defended.

Regardless of the finer points of this distinction, the compositional meaning of the resulting constructions is the same: ‘this time’.

The identification of a source in feminine demonstrative **ðī* is prompted by the quality of the vowel of first syllable, in all cases either /i/ or /u/: Banī Ṣakhr *ðilwān*, Ḥarb *ðulwān*. In this view, forms with /i/ would thus directly reflect etymological */ī/, shortened due to its position preceding the definite article, and those with /u/ would represent a logical secondary development proceeding from the general loss of contrast between inherited */i, u/ in the dialects in question, resulting in a lax high or mid vowel phonetically “colored” by surrounding consonantism (Palva 1980). Though thus the sounder choice from a phonological standpoint, the selection of feminine demonstrative **ðī* over masculine **ðā* is less satisfying syntactically: unlike the Egyptian reflexes of **ðī l-waqt* discussed above, there is no precedent in these Arabic varieties for the somewhat anomalous usage of a feminine demonstrative form preposed to a masculine noun, and thus **ðā* would be the “expected” choice for such a construction. This is not necessarily damning, however, as were it not for the Egyptian forms’ unusually deep record of historical attestation their unanimously accepted **ðī l-waqt* etymology would seem equally far-fetched on the basis of agreement patterns in present-day Egyptian varieties. In light of this state of affairs, it is perhaps preferable to abide by the far better understood principles of phonological change in the area than to posit a sporadic raising and rounding of original */ā/ solely in order to account for an assumed pattern of syntactic agreement. It is not clear whether such a structure should be posited to represent a common innovation with that of historical Egyptian varieties (its only clear parallel) or an independent development in demonstrative agreement patterning; potential precedent might also be found in the use of formally unmotivated feminine distal demonstratives with certain time nouns in Syrian, Moroccan, Kuwaiti, and some

Egyptian varieties, for example Damascus *hadīk* (FSG) *l-yōm* (MSG) 'that day' (Brustad 2000: 127-128).

Greater ambiguity exists in the choice between the related lexemes **ʔawān* and **ʔān* as an element originally meaning 'time' in these forms, currently reflected in both varieties by *-wān*; both are attested in effectively interchangeable meanings since the Classical Arabic period (al-Farāhīdī, *Kitāb al-'Ayn*), though appear only to be active in the modern dialects in the context of fixed expressions (see examples in Hinds & Badawi 1986; Qafisheh 1997). In selecting between the two in the present case, the crux lies in the historical treatment of initial /ʔ/. Two outcomes, neither clearly conditioned, are attested side-by-side in each of the relevant varieties. The first involves deletion of the glottal stop and any following vowel, the second mutation of /ʔ/ to /w/ (Il-Hazmy 1975): Ḥarb *bil* < **ʔibil* 'camels', *wann* < **ʔanna* 'he sighed'. Thus, *-wān* may equally represent the product of **ʔawān* via the first sound change or **ʔān* via the second. As the latter change of /ʔ/ > /w/ would appear relatively rare in the dialect of the Banī Ṣakhr in particular (restricted in the available data to interrogatives *wēn* < **ʔayn* 'where?' and *wiʃ* < **ʔayʃ* 'what?'), the reconstruction of *-wān* < **ʔawān* via deletion, well attested across both relevant varieties, has been tentatively favored.

In terms of further phonological development, the */l/ of the definite article is variably deleted in Ḥarb *ðulwān ~ ðuwān*.

**al-ħīn*

Bahrain *ilħīn*	Bdūl *alħīn*	Dhofar *ilħīn*
Negev *alħīn ~ alħīniy*	Abu Dhabi *lħīn*	Sudayr *alħīn*
Sinai *ilħīn ~ ilħīnih*	Abha *alħīn*	Ẓafīr *alħīn ~ aħīn*

Table 28: 'Now' from **al-ħīn*

Forms reflecting an etymology of **al-ħīn* are found across the southern portion of the Arabian Peninsula, northward through the Persian Gulf region and central Arabia, then west as far as the Negev Desert and the Sinai Peninsula. They consist of the lexeme **ħīn* modified by the definite article **al-*, with a presumed original meaning of 'the time'. As previously described, the lexeme **ħīn* 'time' is current in the contemporary dialects of the Arabian Peninsula, and included already in the earliest lexica of Classical Arabic.
The phonological variation of note across these forms primarily reflects local iterations of the definite article in /al-, il-, l-/: Abha *alħīn*, Dhofar *ilħīn*, Abu Dhabi *lħīn*. In one case, that of Ẓafīr *alħīn ~ aħīn*, /l/ undergoes variable irregular deletion. As discussed for **hā l-ħīn*, above, there is an indication in the Sinai/Negev region of a variant etymology based on an individuated **ħīna*: Sinai *ilħīnih*, Negev *alħīniy*.

**al-ħazza*

Antiochia *alħaz*	Soukhne *alħazz ~ alħaz ~ alħa ~ alħaztēn*	Abu Dhabi *lħazza*

Table 29: 'Now' from **al-ħazza*

Forms reconstructable to **al-ħazza* occur in the environs of Abu Dhabi on the Persian Gulf and significantly further north in the Syrian interior and among the Bedouin inhabitants of Antiochia. They are composed of the lexeme **ħazza* 'time, moment' modified with the definite article **al-*, with an original meaning of 'the time'. The noun **ħazza* in the meaning of 'time' has already been described as traceable to medieval Classical Arabic and current in dialects of the Arabian Peninsula. While lack of sufficient lexical resources makes it difficult to confirm or deny its presence in the peripheral Syrian varieties included above, the noted dialectal continuity between the Central Arabian area and the Syrian Desert (Ingham 1994) renders its occurrence in those regions highly plausible.

The initial vowel (or lack thereof) is dictated by the local form of the definite article: Abu Dhabi *lħazza*, Antiochia *alħaz*. In the latter, final /a/ has been deleted and /zz/ degeminated, both sporadically. In Soukhne, significant irregular erosion is similarly attested word-finally, resulting in the variable forms *alħazz* ~ *alħaz* ~ *alħa* (the nature of Soukhne *alħaztēn* is considered in §3.2.1, below).

**as-sāʕa*

Malta *issa*		

Table 30: 'Now' from **as-sāʕa*

In Malta, the form *issa* seems to reflect an etymology of **as-sāʕa*, consisting of the lexeme **sāʕa* 'time, while, hour' and **s-*, an allomorph of the definite article **al-*. The compositional meaning of the construction is 'the time'. As previously described,

the lexical noun **sāʕa* is attested across the modern Arabic varieties (including Maltese) and in Classical Arabic.

As is typical for Maltese, the vowel of the definite article is /i/. The final /a/ and the pharyngeal approximant are not present. Though deletion of /ʕ/ is regular in Maltese, it is probable that its loss and that of the subsequent /a/ in *issa* precedes this general development: deletion of /ʕ/ is understood to follow the regular raising of word-internal /ā/ (cf. *tīyək* < **tāʕak*), indicating that the etymological /ā/ of *issa* was already in word-final position before the spread of the latter change, and is represented by contemporary short /a/ due to a general loss of vowel quantity distinctions word finally.

**ðāba*

Casablanca *dāba*	J-Fez *daba*	Anjra *dāba*
Tlemcen *dāba*	Tetouan *daba*	
Goulimine *daba*	Marrakech *dāba*	

Table 31: 'Now' from **ðāba*

Across Morocco and far western Algeria we encounter forms reconstructable to **ðāba*. Though unattested in Classical Arabic or other well documented historical forms of the language, the presumed original meaning of the term would be something like 'immediately, right away', on the basis of the modern Sudanese construction *dābin mā* 'as soon as' (Hillelson 1935) and etymological relatives in Egyptian and Syrian *(ya)dōb* 'only just, barely' (Hinds & Badawi 1986; Stowasser & Ani 2004); compare also the historical Palestinian usage *edōb*, presented by Bauer (1926) as a synonym for *etaww* 'just now' (cf. discussion of **tawwa*, below). An additional descendant of the same root

within the Moroccan area may incidentally be found in Anjra *dābsāx* ‘afterwards’. Beyond noting the uniqueness of this term, Vicente (2000) does not attempt an etymology, but based on other attested temporal terms she describes for the dialect a viable proposal might be **dāb s-sā-k*, with *dāb* meaning something like ‘immediately following’, *s-sā* being an attested Anjra operator meaning ‘then, (at) that time’, and **-k* a distal deictic element present in other temporal and spatial adverbs in the variety (becoming *-x* word finally by regular process). The proto-etymon **ðāba* itself may in turn be traceable to the Classical Arabic root *ð-w-b* ‘melt (away), dissolve, disappear’, which in its metaphorical use emphasizes an inchoative change of state:

1) Classical Arabic (Lane 1968: 986)
   mā ðāb -a fī yad -ay -hi min -hu xayr -un
   NEG melt.PFV-3MSG in hand-DU-GEN.3MSG from-GEN.3MSG good-NOM
   ‘No good came to him of it.’ (lit. ‘Good did not melt into his hands …’)

2) *ðāb -a ʕalay-hi l- māl -u*
   melt.PFV-3MSG upon-GEN.3MSG DEF-property-NOM
   ‘He came to owe the property.’ (lit. ‘The property melted upon him’)

Precedent for a semantic link between change of state and immediacy in Arabic exists in the in the derivatives of the Classical Arabic root *ħ-w-l*, which include both the verb *ħāla* ‘change, shift, transform’ and the adverbials *ħālan*, *fī l-ħāl* ‘immediately’ (Lane 1968). Lacking further evidence, however, the specific relationship of reconstructed root **ðāba* and Classical *ð-w-b* are for the moment difficult to establish with any certainty.

The initial consonant, though universally represented by /d/ in the forms included in the present sample (e.g., J-Fez *daba*) , is best reconstructed to */ð/ in light of reflexes

such as *ðaba* noted for contemporary Chaouia and historically recorded Andalusi Arabic, as reported by Heath (2002). The subsequent merger to /d/ would be consistent with regular sound change laws in all the dialects cited above. Variation in the length of internal /ā/ reflects regular dialectal norms.

The etymology presented here is deemed preferable to the traditionally proposed **ʔiðā bi-hi* 'and then' due to a) the tenuous phonological relation of the latter to attested modern forms, such a derivation forcing the researcher to posit the occurrence of multiple idiosyncratic sound changes, and b) its failure to account for the aforementioned and apparently related temporal terms which exist outside the Moroccan/Algerian area.

**tawwa*

Tunis *təwwa*	Fezzan *taww ~ taw*	Benghazi *towwa*
Tripoli *tawwa*	Misīrīyah *tauwa*	Al-Khaburah *taww*
Tozeur *towwa ~ toww ~ tow ~ tū*	Sousse *tawwa*	Mateur *tawwa ~ taw*
B-Kadugli *towwa*	J-Tunis *tawwa*	

Table 32: 'Now' from **tawwa*

Forms descending from **tawwa* occur across Libya and Tunisia, extending southward through the Fezzan and as far as western Sudan. One outlier form is found in Oman (Al-Khaburah *taww*). They originate in an adverb **tawwa*, 'immediately, right away', which is widely attested across multiple modern dialect regions from Egypt to the Persian Gulf to Mesopotamia (Hinds & Badawi 1986; Qafisheh 1997; Woodhead & Beene 1967). The meaning of the adverb in Classical Arabic seems to have been 'one-

fold, singly' (al-Farāhīdī, *Kitāb al-ʿAyn*), but by the early modern period one finds the value 'immediately, just now' entering late stage classical lexica, such as the eighteenth century *Tāj al-ʿArūs* (*apud* Lane 1968: 321).

The vowel of the first syllable varies between /a, ə, o/: Sousse *tawwa*, Tripoli *təwwa*, Benghazi *towwa*. Forms with /ə/ represent a regular reduction of inherited /a/, while those with /o, ū/ likely display an anticipatory rounding under the influence of following */ww/. It is unclear whether Misīrīyah *tauwa* contains a phonemic diphthong or whether is merely an alternative transcription for *tawwa*. Original */a/ is often lost and */ww/ degeminated as part of a broader process of irregular, regressive erosion resulting in the deletion of word-final phones. The full spectrum of outcomes is demonstrated by Tozeur *towwa* ~ *toww* ~ *tow* ~ *tū* (with the final variant showing the furthest degree of lenition, in the form of monophthongization).

3.2.1 Secondary Developments

Several forms present in the sample display evidence of secondary development via the addition of supplemental morphology to a preexisting word for 'now'. As such augmentations may be demonstrated for phonological or morphological reasons to have occurred following the initial development of the terms in question, they will be considered separately from the primary etymological sources described above. Given the focus of the present analysis on the evaluation of possible areal phenomena, only those developments attested across multiple sample points will be discussed.

NOW + *-ēn*

In a region encompassing interior Syria and adjacent portions of Iraq, the sample records three forms consisting of a word for 'now' augmented with a suffix *-ēn*: Aleppo *hallaqtēn*, Soukhne *alḥaztēn*, Khawētnah *hassaʕēn*. These are based on reflexes of the three distinct primary etymologies **hā l-waqt*, **hā l-ḥazza* and **hā s-sāʕa*, discussed above. The origin of the added suffix itself is far from clear; multiple proposals have been tendered, ranging from a historical dual marker to a phonological deformation of an inherited adverbial ending to a borrowed deictic from local varieties of Aramaic. Connections have been drawn to *-ēn* endings augmenting other temporal adverbials in the same dialect area and beyond (Procházka 2000).

Regardless of its ultimate source, of importance to the present analysis is the fact that the addition of *-ēn* seems to have occurred after the initial development of the relevant etyma from their original periphrastic constructions, the clearest case being Khawētnah *hassaʕēn* in which the *-ēn* suffix attaches directly to the final /ʕ/ of the modern stem, rather than the subsequently deleted /a/ of original **hā s-sāʕa* (which would be rendered morphologically as /-at/ in non-final, modified state). Also in support of their secondary nature is the fact that in all cases the forms with *-ēn* exist alongside unaugmented variants representing the same base etyma: Aleppo *hallaq*, Soukhne *alḥaz*, Khawētnah *hassaʕ*. As such, the presence of *-ēn* will not be taken to exclude a given form from classification with unmodified outcomes of the same etymological origin but rather as an additional layer of areality to be analyzed separately.

NOW + *-ay-*

A second morphological augmentation attested by multiple forms in the sample lies in the addition of the diminutive infix **-ay-*, which is applied between the ultimate and penultimate members of a nominal lexeme's consonantal root in the pattern C(V)C*ay*C. Such usage is paralleled cross-linguistically (cf. Spanish *ahora*, *ahorita*), often with a meaning of 'right now, just now'. With the possible exception of Mali *ðrayk*, however (which is alternately glossed in Heath's (2003) texts as 'now' and 'right now'), no such semantic or pragmatic nuance is recorded for the diminutized forms of the present sample, which seem to simply carry a general meaning of 'now' at least synchronically.

The diminutized forms occur in three geographic pockets: a sole attestation in the Malian desert, a pair located in Upper Egypt, and in Palestine. The Malian form appears to descend from the general Saharan source **ðā l-waqt*, while those in Egypt reflect a primary etymology of **ðī l-waqt* and that in Palestine regionally dominant **hā l-waqt*: Mali *ðrayk*, B'ērāt *dilgēti ~ dilgē*, Aswan *dilwakīti*, Jerusalem *halʔēt*. In all but the Timbuktu case, **-ay-* has presumably monophthongized to *-ē-* via regular sound change, and the additional raising of Aswan *-ī-* may be due to the influence of the final high vowel. Previous analyses, focusing on Palestinian iterations of this development, have posited the presence of the diminutive infix in the primary etymological source of the items in question, i.e., deriving them directly from **hā l-waqayt* 'this little time' rather than **hallaqt* + *-ay-* '(diminutized) now' (e.g., Watson 2006). While either alternative is equally plausible for Jerusalem *halʔēt*, characteristics of the Malian and Egyptian forms argue strongly for the identification of *-ay-* in those cases as a secondary development applied to preexisting words for 'now' rather having been present as part of an incipient

periphrastic structure. The final /i/ of both Egyptian forms, as has been described in the discussion of the etymology **ðī l-waqt*, above, represents the idiosyncratically phonologized product of a regular epenthetic process triggered by the final cluster */qt/. Had the *-ay-* infix been present from the earliest stages of development, no such cluster would have existed to account for the genesis of /i/ in the contemporary forms. This leads to the conclusion that the application of diminutive *-ay-* must have occurred following the completion of any processes of epenthesis and phonologization. Further, the irregularly devoiced /k/ of the Aswan from (vs. regular /g/), is difficult to explain save as the product of assimilation to an immediately adjacent /t/, adding additional weight to the contention that an original descendent of unmodified **ðī l-waqt* existed in the variety which was subsequently augmented through the application of diminutive morphology. Perhaps even more clearly, diminutized Mali *ðrayk* is derived not from **ðā l-waqayt* but rather from a direct application of diminutive morphology to pan-Saharan **ðark* (also present in the dialect as *ðark*, an additional form meaning 'now'): the position of *-ay-* between /r/ and /k/ – as opposed to */q/ and */t/ for the Egyptian and Palestinian forms – indicates that at the time the morphology was applied **ðark* had already come to be analyzed as a monomorphemic entity with triconsonantal root *ð-r-k* (hence *ðrayk* via the regular diminutive pattern CC*ay*C).

For the above reasons, the Malian and Egyptian diminutized forms will be treated as definitively secondary in nature. Such an account is equally plausible for their Palestinian counterpart, though the lack of similarly conclusive diachronic evidence in that case means the present analysis means that the relative ordering of development cannot be known with certainty.

3.3 ANALYSIS

Having presented etymologies for the temporal adverbs 'now' attested by the Arabic dialect sample, as well as having worked to distinguish primary etymological sources from secondary morphological developments, I now turn to examine the data for evidence of CIG on the basis of the study's three-part heuristic. In the first subsection, I address the heuristic's first and second conditions by evaluating the developments observed above as potential products of grammaticalization and assigning them, when possible, to higher level groupings of evolutionary pathway. I then proceed to the heuristic's third condition by analyzing the geographic incidence of any multiply-attested grammaticalization paths uncovered.

3.3.1 Evaluation of Grammaticalization Status and Multiply Attested Pathways

Below, I utilize Heine's four component processes of desemanticization, extension, decategorialization and erosion to identify examples of grammaticalization present among the etymological derivations identified in §3.2 and address condition (i) of the heuristic. As appropriate, I organize these derivations as analogous, individual realizations representing more general paths of development, of which three are identifiable within the data; these groupings are of relevance to the evaluation of our heuristic's condition (ii).

3.3.1.1 'Now' from Time Noun and Demonstrative (NOW < THIS TIME)

By far the largest single source of Arabic temporal adverbs 'now' consists of source constructions signifying something approximating 'this time', with three distinct demonstrative elements combining with five different lexemes meaning 'time' to create

eight specific realizations attested in the current sample. This abundance is perhaps unsurprising given the importance of exactly such zero-marked NPs consisting of a time noun and modifier as a diachronic source of temporal adverbials worldwide, as noted by Haspelmath (1997). From the grammaticalization literature, this trajectory falls under the broader heading sketched by Heine and Kuteva's (2002) grammaticalization pathway TEMPORAL < TIME.

As to whether these Arabic developments may be taken as examples of grammaticalization, evidence for processes of desemanticization, extension, decategorialization and erosion is observed as indicated in Table 33.

Source	***Desemanticization?***	***Extension?***	***Decategorialization?***	***Erosion?***
**ðī l-waqt*	9/9	9/9	9/9	7/9
**ðā l-waqt*	26/26	26/26	26/26	26/26
**hā l-waqt*	12/12	12/12	11/12	11/12
**ðā l-ħīn*	14/14	14/14	10/14	10/14
**hā l-ħīn*	19/19	19/19	7/19	7/19
**hā s-sāʕa*	15/15	15/15	15/15	15/15
**hā l-ħazza*	2/4	2/4	1/4	1/4
**ðī l-ʔawān*	2/2	2/2	2/2	2/2

Table 33: Occurrence of Diagnostic Grammaticalization Processes, NOW < THIS TIME

With the exception of Bahrain *halħazza* and Kuwait *halħazza*, discussed separately below, all reflexes representing products of the path NOW < THIS TIME may be seen to have undergone desemanticization through bleaching of the specifying deictic

function of the original demonstrative; the result is an unrestricted temporal adverb, the reference of which spans the full range from unrestricted, point-specific instantaneity to unrestricted, generally obtaining present state. The analogous loss of specification in the development of Latin *hac hora* '(at) this hour' to Portuguese *agora* 'now', Spanish *ahora* 'now' is used by Lehmann (2015: 136) as an illustrative example in his description of desemanticization as a diachronic process. This loss of pragmatic force in the Arabic examples is perhaps demonstrated most clearly in dialects of the Sudanese area, where generic forms for 'now' from THIS TIME are reinforced with an additional, synchronically productive demonstrative element to further specify their reference: Khartoum *hassaʕ* 'now', *hassaʕ da* 'right now' (lit. 'this now'). Semantic generalization would seem to have affected reflexes of **hā l-ħazza* to a significantly lesser degree than those of the other etymological sources, Bahrain *halħazza* and Kuwait *halħazza* both being glossed as 'at this moment' and Abu Dhabi *halħazza ~ alħazza* alternately as 'now, at this moment' and 'now'.

Regarding extension, the generalization in value observed across the body of NOW < THIS TIME forms is accompanied by the expansion of their use to novel contexts covering the full semantic range described above:

3) Dhofar (Davey 2016: 263)
   ftʿan -t il- qisʿsʿa ðaħħīn
   remember.PFV-1SG DEF-story now
   'I remember the story now.' (lit. 'I remembered the story (just) now.')

4) *mā ħad yi-staxdām hāðī l- kitāb ðaħħīn*
   NEG one 3- use.IPFV DEM.SG DEF-book now
   'No one uses this book now.'

As seen in Table 33, the majority of reflexes also undergo decategorialization. The clearest evidence of this generally rests in the loss of compositionality of the original noun phrase, as convincingly demonstrated through the erosion of phonological form such that the three component elements of the source construction – demonstrative, definite article, and time noun – are no longer discretely identifiable. Examples of such degraded morphosyntactic autonomy of internal elements include Algiers *dork* < **ðā l-waqt*, Ḥarb *ðuwān* < **ðī l-ʔawān*, Ẓafīr *hāħīn* < **hā l-ħīn*, among others. This loss of compositionality is most obvious among the reflexes of **ðī l-waqt*, **ðā l-waqt*, **ðī l-ʔawān*, and the Sudanese descendants of **hā s-sāʕa*, which all present schemes of demonstrative construction developmentally distinct from those synchronically productive in the dialects in which they are used – though at least in those areas where a written record is available, these schemes have been verified as having existed at an earlier stage (Zack 2009 for Egypt; Bin Sharīfah 1987 for western North Africa). The departure of these former noun phrases from nominal categorial status is also demonstrated by the unacceptability of their use in contexts amenable to their respective source constructions. An example is quantification: in the dialect of Damascus, the compositional phrase *hā l-waʔt* 'this time' may be quantified with *kill* 'all' as in the title of the Noura Murad (2000) play *Baʕd kəll hā l-waʔt* 'After all this Time'; the corresponding structure with cognate Damascus *hallaʔ* 'now' in place of *hā l-waʔt*, however, is deemed ungrammatical (Rama Hamarneh p.c.).

As seen in the above table, phonological erosion is displayed by reflexes of all eight etyma discussed here, though those of **hā l-ħīn* and **hā l-ħazza* are affected at a noticeably lower rate (7/19 and 1/4, respectively); in the case of **hā l-ħazza*, at least, this fact may be related to their lower degree of desemanticization, as the only eroded reflex, Abu Dhabi *alħazza*, is also reported to share more generalized semantics. As described

in detail in §4.2, erosion most commonly consists of the deletion of initial (Abu Dhabi *alħīn*), medial (Aden *daħīn*) or final phonological material (Baghdad *hassa*), but under the broader heading of idiosyncratic change may also be seen the irregular deformation of consonants (Tlemcen *derwaq*, Kabābīsh *taħīn*) or the sporadic phonologization of an anaptyctic vowel (Cairo *dilwaʔti*).

In light of the evidence identifying processes of desemanticization, extension, decategorialization and erosion across the reflexes of the various etymological derivations representing the diachronic path NOW < THIS TIME, it appears warranted to identify the former as products of grammaticalization and the latter as a cross-linguistically corroborated grammaticalization path. Moreover, this pathway is attested by multiple discrete developments based on distinct but synonymous etymological sources. Therefore, conditions (i) and (ii) of this study's heuristic are met, and the evaluation of this family of forms as potential products of CIG will continue with the analysis of their geographic distribution in §4.3.2.

3.3.1.2 'Now' from Time Noun and Definite Article (NOW < THE TIME)

An additional diachronic source of temporal adverbs 'now' in the Arabic data is that of constructions originally meaning 'the time', composed of a noun with a meaning of 'time' modified by the definite article in a form consistent with its general local realization. Similarly to the path NOW < THIS TIME examined above, NOW < THE TIME is consistent with the cross-linguistically observed pattern of deriving temporal adverbs from zero-marked, modified NPs (Haspelmath 1997), and likewise is subsumed by the path identified as TEMPORAL < TIME in Heine and Kuteva's (2002) survey of grammaticalization pathways worldwide.

Relevant to the evaluation of the Arabic reflexes as examples of grammaticalization, the apparent incidence of processes of desemanticization, extension, decategorialization and erosion in the data are summarized in Table 34.

Source	***Desemanticization?***	***Extension?***	***Decategorialization?***	***Erosion?***
**al-ħīn*	12/12	12/12	1/12	1/12
**al-ħazza*	6/6	6/6	2/6	4/6
**as-sāʕa*	1/1	1/1	1/1	1/1

Table 34: Occurrence of Diagnostic Grammaticalization Processes, NOW < THE TIME

Desemanticization is evidenced across the body of examples through the loss of specification originally indicated by the definite article **al-*, the referent of the time noun which it modifies being neither uniquely specified nor identifiable on the basis of previous context. This clear loss of original definite article's semantic content is accompanied by the extension of the forms' use to pragmatic contexts in which a significant part of their function is to signal the introduction of new information, thereby clearly demonstrating a lack of connection to any previous referential frame:

5) Negev (Henkin 2016: 295-296)
dayman b- yi-sʔal -uw: līʃ? līʃ? līʃ? līʃ intuw kiðiy? …
always CNT-3- ask.IPFV-PL why why why why SBJ.2PL so
'They keep asking: Why? Why? Why? Why are you like that? …
maʕ inn -ih fīh aʃyāʔ alħīn itɣayyarˁ -at kiθīr
with COMP-OBJ.3SG EXIS things now change.PFV-3FSG much
Although there are things that have changed considerably now.'

While desemanticization and extension may be shown in this manner, definitive evidence of decategorialization among the products of NOW < THE TIME is noticeably more scarce. The test of compositionality employed to evaluate the products of NOW < THIS TIME above is not readily applicable in this case, as the smaller number of constituent parts and generally lower rate of word-internal erosion result in fewer unambiguously noncompositional forms in which the morphosyntactic operators of the original source construction are no longer individually analyzable. In spite of this fact, three such forms are identifiable in the current data set and may be interpreted as having undergone decategorialization: Ẓafīr *aħīn* (< **al-ħīn*) and Soukhne *alħaz*, *alħa* (< **al-ħazza*). In the first form, the original definite article is eroded such that the multimorphemic interpretation of the source construction is comprised. In the latter two, the erosion affecting word final segments has progressed to the degree that the resulting sequences /ħaz, ħa/ are too abbreviated to constitute viable nouns vis-à-vis the attested morphological templates of the language (Behnstedt 1994); thus, an analysis of *alħaz*, *alħa* as sequences of definite article and noun is not possible and the source lexemes must be viewed as having lost their nominal categorial status. Beyond these examples, Malta *issa* may be shown to have decategorialized through the lack of expected combinatorial allomorphy of the original definite article as compared with its source construction. While nominal *is-siɛa* 'the hour' combines with the prepositional proclitic *sa-* 'up until' to produce *sa-s-siɛa* with elision of the /i/ of the definite article (Borg & Azzopardi-Alexander 1997), the combination of *sa* and *issa* results in *s-issa* 'so far, until now' (Camilleri 1997).

The results of phonetic erosion are noted in the majority of reflexes of **al-ħazza* and **as-sāʕa*, generally occurring regressively from word final position: Antiochia *alħaz*,

Malta *issa*. This process is, however, notably absent from the reflexes of **al-ħīn*, with only Ẓafīr *aħīn* displaying evidence of irregular sound change.

In light of these findings, the derivatives of **al-ħazza* and **as-sāʕa* appear to be strong candidates to represent the products of grammaticalization, as they show signs of the four diagnostic processes of desemanticization, extension, decategorialization and erosion. Reflexes of **al-ħīn* have undergone the first two of these processes relating to changes in semantic value and incidence, but on the whole show only minimal evidence of those involving formal aspects, whether morphosyntactic or phonological. In this respect, they may be seen as analogous to the **ydawr*-derived 'want' futures discussed in Chapter 2, which were viewed as representing a relatively less advanced stage of grammaticalization than other FUT < WANT counterparts (§2.3.1.2). As in that case, despite the lack of formal change, the evident processes of desemanticization and extension are enough to support a claim of (at least incipient) grammaticalization of **al-ħīn*, as the grammaticalization literature typically views decategorialization and erosion as (chronologically) secondary processes to desemanticization and extension (cf. Heine 2007; Haspelmath 1999), and numerous cases of grammaticalization worldwide have been accepted that rested on the latter two alone. Thus, all three realizations of the multiply-attested NOW < THE TIME pathway may be seen to meet conditions (i) and (ii) of the heuristic and will be considered in the analysis of geographical distribution.

3.3.1.3 'Now' from 'Immediately, Right Away' (NOW < IMMEDIATELY)

The third and final source of terms signifying 'now' identified in the sample is that consisting of adverbs originally meaning something like 'immediately, right away'. This evolutionary trajectory is not well attested in the cross-linguistic literature, though it

should be noted that the only major study of the diachronic origins of temporal adverbs globally (Haspelmath 1997) makes a methodological choice to limit its scope of inquiry to non-adverbial source constructions only. Thus, it may be that the lack of reported cross-linguistic parallel for the NOW < IMMEDIATELY path is at least partially an artifact of the structure of the limited available research.

Two representatives of this proposed pathway are found in the current sample, and are evaluated below for evidence of desemanticization, extension, decategorialization and erosion, starting with the results presented in Table 35.

Source	***Desemanticization?***	***Extension?***	***Decategorialization?***	***Erosion?***
**ðāba*	7/7	7/7	2/7	0/7
**tawwa*	16/16	16/16	4/16	6/16

Table 35: Occurrence of Diagnostic Grammaticalization Processes, NOW < IMMEDIATELY

Desemanticization is evidenced in the data via the generalization of immediate time reference to a broader value of present state, a process analogous to that identified for the desemanticization of immediate or imminent future tense markers discussed in §1.3.1 and §2.3.1.1. This increased semantic abstraction and bleaching of pragmatic force is accompanied by extension to novel contexts in which the former value 'immediately' could not occur, such as in the description of atelic, generally obtaining states:

6) Marrakech (Sánchez 2014: 246)
dāba hād əf- ʃi ma bqa -ʃ
now DEM DEF-thing NEG remain.PFV-NEG
'Now this thing doesn't exist anymore.'

7) Sousse (Talmoudi 1981: 54)
u tawwa msimm -īn -u li- mbʿāwitʿ
and now name.PTCP-PL-OBJ.3MSG DEF-pop.eyed
'And now we call him the pop-eyed.'

This semantic bleaching is further evidenced in several dialects by use of reduplicated forms to express a reinforced sense of immediacy, as in J-Tunis *tawwa tawwa*, Casablanca *daba daba* 'immediately, right now'. The more specified, imminent meaning is thus replaced by a broader value of simultaneous time reference, and the contextual distribution of the items expands accordingly.

In comparison to these relatively transparent processes of desemanticization and extension, decategorialization of the products of NOW < IMMEDIATELY is significantly more difficult to demonstrate because both the proposed source constructions and the modern reflexes are and remain temporal adverbs. This is not to say, however, that all temporal adverbs are categorically equivalent, and the deictic quality of 'now' brings with it a novel referential character not shared with 'immediately, right away'. Owens discusses formal consequences of this function for Benghazi *tawwa*, noting that it (along with a small set of similarly deictic temporal adverbials such as *amis* 'yesterday') shares many distributional features with nominals and may consequently appear in complementation structures off limits to other adverbs, for example *laʕanid tawwa* 'until now' (Owens 1984: 188). This shift away from the distributional patterns of lexical

adverbs and toward that of a closed class of temporal deictics appears to represent a form of decategorialization associated with the assumption of a less lexical, increasingly functional role. Though not explicitly analyzed in other descriptions of dialects containing products of NOW < IMMEDIATELY, an examination of those sources containing substantial texts reveals evidence of a similar development in at least five additional Arabic varieties: Tripoli *zēy təwwa* 'like now, for example' (Pereira 2008: 391), Casablanca *mən daba* 'from now' (Harrell 1966: 18), Marrakech *l dāba* 'till now' (Sánchez 2014: 314), Tunis *ħattā l-tawwā* 'until/up to now' (Ben Abdelkader, Ayed & Naouar 1977: 259), Sousse *l-tawwa* 'till now' (Talmoudi 1981: 32). Given the relatively common occurrence of this change in varieties for which sufficient textual material exists to attempt an assessment – six of eight sources surveyed here – it is likely that the actual rate of decategorialization among the reflexes of **ðāba* and **tawwa* is significantly higher than indicated by the results presented in Table 35, though a definitive determination for the remaining reflexes awaits additional data.

Phonological erosion is attested at a comparatively low rate among the products of NOW < IMMEDIATELY, with a marked distinction between the evolution of reflexes of the two individual etymologies. While six of sixteen reflexes of **tawwa* show erosion in the form of regressive, irregular deletion of word-final segments, as demonstrated by variable Tozeur *towwa* ~ *toww* ~ *tow* ~ *tū*, no reflexes of **ðāba* display evidence of phonological adaptation outside the framework of regular sound change. It is conceivable that the continued association of these reflexes with the broader class of adverbs has exerted a conservative influence on the progress of word-final erosion, as the adverb-marking suffix *-a* thus remains a valid indicator of word class throughout the course of evolution and may resist deletion on etymological grounds. This interpretation does not, however, account for the observed differences between reflexes of **ðāba* on the

one hand and **tawwa* on the other as adverbial *-a* is active in dialects containing both: Casablanca *dīma*, Tripoli *dīma* 'always', and therefore represents only a partial explanation at best.

The products of NOW < IMMEDIATELY show clear evidence of desemanticization and extension, and further display indicators of decategorialization among those reflexes for which extensive textual evidence is available. Erosion is attested only among reflexes of **tawwa*, and then at a rate lower than fifty percent. The occurrence of the first three diagnostic processes of grammaticalization, combined with the marginal occurrence of the fourth, recommends that these forms be identified as the products of grammaticalization on a basis similar to that described for the derivatives of **al-ħīn* above, thereby meeting condition (i) of the study's heuristic. Condition (ii) is subsequently met by the presence of two etymologically distinct realizations of the relevant grammaticalization pathway, and the products of NOW < IMMEDIATELY will therefore be addressed in the geographic analysis presented in the following section.

3.3.2 Evaluation of Geographic Distribution

The preceding analysis has identified three multiply-attested grammaticalization pathways which successfully pass conditions (i) and (ii) of the study's heuristic in the search for evidence of CIG, together comprising the entirety of forms signifying 'now' encountered in the current sample: NOW < THIS TIME, NOW < THE TIME and NOW < IMMEDIATELY. The geographic incidence of the individual realizations of these pathways is presented below in Figure 2, following the same conventions as utilized in previous isogloss maps.

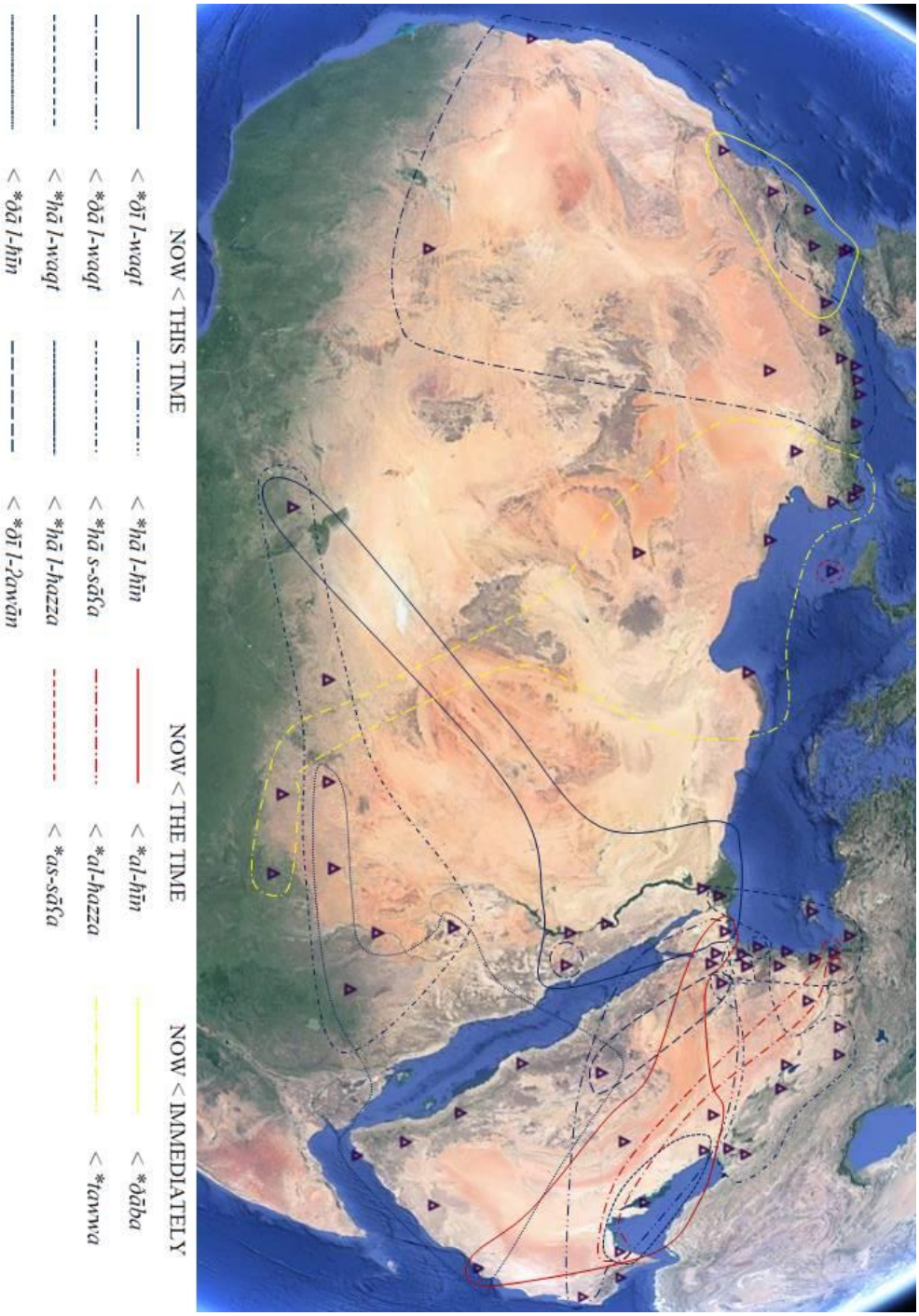

Figure 2: Geographic Distribution of Temporal Adverbs ‘Now’ (Map data: Google, S.O., NOAA, U. S. Navy, NGA, GEBCO; Image: Landsat/Copernicus)

As shown in Figure 2, grammaticalizations of NOW < THIS TIME predominate geographically as well as numerically, representing 101 of 143 reflexes considered and attested from Oman to Mauritania. Despite this impressive geographic range, products of the NOW < THIS TIME path remain geographically contiguous with one another across their entire area. In the east, derivatives of **hā s-sāʕa* typifying the Iraqi area bridge the Syrian desert to meet reflexes of **hā l-waqt* in the Levant, those of **ðī l-ʔawān* in Northwest Arabia, and the family of **hā l-ħīn* forms covering the northern portion of the Arabian Peninsula, dipping as far southeast as the territory of the Āl Wahībah in Oman and extending westward into the Sinai. In the eastern part of their range along the coast of the Persian Gulf, the latter forms occur alongside those deriving from **hā l-ħazza*. Mirroring the reflexes of **hā l-ħīn* in the southern portion of the Arabian Peninsula are reflexes of **ðā l-ħīn,* which also occur in adjacent regions across the Red Sea in eastern and central Sudan.

In the Sinai and the Nile Delta, reflexes of **hā l-waqt* and **hā l-ħīn* overlap with the **ðī l-waqt* forms typical of the broader Egyptian region to extend the incidence of items of the NOW < THIS TIME type. These in turn border a second set of **hā s-sāʕa* forms covering the greater Sudanic area and are co-territorial with a single reflex of **ðā l-waqt* in use among the ʻAbābdah near the Red Sea coast. In the far west of the Sudanic region we find reflexes of both **hā s-sāʕa*, characteristic of Sudanese Arabic at large, and typically Egyptian **ðī l-waqt*; the Egyptian connection is made all the more plausible here by the existence of multiple dialectological isoglosses linking this area with portions of Upper Egypt (see Owens 2003). Moving westward from here to the Saharan area, we encounter forms of **ðā l-waqt* which extend to from Mauritania and southern Morocco in the west to Mali in the east to Algeria in the north. Adjacent to the latter part of this area is a single, isolated attestation of **ðā l-ħīn* recorded in the Judeo-Arabic variety of Tunis.

Realizations of the pathway NOW < THE TIME similarly occupy a geographically contiguous area, with the exception of the sole derivation of **as-sāʕa*, Malta *issa*, isolated far to the west. The remaining forms are located in a region spanning much of the Arabian Peninsula and the Syrian interior, extending as far westward as Antiochia and the Sinai. The greater part of this zone is covered by reflexes of **al-ħīn*, which we find across much of Arabia stretching to include both the Red Sea and Persian Gulf coasts, as well as further north into the Sinai Peninsula and southern Jordanian territory. Neighboring these forms across the Syrian desert are derivatives of **al-ħazza*, which also attest a single reflex overlapping the opposite end of the **al-ħīn* area in Abu Dhabi.

The two realizations of NOW < IMMEDIATELY exist geographically separated from one another, though in a manner that suggests this may not always have been the case. Reflexes of **ðāba* occur ubiquitously across Morocco and neighboring Tlemcen in western Algeria, while we find forms of **tawwa* in an area covering modern Tunisia and Libya and extending an arm southward to include a portion of the western Sudan. In addition to the latter, one isolated reflex of **tawwa*, Al-Khaburah *taww*, occurs a significant distance away on the northern Omani coast. Though it is not possible to reconcile this last instance with an account of geographic diffusion linking it to the previous two areas, the latter may plausibly be seen to have once represented a contiguous zone of NOW < IMMEDIATELY reflexes reaching across North Africa, the original border between the **ðāba* and **tawwa* etyma now overlaid by the northward spread of **ðā l-waqt* forms moving upward from the Sahara. The striking uniformity of idiosyncratic phonological developments among the North African **ðā l-waqt* reflexes may be taken to support this interpretation, appearing as they do to represent the products of a single, shared process of etymological evolution and thereby simplifying an account of spread either from the Saharan region into a formerly NOW < IMMEDIATELY

dominant Algeria or from a single Algerian origin point outward to the distribution observed today. Thus, while we cannot identify the reflexes of NOW < IMMEDIATELY as geographically adjacent *a priori*, a possible history of contiguity may be inferred with only minimal secondary interpretation.

The results of this geographic analysis show that of the three grammaticalization pathways which meet our heuristic's first and second conditions, two also meet condition (iii) in a straightforward manner. The products of NOW < THIS TIME, by far the largest single source temporal adverbs 'now' in the Arabic dialect data, display complete geographic contiguity. Similarly, the distribution of forms of the NOW < THE TIME type is consistent with a history of areal diffusion across the Arabian and Syrian regions, though Malta *issa* is best excluded from this account. The two realizations of NOW < IMMEDIATELY do not meet condition (iii) in their present distribution. Though the data readily allows for an interpretation of adjoining geography in an earlier time period, until conclusive evidence is brought to light this proposal must remain speculative, and for the purposes of this evaluation this last set of forms cannot be said to meet the heuristic's condition (iii) in a strict sense.

3.4 Discussion and Conclusions

This chapter addressed the development of Arabic temporal adverbs meaning 'now' and evaluated the forms presented in the current sample for evidence of their evolution and diffusion via processes of CIG. First, each of the one hundred and forty-three reflexes considered was assigned an appropriate etymology, of which a total of thirteen were identified. In some cases, significant work was necessary to separate secondary developments from primary etymological origins. Once the details of these relationships

were established, the resulting derivations were considered with respect to the three conditions of the study's heuristic to assess their consistency with predicted characteristics of the products of CIG.

First, these results were studied for evidence of grammaticalization in their diachronic development, in the form of the component processes of desemanticization, extension, decategorialization and erosion, and grouped into a smaller set of semantically parallel evolutionary pathways. The individual realizations comprising the pathway NOW < THIS TIME, consisting of reflexes of **ðī l-waqt*, **ðā l-waqt*, **hā l-waqt*, **ðā l-ħīn*, **hā l-ħīn*, **hā s-sāʕa*, **hā l-ħazza* and **ðī l-ʔawān*, attest each of the diagnostic processes in good measure and are readily identifiable as grammaticalization processes, thereby meeting condition (i) of the heuristic. The realizations of NOW < THE TIME, **al-ħīn*, **al-ħazza* and **as-sāʕa* similarly meet the condition by the established criteria, though the comparatively lower rate of decategorialization and erosion is noted among reflexes of **al-ħīn*. The etymological derivations making up the third and final pathway NOW < IMMEDIATELY, **ðāba* and **tawwa* show convincing evidence of desemanticization and extension, but the identification of decategorialization processes is complicated by the adverbial nature of the source forms and the lack of extensive textual resources through which to analyze the items in question. However, signs of decategorialization are consistently found in reflexes for which sufficient data is available for analysis, and this suggests that the process may be significantly more prevalent than reported here. While phonological erosion is not uncommon in reflexes of **tawwa* it remains strikingly unattested among those of **ðāba*. As discussed, however, this fact does not necessarily preclude their identification as the products of grammaticalization as long as the evidence of desemanticization and other chronologically secondary processes is strong. Thus, the constituent realizations of NOW < IMMEDIATELY are claimed to

meet the heuristic's condition (i), while recognizing that the case for the reflexes of **ðāba* comprises less supporting evidence than that for those of developmentally analogous **tawwa*.

Regarding condition (ii), the requirement for multiply-attested grammaticalization pathways, all three evolutionary trajectories meet this criterion, with NOW < THIS TIME comprising eight distinct etymological realizations, NOW < THE TIME three and NOW < IMMEDIATELY two. In terms of the heuristic's third condition, that of a contiguous geographic distribution consistent with a history of diffusion via areal contact, all reflexes representing NOW < THIS TIME occur in either adjacent or overlapping fashion on the map and thereby meet the requirement. Reflexes belonging to the NOW < THE TIME grammaticalization pathway also show a high degree of geographic contiguity, the sole exceptional, isolated form being Malta *issa*, which apparently represents an independent development. Forms associated with the final pathway, NOW < IMMEDIATELY, do not meet the heuristic's third condition as such: the path's North African **ðāba* and **tawwa* reflexes are separated by a zone of **ðā l-waqt* derived forms in Algeria, and a lone reflex of **tawwa* exists significantly removed in Oman. While there are indicators that this geographic separation may not have obtained historically, we must for the moment refrain from identifying the geographic incidence sought by the heuristic and leave condition (iii) unmet in this case.

Despite this final shortcoming, the evidence summarized here provides sufficient basis to identify the forms comprised by the NOW < THIS TIME and NOW < THE TIME paths of development as the probable products of CIG. Under this scenario, at least one initial, innovative grammaticalization representing each of these paths was repeatedly replicated cross-dialectally until the respective results came to occupy their current areal distributions. Following from the geographic findings reported above, it is not possible

to say the same for the products of the path NOW < IMMEDIATELY, though for reasons previously discussed the door remains open to future interpretation in light of additional evidence. Regardless, the nature of the results of the first two grammaticalization pathways is sufficient to propose a significant role for CIG in shaping the modern Arabic forms for ‘now’.

Chapter 4: Genitive Exponents

4.1 Introduction

In this chapter, I assess the forms and functions of the Arabic genitive exponents for evidence of CIG as a mechanism underlying their diachronic development. The label "genitive exponent" is typically applied by Arabists to the set of analytic genitive particles which have emerged nearly ubiquitously across the body of modern Arabic varieties as an alternate means of marking possessive relationship operating in parallel to the inherited system of enclitic possessive pronouns. Alongside the future tense markers just discussed, the genitive exponents are among the most commonly cited examples of Arabic pluriform development, and the description of their formal and functional properties has been an object of significant interest in the field. Perhaps the most authoritative work on the subject to date is that of Eksell Harning (1980), who provides a comprehensive synchronic catalogue of attested genitive exponent forms with an emphasis on etymological origin and morphosyntactic behavior; this work is supplemented by a further diachronic investigation into the origin of that subset of exponents involving etymological relativizers (Eksell Harning 1984). Eksell Harning's reconstructions of source forms have contributed greatly to the analysis of etymological origins provided below and the results given in §3.2 build upon her earlier findings, though the present sample includes data points not examined in Eksell Harning's study and will additionally depart from her analyses in some significant ways. Complementing Eksell Harning's work, syntactic and semantic characteristics of the genitive exponents are examined in a comparative frame by Brustad (2000). Of particular relevance to the current investigation, the role of grammaticalization in the historical development of these items is briefly touched on by Rubin (2005), although his conclusions shall be seen to differ from those presented here in some key respects.

Following the format established in the previous chapters, I begin the examination of the genitive exponents by first providing the complete set of forms attested by the sample organized on the basis of etymological origin, tracing any observed formal changes linking reconstructed source forms to modern reflexes. Once this presentation and explication of the data is complete, I proceed to evaluate the results for their compatibility with an account of diachronic development based in CIG, utilizing the three-part heuristic laid out in §1.4. This includes consideration of the products of each etymological source as examples of grammaticalization, based on the identification of cross-linguistic counterparts as well as the presence of the characteristic component processes of desemanticization, extension, decategorialization and erosion. Subsequently, any multiply-attested pathways of development which plausibly comprise the products of replication via CIG are analyzed geographically to determine whether their contemporary distribution tallies with a historical account of areal diffusion. This analysis is followed by discussion of the collected evidence and a final assessment of the role of CIG in the development of the Arabic genitive exponents.

4.2 Attested Forms by Etymology

Provided below is a complete accounting of genitive exponent forms observed in the study's sample. These are organized on the basis of putative source etyma, accompanied by a description of lexical/structural origins and a brief sketch of geographical incidence. The relevant phonological and morphological changes resulting in the contemporary realization of individual reflexes are illustrated, as well as additional features noteworthy from a descriptive standpoint.

**matāʕ*

Algiers *mtaʕ*	Tozeur *mtāʕ ~ tāʕ ~ t ~ aʕ ~ ntīʕ*	M'zab *mtāʕ* [+]
Beirut *tāʕ ~ tāʕūl*	Kadugli *bitā*	Saïda *ntāʕ* [+]
B'ērāt *ibtāʕ* [+]	Fezzan *mtāʕ* [+]	Saoura *mtāʕ ~ ntāʕ*
Cairo *bitāʕ* [+]	Aswan *bitāʕ*	Djidjelli *mtāʕ*
Nouakchott *ntaʕ*	Kharga *bitāʕ*	J-Fez *ntāʕ*
Amman *tāʕ* [+]	'Awāmrah *bitāʕ*	Benghazi *imtāʕ* [+]
Khartoum *bitāʕ* [+]	Shukrīyah *bitāʕ ~ butāʕ* [+]	'Abābdah *bitāʕ*
Malta *ta*	Sousse *mtāʕ*	Azru *ntaʕ*
Casablanca *mtāʕ*	J-Tunis *ntāʕ ~ tāʕ*	Marrakech *ntāʕ ~ tāʕ* [+]
Sinai *btāʕ ~ tāʕ*	Cherchell *ntāʕ*	Mali *ntāʕ* [+]
Tripoli *mtāʕ* [+]	Tlemcen *ntāʕ*	Dellys *ntaʕ ~ taʕ* [+]
Tunis *ntāʕ*	Larbaâ *ntāʕ* [+]	Mateur *mtāʕ*

Table 36: Genitive exponents from **matāʕ*

Genitive exponents deriving from **matāʕ* are extremely common in the sample, attesting a total of 47 distinct reflexes. The source form is a noun meaning 'property, belonging(s)'. These forms occur in broad area encompassing the entirety of North Africa, extending southward across the Sahara to include Mali and Mauritania. Their occurrence continues across the Nile Valley as far south as the central Sudan and extends into the Levantine area to reach Lebanon and Jordan. The word *matāʕ* is attested from Classical Arabic in the meaning of 'property, belonging(s), possession(s)'. The intent is clearly physical, tangible belongings as they are often described as being "gathered" or

"presented", a sense further reinforced by the word's common collocation with the lexemes *bayt* 'house(hold)' and *ɣanāʔim* 'livestock': *matāʕ-u l-bayt-i* (property-NOM DEF-house-GEN) 'household belongings', *al-matāʕ-u wa l-ɣanāʔim-u* (DEF-property-NOM and DEF-livestock-NOM) 'property and livestock'; the lexical root from which **matāʕ* derives, **m-t-ʕ*, is associated with verbs meaning 'enjoy', and indeed classical definitions often reference this semantic connection by further describing *matāʕ* as 'the things which one enjoys, has privilege of' (al-Farāhīdī, *Kitāb al-ʿAyn*). As a lexical noun, **matāʕ* is obsolete and largely unattested in modern Arabic varieties: vestiges of substantive usage remain in frozen expressions in Morocco – *huwa bə-mtaʕ-u* (SBJ.2MSG with-property-GEN.3MSG) 'he is pretty well off' (lit. 'he is with/by his possessions/wealth') (Harrell 1966: 89) – and Hinds and Badawi (1986) record *matāʕ* 'luggage' for Egypt, though simultaneously flagging the term as a potential borrowing from Modern Standard Arabic by indicating that it is typical of "educated" speakers. Thus, the proposed development from the original source noun would need to have begun sufficiently early as to precede this subsequent lexical loss in the modern dialects.

In terms of phonological development, changes to the first syllable are common, at times resulting in deletion of said syllable entirely: Marrakech *tāʕ*, Malta *ta*, Amman *tāʕ*. More common, though, are changes to the realization of the initial consonant and vowel. Regarding the vocalic evolution, involving the */a/ of **matāʕ*, it is notable that no reflex drawn from the current sample maintains this */a/ as /a/. In some cases, localized to the greater Nile Valley, a quality change occurs whereby original */a/ is reflected by modern /i/ (Cairo *bitāʕ*, Khartoum *bitāʕ*) and in one case /u/ (Shukrīyah *butāʕ*). This change is irregular in all cases, save that Shukrīyah *butāʕ* could be derived from an intermediate **bitāʕ* via regular means (and, indeed, *bitāʕ* is variably attested in the dialect). The resyllabified forms Benghazi *imtāʕ* and B'ērāt *ibtāʕ* may similarly both be

accounted for via regular processes if a prior sporadic change to **bitāʕ* is posited. Most commonly, the vowel of the first syllable is deleted altogether, as in Fezzan *mtāʕ*, Azru *ntaʕ*, Sinai *btāʕ*. Across coastal North Africa, this change can be attributed the general deletion of unstressed short vowels, in the Saharan area to the deletion of short vowels in non-final open syllables, and in Sinai to the loss of /i, u/ specifically in open initial syllables (this last case, then, providing indirect evidence of a previous change to **bitāʕ*). Though no traces of original */a/ thus survive in the modern data, it is corroborated by medieval and early modern attestations of the Andalusian reflex of the term in both Arabic and Latin characters: <ħubzena matá culliém> *xubz^{i}-nā matāʕ kull yōm* (bread-GEN.1PL *matāʕ* every day) 'our daily bread' (Corriente 1977: 125). Additionally, as late as 1812 Mīkhā'īl Ṣabbāgh gives in his description of the spoken Arabic of Egypt and the southern Levant *matāʕ ~ batāʕ*, provided in the Arabic script of the 1886 published version with a clearly voweled /a/ (Ṣabbāgh 1886: 26, 27).

The initial consonant of the reflexes in question is also frequently affected by sound change, in this case always irregular. Common throughout the Nile Valley is a sporadic denasalization of initial */m/ > /b/, as in 'Abābdah *bitāʕ*. While idiosyncratic change of */m/ > /b/ is encountered for a few individual lexemes in the southern part of this area, for example Khartoum *bakān* < **makān* 'place' (Tamis & Persson 2013), its occurrence is quite rare and not typical of the majority of the geographic zone in question. The evidence just cited from Ṣabbāgh would indicate that initial /m/ and /b/ in these forms were in variation in the usage of "some" Egyptian and Levantine speakers in the early decades of the nineteenth century: "*wa ba'ḍu ahli miṣra ma'a ahli l-shāmi yalfaẓūna hādhihi l-mīma min matā' bā'an fa-yaqūlūna batā'ī*" 'and some of the people of Egypt along with the people of the Levant pronounce this *miim* of *matā'* as a *baa'*, so they say *batā'ī* (mine)' (Ṣabbāgh 1886: 27). Common in North Africa from the coast to

the interior we also find sporadic regressive place assimilation of */m/ > /n/ under the influence of following /t/, as demonstrated by Larbaâ *ntāʕ*.

Phonological change beyond the first syllable of these forms is relatively rare. A general loss of contrastive vowel length has affected some North African forms such as Algiers *mtaʕ*, and both Malta *ta* and Kadugli *bitā* demonstrate the results of a regular loss of /ʕ/. Tozeur attests an array of forms displaying varying degrees and routes of phonetic erosion, as well as one case of unexplained shift from */ā/ > /ī/. Of these, *mtāʕ* and *tāʕ* show counterparts elsewhere, while *t, aʕ* and *ntīʕ* are unique in the sample.

Forms marked with the symbol [+] display morphosyntactic agreement with the noun heading the genitive phrase in which they operate, as seen in the inflected forms of Saïda *ntāʕ* displaying concord with feminine singular *bagra* 'cow' and plural *bgər* 'cows': *el-bagra ntāʕ-ət qaddūr* (DEF-cow *ntāʕ*-FSG Caddour) 'the cow of Caddour', *el-bgər ntāwʕ əd-dawwār* (DEF-cows *ntāʕ*.PL DEF-camp) 'the cows of the camp' (Marçais 1908: 175). Given the nature of the data available it is likely that such agreement properties are significantly underreported, as a great number of descriptive sources simply provide a masculine singular citation form without explicit discussion of whether or not agreement phenomena are observed (even when they may subsequently be revealed in texts or numbered examples).

When agreement properties are displayed, the resultant paradigms most often cover the full range of values displayed by other nominal modifiers such as demonstrative adjectives. As for Saïda just above, this generally entails a masculine singular form, a feminine singular and a common plural, though in at least two cases, those of Benghazi and Khartoum, a distinct feminine plural agreement is documented; this is in keeping with general nominal agreement patterns in Benghazi and in traditional Khartoumi usage, though Dickins (2006) indicates feminine plural agreement,

presumably inclusive of the genitive exponent, is obsolescent in contemporary Khartoum. Dellys *ntaʕ ~ taʕ* is reported to show agreement for number only and Amman *tāʕ* for gender only, while the two Libyan varieties of Tripoli and Fezzan are reported to distinguish number in the feminine but not in the masculine, for example Fezzan *mtāʕt* (FSG) / *mtāʕāt* (FPL) but invariant *mtāʕ* (MSG and MPL). In at least the cases of Saïda *ntāʕ* and Marrakech *ntāʕ ~ tāʕ*, agreement patterning is reported to be optional (though it is possible that a pragmatic conditioning exists, as will be discussed for the better understood example of Damascus *tabaʕ* below). As for form, feminine singular and plural inflections match typical nominal/adjectival agreement morphology and are inflected with the local reflex of construct state feminine singular *-at*, *-it*, *-ət*, *-t* and feminine plural *-āt*, respectively. Masculine and common plural forms, however, display three (possibly four) distinct developments. Such forms in Mali, the Sudan, Upper Egypt and western Libya show "sound" affixal plural morphology in the shape of *-īn*, typical of derived nominals (e.g., Mali *ntāʕīn*). It is also plausible to interpret invariant Beirut *tāʕūl* as reflecting an earlier plural formed through addition of a suffix *-ūl*, as discussed below for Damascus *tabaʕūl* (CP), thereby indicating that a number distinction may once have been active in the dialect. More common than affixal morphology is the use of a "broken" plural pattern consisting of changes to the underlying morphological template. Across North Africa, predating regionally specific changes of */ā/ > /a/ and */a/ > /ə/, the pattern used is reconstructable as *(C)C*āwa*C (M'zab *mtāwaʕ*, Larbaâ *ntāweʕ*, Dellys *tawəʕ*), while in the Lower Nile Valley C*i*C*ū*C predominates (Cairo *bitūʕ*). It is of note that none of these correspond to the pattern *ʔa*CC*i*C*a*, that of the attested plural *ʔamtiʕa* shown by the original substantive in Classical Arabic and those dialects which retain it in its lexical sense.

**tabaʕ*

Beirut *tabaʕ*	Amman *tabaʕ* [+]	Aleppo *tabaʕ*
Damascus *tabaʕ* ~ *tabaʕīt* [+]	Negev *tabaʕ* [+]	Jisr az-Zarqa *tabaʕ* [+]
Jerusalem *tabaʕ*	Soukhne *tabaʕ*	

Table 37: Genitive exponents from **tabaʕ*

Throughout the Levant, genitive exponents are attested deriving from **tabaʕ*. I follow Eksell Harning (1980) and Rubin (2005) in reconstructing these item to an original substantive meaning 'belonging, possession', derived from the verbal root **t-b-ʕ* 'follow', though such a nominal is not currently in use in lexical form in the modern Arabic dialects. The Classical Arabic cognate *tabaʕ* is most often used in relation to human referents to express 'subject, follower' (al-Farāhīdī, *Kitāb al-ʿAyn*), though as Rubin observes the semantic extension of other derivations of the same root (such as participial *tābiʕun li-*) to encompass relationships of inanimate possession or belonging is already seen in classical usage (2005: 54; supported also by Lane 1968) and widespread in modern dialects as well (cf. Hinds & Badawi 1986; Qafisheh 1997).

As all modern reflexes captured by the current sample directly reflect the inherited phonological form of original **tabaʕ*, no interpretation of historical sound change is necessary in this case (on Damascus *tabaʕīt*, see immediately below). In terms of morphological development, four of the nine forms listed here are described as displaying agreement features relating them to the nouns which they modify. Jisr az-Zarqa *tabaʕ* and Negev *tabaʕ* show agreement for both number and gender, while Amman *tabaʕ* is recorded as agreeing only in gender and Damascus *tabaʕ* only in

number. Feminine singular value (or purely feminine, in the case of Amman), is marked by the local realization of the feminine singular construct state morpheme *-*at*: Jisr az-Zarqa *tabʕat*, Negev *tabaʕat*, Amman *tabaʕat*. This may also account for the irregular shape of common singular Damascus *tabaʕīt*, if the former is viewed as deriving from a historical feminine **tabaʕit*, with the sporadic lengthening of the last vowel perhaps occurring under the influence of the frequently attached clitic pronoun -*ī* (1SG), **tabaʕit-ī* > *tabaʕīt-ī* 'my, mine'. Plural forms are all suffixal and show considerable variation, in no cases corresponding to the broken plural of the substantive in its Classical Arabic form *ʔatbāʕ*. Jisr az-Zarqa *tabʕīn* marks plural with the regular sound plural marker -*īn*. The limited textual resources available for Negev Arabic attest the *tabaʕāt*, formed via the addition of the regular feminine plural suffix -*āt* – as the referent in this case is in fact feminine plural, it is unclear whether this inflection indicates gender as well or conveys a generalized plural value. The latter is the case for Damascus *tabaʕāt*, which despite its origin may be used in reference to either masculine or feminine plural nouns. More commonly attested as a common plural in Damascus is *tabaʕūl*. The origin of the suffix -*ūl* is far from clear, but a possible source of analogy might lie in the plural demonstrative *hadōl* 'these'. Were this hypothesis to be supported, it may in turn provide an alternative account for the anomalous /ī/ of *tabaʕīt*, hinting at a previous paradigm *tabaʕ* (MSG) / *tabaʕīt* (FSG) / *tabaʕūl* (CPL) on the pattern of *hāda* (MSG) / *hādī* (FSG) / *hadōl* (CPL). Lacking further evidence, however, this proposal remains speculative. In the contemporary dialect of Damascus, the application of number agreement morphology is reported to be variable; Brustad (2000) argues convincingly that such variation is motivated by the interaction of semantic and pragmatic factors, reflecting the intended degree of individuation of the plural head noun in a given circumstance.

**ħaqq*

Bahrain *ħagg*	Sana'a *ħagg*	Shukrīyah *ħagg* [+]
Khartoum *ħagg* [+]	Sinai *ħagg*	Abu Dhabi *ħagg* [+]
Kuwait *ħagg*	Kadugli *hagg*	Abha *ħagg* [+]
Mecca *ħagg* [+]	Ḥarb *ħagg* [+]	Aden *ħaqq*
Sudayr *ħagg*	Banī Ṣakhr *ħagg*	Dhofar *ħaqq* [+]

Table 38: Genitive exponents from **ħaqq*

Genitive exponents deriving from a noun **ħaqq* are found throughout the Arabian Peninsula as far north as the Northwest Arabian area, as well as across the Red Sea in Sudan. The reconstructed meaning of the original substantive is 'property, right, claim'. Classical Arabic *ħaqq* is found in these meanings from its earliest strata of documentation (al-Farāhīdī, *Kitāb al-'Ayn*), and cognates with similar definitions are extremely widespread in modern dialects from Morocco to the Persian Gulf (e.g., Harrell 1966; Woodhead & Beene 1967; Qafisheh 1997). In several varieties, perhaps significantly in areas – though not all areas – where reflexes of **ħaqq* have developed as genitive exponents, the deontic element entailed in the possessive relationship has largely been lost, leaving a more general meaning of 'property, possession' (as noted by Hillelson 1935 for the Sudan, Davey 2016 in Oman).

Phonological derivation of the various reflexes from the source etymon is fairly straightforward. The majority of forms follow general patterns of local sound change in realizing */q/ as /g/ (e.g., Banī Ṣakhr *ħagg*), the two exceptions being Aden *ħaqq* and Dhofar *ħaqq* in which */q/ is regularly maintained. The shift of /ħ/ > /h/ in Kadugli *hagg* is similarly a regular process.

Forms marked the symbol [+] above display morphological agreement with modified nouns. In the dialects of Abu Dhabi, the Ḥarb and the Shukrīyah, agreement is for gender only with feminine forms marked by the standard nominal/adjectival feminine singular suffix *-at*, for example Ḥarb *ħaggat*. In other varieties, agreement indicates both gender and number. In Mecca and Dhofar, this comprises a masculine singular, a feminine singular, and a common plural form, while in Khartoum and Abha masculine and feminine plural values are further differentiated. Feminine plurals in both of the latter varieties are formed via the addition of the general feminine plural suffix *-āt*, which in Abha is variably accompanied by the irregular change of /a/ > /u/ to result in the forms *ħaggāt* ~ *ħuggāt*, perhaps linked to stress shift. Masculine plural agreement is present in a variety of forms, none of which reflect the broken plural **ħuqūq* universally attested for the original substantive. In Khartoum and Abha, this value is expressed via the regular nominal/adjectival masculine plural suffix *-īn*, Abha once again showing the variation *ħaggīn* ~ *ħuggīn*. In Mecca, the common plural is expressed by *ħaggōn*, with one source (Schreiber 1970) mentioning a minority variant *ħaggīn* comparable to the forms just discussed.

The origin of the suffix *-ōn* is not obvious. While it may be tempting to relate it to the Classical Arabic nominative sound plural marker *-ūna*, this suffix is in no other way active in the dialect and such a derivation thus seems unwarranted; the same should be said of the verbal third person plural suffix *-ūna*, which is active in several other Peninsular varieties as *-ūn* but not in Mecca, which shows only *-ū*. The passing similarity to the plural demonstrative form *hadōl* may be worthy of mention; indeed such a parallel would open the door to possible comparison with the *-ūl* of Damascus *tabaʕūl* and Beirut *tāʕūl*, discussed above. In terms of accounting for the final /n/ in such a scenario, occasional examples of /l/ > /n/ word finally are encountered in the Meccan materials

presented by Ingham (1971), or alternatively this could represent influence from the more readily parsable *ħaggīn* variant recorded by Schreiber. In Dhofar, the common plural is indicated by *ħaqqūt ~ ħaqqōt*. Once again, a clear etymology is not evident for the suffix *-ūt ~ -ōt*. It may represent a blended form of an original feminine plural **ħaqqāt* influenced by /ū/-plural inflecting nominal modifiers like the demonstrative *hāðūn ~ hāðūl* (MPL); supporting such a mixed origin might be the observation that, despite the feminine plural existing as a distinct value across the remainder of Dhofari inflectional system, with regard to the genitive exponent the form *ħaqqūt ~ ħaqqōt* alone is used in both masculine and feminine contexts. Whatever the case, the origin of this form is almost certainly linked to that of the identical suffix in the more widespread plural exponent form *mālūt ~ mālōt*, which is in use in Dhofar and further afield and is discussed in greater detail below.

In at least Abu Dhabi and Dhofar, application of the agreement patterns just described is reported to be nonobligatory, though no attempt at a sociolinguistic or pragmatic account of this variable usage is made.

**māl*

Baghdad *māl* [+]	Kuwait *māl* [+]	Basra *māl* [+]
J-Baghdad *māl*	Sudayr *māl*	Abu Dhabi *māl*
Bahrain *māl* [+]	Khawaytnah *māl* [+]	Dhofar *māl* [+]
Khuzestan *māl* [+]	Hit *māl* [+]	Mosul *māl*

Table 39: Genitive exponents from **māl*

Genitive exponents from **māl* are encountered throughout Mesopotamia and the eastern portion of the Arabian Peninsula. Their source is an original substantive meaning 'property, possession(s)'. Classical Arabic *māl* occurs with this meaning from the earliest documented stratum of the language (al-Farāhīdī, *Kitāb al-'Ayn*), and reflexes of the same source and significance are ubiquitous across modern Arabic varieties. In many modern dialects, an additional meaning of 'money, (specifically monetary) wealth' has emerged alongside the more general meaning of 'property, possession(s)' (e.g., Harrell 1966; Hinds & Badawi 1986).

As all modern forms remain phonologically identical to the source etymon, it is not necessary to explicate any products of sound change here. With regard to morphological evolution, eight of twelve reflexes above definitively show morphosyntactic agreement properties not characteristic of the original noun. In the dialects of Bahrain, Hit and the Khawaytnah, gender agreement is displayed by way of the local reflex of the feminine singular suffix *-at*: Bahrain *mālat*, Hit *mālit*, Khawaytnah *mālət*. The exponents observed for Baghdad, Khuzestan, Kuwait, Basra and Dhofar additionally agree with their head nouns in number. Forms from neighboring Khuzestan and Basra inflect for both masculine plural and feminine plural values, utilizing the productive nominal/adjectival suffixes *-īn* and *-āt* to deliver such complete arrays as Basra *māl* (MSG), *mālat* (FSG), *mālīn* (MPL), *mālāt* (FPL). In Baghdad, the presumably once feminine plural *mālāt* has generalized to both masculine and feminine plural contexts. The Dhofar common plural *mālūt ~ mālōt* is analogous to *ħaqqūt ~ ħaqqōt* form previously discussed. Intriguingly in light of this form's unusual morphological shape, Brustad's data from Kuwait attests a clear example of a similar form *mālōt* modifying a nonhuman plural referent (2000: 82). Potentially problematic though this may be for the idiosyncratic blended form hypothesis presented above, especially given

the not insignificant distance between the two areas, no more attractive proposal currently presents itself. None of the above plural forms correspond to the broken plural of the original substantive, as represented by Classical Arabic *ʔamwāl*. In both Dhofar and Kuwait, both number and gender agreement are reported to be nonobligatory, and at least for the case of Kuwait Brustad demonstrates a pragmatic basis for this variability related to individuation, as described for Damascus *tabaʕ* above.

It should also be mentioned here that a genitive construction involving *māl* exists in neighboring Persian. The noun *māl* 'property, belonging', apparently borrowed from Arabic as a substantive in light of its (non-Persian) broken plural *amvāl* (< Arabic *ʔamwāl*), is utilized in the Persian *ezāfe* attributive construction to express predicative possession: *in ketāb māl-e man ast* (DEM book *māl*-GEN 2SG be.3SG) 'this book is mine' (lit. 'this book property-of me is') (Lambton 1974: 10). Though numerous differences between the Arabic and Persian structures exist and the development of each may be accounted for on purely internal grounds, the restricted occurrence of Arabic *māl* to the precise zone bordering the Persian language area should not go unnoticed and the investigation of potential links between functional uses of Arabic and Persian *māl* would represent a rich target of future investigation.

**jənā*

Fezzan *jnā* ~ *jən* [+]		

Table 40: Genitive exponents from **jənā*

A genitive exponent derived from **jənā* occurs in Fezzan in the Libyan interior alongside derivatives of **matāʕ*. Though lexical resources relating to this variety are

virtually nonexistent, the source would seem to be a noun **jənā* meaning something like 'what is gathered, harvest, rightful gains' on the basis of Classical Arabic *janī* of the same meaning (Ibn Manẓūr, *Lisān al-ʿArab*). Cognate forms in surrounding modern dialects – Egyptian *ganī,* Sudanese *ɟanī,* Moroccan *ʒnī* – more often function as verbal nouns, signifying '(the act of) gathering, harvesting' rather than its results (Hinds & Badawi 1986; Tamis & Persson 2013; Harrell 1966). Such a reading is possible for the Classical Arabic form as well, though the former meaning would seem to be the default interpretation, as in the line of poetry attributed to a third century Lakhmid king:

1) Classical Arabic (Ibn Manẓūr, *Lisān al-ʿArab*)
   hāðā janā -ya wa xiyār-u -hu fī -hi
   DEM.MSG gathered-GEN.1SG and best -NOM-GEN.3MSG in-GEN.3MSG
   'This is what I have gathered and its best is in it.'

An interesting semantic extension of the same root is found neighboring the Fezzan area in Sudanese *janā* 'baby, offspring' (Hillelson 1935).

Deletion of /ə/ in *jnā* may be explained by a regular loss of short vowels in initial open syllables described for at least some varieties of the Fezzan area; its preservation in the variant form *jən* is thus likely a consequence of the irregular deletion of final /ā/, resulting in a closed first syllable. Morphologically, the item is reported to have developed agreement properties, inflecting for the number and gender of the head noun of the genitive phrase. Attested inflections include *jnā ~ jən* (MSG), *jənt* (FSG), *jnī* (MPL), *jnāt* (FPL). Both feminine forms would appear to be based on the *jən* variant of the masculine singular, through the addition of the regular agreement morphemes *-t* and *-āt*. The masculine plural *jnī* could be interpreted as an eroded form of *jən* modified by the

regular masculine plural suffix *-īn*, but perhaps preferably is viewed as a pattern alternating plural of singular *jən*. Were this the case, then the entire inflected paradigm could be interpreted as directly analogous to that of the similarly phonologically and morphologically irregular **bən* 'son'; though only the feminine singular *bənt* 'daughter' and a diminutized feminine plural *bnaiyāt* 'little daughters' are directly attested in the extremely limited materials available for Fezzan Arabic, on the basis of neighboring dialects and sound changes described for the variety this paradigm would be constructed as **bən* (MSG), *bənt* (FSG), **bnī* (MPL), **bnāt* (FPL), thus directly parallel to attested *jən* (MSG), *jənt* (FSG), *jnī* (MPL), *jnāt* (FPL). Though this argument is primarily based on similarity in word form, any relevant analogy may have been bolstered by the common metaphorical use of 'son/daughter' across Arabic varieties to denote abstract semantic relations, often verging on genitive: Classical Arabic *ibn-u tˁ-tˁarīq-i* (son-NOM DEF-road-GEN) 'the traveler' (lit. 'the son of the road'), *ibn-u l-ħarb-i* (son-NOM DEF-war-GEN) 'the warrior' (lit. 'the son of war') (Lane 1968: 263).

**hana*

B'ērāt *ihnīn* [+]	Soukhne *hanayyi* [+]	Cameroon *hanā*
B-Kadugli *hān* [+]	Khawaytnah *hnīt* [+]	Nigeria *hana* [+]
Aswan *ihnīt*	Abéché *hana*	

Table 41: Genitive exponents from **hana*

In the West Sudanic area, Upper Egypt and interior Syria genitive exponents are encountered which derive from a noun **hana*, with a lexical meaning of 'thing'. This reconstructed source is based on the Classical Arabic *hana* 'thing' (al-Farāhīdī, *Kitāb al-*

‘Ayn). Cognates of this form are generally not active as lexical nouns in modern Arabic varieties (including those with **hana*-derived genitive exponents), though in some regions of contemporary Saudi Arabia *hana* and *han* are attested with a meaning of ‘thing, trifle’ (Behnstedt & Woidich 2012: 168).

The significant phonological variation observed across the modern reflexes would seem to stem from the unusual fact that the source etymon contains only two root consonants rather than the generally requisite three (indeed, Classical Arabic *hana* features prominently in discussions of the biconsonantal root in traditional Arabic grammar). Numerous strategies are evidenced to morphophonologically regularize inherited **hana* in such a way as to provide a third analyzable root consonant. In the dialect of the Baggārah population of Kadugli, the first /a/ is lengthened to /ā/, thereby providing a viable third root analyzable as underlying /w/ or /y/. Cameroon similarly reinterprets the final /a/ as /ā/, facilitated by the breakdown of vowel length distinctions word finally – forms involving clitic pronouns show Nigeria *hana* and Abéché *hana* to have historically undergone the same process, for example Nigeria *hanā-hum* ‘their’ (combined with 3MPL possessive pronoun -*hum*). In B‘ērāt and Soukhne and among the Khawaytnah, the final /a/ would seem to have been reanalyzed as an underlying /y/, once again increasing the number of valid roots to three. In B‘ērāt this appears to have been additionally reinforced through the reduplication of the original final root consonant /n/ (though this additional /n/ is not present in all inflected forms of the item, as shall be seen below). The Soukhne and Khawaytnah forms may be products of a secondary feminization through the addition of the morpheme *-*at*, perhaps under pressure from an inherited feminine value of the source lexeme incompatible with the analysis of final /a/ as /y/; in this scenario, Soukhne *hanayyi* would appear to represent a masculine back formation of its feminine singular inflection *hanayyt*. In terms of other phonological

processes, the final /i/ of Soukhne *hanayyi* is likely anaptyctic in origin, and the initial resyllabification of B‘ērāt *ihnīn* and Aswan *ihnīt* is the result of regular sound change. The final /t/ of the Aswan form is best understood as morphological in nature (see below).

All forms save those of Abéché and Cameroon show agreement patterning of some kind. Soukhne *hanayyi* agrees purely for gender through the addition of the feminine singular morpheme *-t* in the manner just seen, while Khawaytnah *hnīt* agrees for number but not gender via the addition of a formally feminine plural suffix *-āt*, delivering *hniyāt* (structurally quite similar to the attested Classical Arabic plural *hanawāt*, the difference lying in the reinterpretation of final /a/ as underlying /y/ vs. /w/). B-Kadugli *hān* inflects for gender and number in *hint* (FSG) and *hinē* (CPL). The former clearly involves the addition of feminine singular *-t*, and the latter likely represents a broken plural pattern applied to an extrapolated root involving final /y/ (/ē/ regularly reflecting */ay/). Comparable forms inflecting for gender and number are found in Nigeria. Though the typical feminine singular member of this paradigm actually represents a suppletive derivative of the etymology **hū l-*, discussed below, dialectal variants show a feminine singular in *hintá*, the distinctive stress assignment indicating an origin in **hintat* followed by semi-regular final /t/ deletion. This form may be interpreted as a double application of feminine morphology, perhaps prompted by the reanalysis of an original feminine *-t* as a the sought after third root, as is well documented for other biconsonantal root words such as *bint* ‘girl, daughter’ and *ʔuxt* ‘sister’ as a pan-Arabic phenomenon and, in fact, for Classical Arabic *hana* itself in a dialectal alternant *hant* (al-Farāhīdī, *Kitāb al-‘Ayn*). This proposal is bolstered by a minority variant of the common plural form, *hintāt*, which seems to similarly represent the application of (subsequently generalized) feminine plural morphology on top of a reanalyzed feminine singular

marker. More frequent as the plural of Nigeria *hana* though is *hinē*, on the same pattern as observed for B-Kadugli. B‘ērāt *ihnīn* shows a full range of gender and number values in its inflected forms *ihnīt* (FSG), *ihniyyīn* (MPL), *ihniyyāt* (FPL), which appear to reflect the addition of morphologically regular nominal/adjectival agreement suffixes to a reanalyzed base *ihniyy*-, absent the reduplicated final /n/ of the masculine singular. Though only documented while modifying a feminine singular referent, Aswan *ihnīt* hints at a similar system of inflection in matching the form and value of neighboring B‘ērāt *ihnīt* (FSG).

**ʃayyit*

Cyprus *ʃayt* [+]	Damascus *ʃīt* [+]	Khawaytnah *ʃīt* [+]
Jisr az-Zarqa *ʃīt*		

Table 42: Genitive exponents from **ʃayyit*

Throughout the Levant, genitive exponents occur deriving from an original **ʃayyit*. This would appear to represent a feminized form of pan-Levantine *ʃī* ‘thing’ < **ʃay* < **ʃayʔ*, with Cyprus *ʃayt* preserving a more conservative vocalism (hence the ultimate reconstruction to **ʃayyit*). Classical Arabic *ʃayʔ* is the variety’s most frequently attested word for ‘thing’ and known from an early date (al-Farāhīdī, *Kitāb al-ʿAyn*), and **ʃayʔ* would appear to be reconstructable to a common node of the classical language and all modern Arabic dialects, the majority of which contain reflexes of the form as a lexeme ‘thing’ (Behnstedt & Woidich 2012: 166-167) and many more as a grammaticalized negative polarity item turned negator (Lucas 2010). The addition of feminine -*it* is likely reflective of a process of specification/individuation (Brustad 2000), and may have been

additionally motivated by the morphophonological difficulty of attaching clitic pronouns to a base terminating in /ay, ī/ following the regular loss of word final /ʔ/. In all cases save Cyprus, the **ʃayyit*-derived exponents mentioned here are documented as minority forms, in retreat to **tabaʕ*-based forms in the southern part of their range and to numerous regional alternatives among the Khawaytnah.

The monophthongal forms observed for the mainland Levantine forms are indicative of an underlying structure of **ʃiyye*, which predictably becomes **ʃīt* in the construct state: compare the non-construct/construct alternation of Damascus *miyye*/*mīt* 'hundred'. The germination of /y/ here likely represents an attempt to restore a triconsonantal root to the lexeme following the deletion of /ʔ/, thus paralleling the permutations observed for forms of **hana* above. It is ambiguous whether Cyprus *ʃayt* sat out this process altogether or underwent similar gemination followed by the simplification of an unallowable final cluster (cf. Tsiapera 1969).

Morphologically, all forms show agreement phenomena save Jisr az-Zarqa *ʃīt* (though given the scarce documentation of this item this may well be the artifact of a lack of data). Damascus *ʃīt* and Khawaytnah *ʃīt* display number agreement with head nouns via plural forms *ʃyāt*, *ʃiyāt*, respectively. This could either be the product of application of the regular feminine plural prefix *-āt* prior to gemination of the final /y/ or of the sporadic degemination of said /y/. It is of note that among the Khawaytnah this plural closely parallels the observed plural form of the genitive exponent *hnīt, hniyāt*. Cyprus *ʃayt* displays a plural form *ʃat*, which could well be related to the mainland forms through the irregular deletion of /y/ and loss of vowel length. Uniquely among the forms listed above, singular Cyprus *ʃayt* may also be seen to convey a gender value as it is employed only with feminine singular referents, masculine singular referents modified by a suppletive item *tel* discussed below.

**ħājit*

Aswan *ħājt*		

Table 43: Genitive exponents from **ħājit*

Personal communication with Dr. Slah Alrawy of the Egyptian Academy of Arts, a native of Aswan and expert in Egyptian dialect poetry and lexicography, revealed a genitive exponent *ħājt* not previously reported for the area to my knowledge. Provided as a direct equivalent for variant Aswan genitive exponents *bitāʕ* and *ihnīt*, *ħājt* would seem to be a relatively local evolution, as no cognate forms are reported from the other Upper Egyptian sites included in the current study's sample. Aswan *ħājt* presumably relates to a form **ħājit*, the construct state of the Aswan lexical noun *ħāja* 'thing'. An apparent lexical extension of the meaning 'need, needed thing' attested by Classical Arabic *ħāja* and many modern dialect reflexes, the innovation of **ħāja* as 'thing' is characteristic of a broad swath of contemporary Arabic varieties encompassing North Africa, the Nile Valley, and western and southern portions of the Arabian Peninsula (Behnstedt & Woidich 2012: 166-167).

As all examples provided came immediately followed by a vowel-initial clitic pronoun (*-ī* (1SG) or *-ak* (2SGM)), it is unknown whether the loss of /i/ from presumed **ħājit* is fixed as a diachronic change or rather the simple product of regular synchronic rules governing syllable structure. Knowledge of the item's morphosyntactic agreement properties is similarly incomplete, as all contextualized examples modify a feminine singular referent (either *il-ʕarabiyya* 'the car' or *dī* 'this (F)').

**ʃuɣl*

Negev *ʃuɣl* [+]	Sinai *ʃuɣl* [+]	Kharga *ʃaɣl*

Table 44: Genitive exponents from **ʃuɣl*

Genitive exponents deriving from a noun **ʃuɣl* are found in the neighboring regions of the Negev and the Sinai as well as slightly further afield in the Egyptian oasis of Kharga. Their source appears to be a substantive **ʃuɣl* with a meaning of 'thing', itself an extension of an original significance 'work, occupation' common throughout much of the central Arabic-speaking world (e.g., Hinds & Badawi 1986; Stowasser & Ani 2004) and attested for Classical Arabic (Lane 1968). It is equally possible that these forms, or a subset of them, derive from a feminine version of the lexeme, **ʃuɣla*, as both masculine and feminine forms are extant in Arabic cross-dialectally. The feminine may be seen to be preferable in that it is the lexical form attested in the Northwest Arabian region where the genitive exponents in question are most focused (cf. Negev *ʃiɣlih* 'thing'), though at the same time documented lexical forms nearer to Kharga are masculine, with *ʃuɣl* and *ʃoxol* known from the Sudan and Chad, respectively (Behnstedt & Woidich 2012: 168). A masculine source form has been tentatively favored here in light of the fact that the inflected morphological variants described below are more simply accounted for as derived from a masculine base, with the acknowledgment that the feminine version of the etymology still remains a perfectly plausible alternative.

Despite its irregular status, the sporadic shift of */u/ > /a/ in Kharga *ʃaɣl* is best understood as a variation in lexical input rather than a stage in the development of the genitive exponent itself, as variation between /u/ ~ /a/ in the forms meaning 'thing' is attested both across and within dialects (Behnstedt & Woidich 2012: 168) and even

within the Classical Arabic corpus (Lane 1968). Negev *ʃuɣl* and Sinai *ʃuɣl* display agreement patterning with a modified noun, attesting the inflected forms Sinai *ʃuɣlit* (FSG), *ʃuɣlīn* (MPL), *ʃuɣlāt* (FPL), Negev *ʃuɣlit* (FSG), *ʃuɣlīn* (CPL) marked by means of generally productive nominal/adjectival agreement suffixes. No specifically feminine plural form is described for Negev in the data available, despite the general vitality of such a distinct value in the dialect otherwise (indeed, even for the variant genitive exponent Negev *tabaʕāt* (FPL)). No agreement properties are specifically noted for Kharga *ʃaɣl*, though the survey format of the relevant descriptive material is quite vague on this point.

**aððī li-*

Algiers *dyal*	Cherchell *dyāl*	Tetouan *dyal ~ d*
Anatolia *ðīla ~ ðīl ~ ðēla ~ ðēl*	Tlemcen *dyāl ~ dī ~ əddi*	Azru *dyāl ~ d*
Cyprus *tel ~ te*	Djidjelli *əddil ~ əddi ~ dyāl*	Marrakech *dyāl* [+] *~ d ~ t*
Nouakchott *dyal*	Goulimine *dyal ~ d*	Anjra *dyāl ~ d*
Casablanca *dyal ~ d*	J-Fez *dyal ~ di*	Dellys *dyal* [+]

Table 45: Genitive exponents from **aððī li-*

Occurring in western North Africa and the northern fringe of the Levantine area are genitive exponents deriving from an original periphrastic construction **aððī li-*. This phrasal source consists of an archaic form of the relative pronoun **aððī* followed by the proclitic dative preposition **li-* 'to, for', thus with a presumed original significance of 'which is for, belonging to'. A reconstruction of the relative in **aððī* rather than **ðī*

(attested elsewhere in Semitic – cf. Rubin 2005) has been preferred here in order to plausibly account for the entire suite of observed, formally similar items via a single etymological source, though it is theoretically possible that the reflexes of two distinct relative elements **aððī* and **ðī* are active in the data. Though extremely rare in contemporary Arabic dialects, a few examples of relative pronouns reconstructable to **aððī* are attested down to the modern era, including several dialects recorded above as displaying genitive exponents of an apparent **aððī li-* origin, including Djidjelli *əddi*, J-Fez *di*, Cyprus *ta* among others (for comprehensive discussion of Arabic relatives, see Retsö 2004; Pat-El 2017). Eksell Harning (1984) reports that such forms of the relative pronoun have greater currency in the (admittedly sparse) documentation of the medieval and early modern Arabic as spoken in western North Africa and Andalus. It is debated whether such forms are directly relatable to Classical Arabic *allaðī* (for discussion, see Stokes forthcoming), though the verdict is not of strict relevance to the present investigation. Dative **li-* may be reconstructed as a shared inheritance of all Arabic varieties, modern and historical, as described in detail in §4.3.1.3 below.

All forms listed above, save Anatolia *ðīla* and variants, have undergone regular despirantization of */ð/ > /d/, and Cyprus *tel* ~ *te* subsequently shows the devoicing of original voiced stops common in the dialect (see Tsiapera 1969). All outside of Tlemcen and Djidjelli also display loss of the initial vowel (the /ə/ of those dialects representing the regular reflex of */a/) and degemination of the original */ðð/ sequence. As the source of the medial vowel /ā ~ a/ appearing between the **aððī* and *li-* elements in most North African forms (e.g. Azru *dyāl*), Eksell Harning proposes a historical Moroccan variant *yal-* of the preposition *li-*, used in combination with enclitic personal pronouns. Though attestation of this form is scanty, primarily known from a 16th century letter published by Colin (1945), it is not without cross-dialectal precedent: the prothesis of an initial CV

syllable is a common morphophonological augment for **li-* when hosting the clitic pronoun series, as seen for example in Beirut *ʔəl-*.[6] The unmotivated shift of */ī/ to /e/ in Cyprus *tel ~ te* may be the result of a root reanalysis identifying an underlying sequence /ay/; given the conservation of surface /ay/ in Anatolia, however, the /ē/ Anatolia forms *ðēla ~ ðēl* remains unaccounted for. The optional final /a/ of Anatolia *ðīla ~ ðēla* may represent a vestige of the Levantine variant *la-* of the preposition *li-*, the previously proclitic status of which may explain the lack of regular word final raising */a/ > /e/.

In considering the absence of final /l/ in the "short' forms presented above, it is important to note an important distributional characteristic of these variants. The "long" forms with final /l/, for example Tetouan *dyal*, Djidjelli *əddil*, Cyprus *tel*, are primarily utilized when followed by the clitic personal pronouns, while corresponding short forms such as Tetouan *d*, Djidjelli *əddi* and Cyprus *te* occur before nominals. In Tetouan, Djidjelli, Cyprus, Tlemcen, J-Fez, this division is reported to be hard and fast, while in Casablanca, Goulimine, Marrakech and Anjra it is described as more of a distributional tendency with the use of long forms possible if not frequent with nominal complements (though the reverse is not true). No short forms are reported at all for Algiers, Anatolia, Nouakchott, Cherchell and Dellys, though Dellys *dyal* is restricted to pre-pronominal contexts. Eksell Harning attributes this distribution to a historical ellipsis of the *li-* element when preceding nominals as opposed to the clitic pronouns. Though she does not elaborate on the cause of this deletion, I propose as a potential contributing factor a dissimilative process away from the /l/ of a following definite article, which would head

[6] Rather than deriving such resyllabified forms from an allative **ʔilā* 'to' with unmotivated loss of the final long vowel – as alluded to by Cowell (1964) *inter alia* – I prefer to link them to other, broadly observed processes of morphophonological augmentation of **li-* when occurring with attached clitic pronouns; in addition to the described prothesis, these include vowel lengthing, as Cairo *lī-* (Woidich 2006), and consonant reduplication, as in modern Moroccan *lil-* (Heath 2002). Similar change is attested for **bi-* 'with, by', the other common proclitic preposition cross-dialectally, in the same environment.

the great majority of nominals following the exponents in question (Eksell Harning 1984; Brustad 2000).

At this point, my analysis diverges from that of Eksell Harning (1984) in one critical respect. Following the contraction of the short forms prenominally to a reduced state newly resembling the original relative pronoun, Eksell Harning insists on a reidentification and reintegration of these forms with the relative pronoun itself, thereby resulting in a polyfunctionality of the single item as a marker of both relative and genitive qualifiers. This development is proposed to have preceded the encroachment of the more modern relative pronoun form **allī* into the North African area. When **allī*, which is used to mark canonical relative qualifiers only, was adopted into the dialects which display its reflexes today, Eksell Harning claims that genitive uses of earlier **aððī* were effectively stranded and that reflexes of **aððī* surviving in this function were subsequently reinterpreted again as stand-alone genitive exponents.

I differ with aspects of this account on both methodological and theoretical grounds. Instead of positing the emergence of distinct short form genitive exponents twice, I prefer to advance the date of the development of the /l/-less forms to follow the adoption of **allī* as a relative pronoun in those dialects which have it. If functional extension of relative **aððī* to subsume the abbreviated genitive forms does then occur in the remaining dialects, this development takes place in these varieties and these varieties only. Such a split may be reflected in the previously unremarked distributional properties of the short vs. long forms as described above, in that those varieties which display a firm division between short forms before nominals and long forms before pronouns are precisely those which preserve an active **aððī* relative (Tetouan, Djidjelli, Cyprus, Tlemcen, J-Fez), while those using the more common **allī*-based relativizer show a "softer," frequency based split. The behavior of the short forms in the former case may

constitute evidence of their fundamental realignment, having divorced from the long genitive forms in mental representation and come instead to be reassociated with the relative pronoun in the manner described by Eksell Harning. In the latter case, though, the distributional aspects speak instead to the status of the short genitive forms as abbreviated contextual allomorphs of their full length counterparts, the contracting tendencies of which are just that: tendencies.

An additional problem with the narrative of universal reabsorption of the short forms into the relative pronoun followed by secondary split in those **allī*-adopting dialects considered above is the fact that in none of these dialects does **allī* take on the role of genitive marker. From the perspective of language/dialect contact theory, were the short form genitive exponents fully integrated into a polysemous **aððī* relative pronoun it would be striking that in none of the instances of adopted **allī* was the full functional range of the former imposed on the latter (cf. Coetsem 2000). Though not definitive evidence in either direction, this state of affairs is problematic for Eksell Harning's argumentation, a fact she herself mentions as an open question regarding her account (1984: 29). Finally, the status of the short form genitive as distinct from the relative pronoun is supported phonologically at least in the case of Cyprus *te*, which clearly maintains the vowel quality of uneroded *tel* in contrast to the relative pronoun *ta*. No such conclusive formal evidence is available regarding the North African reflexes, where the reduction of the short forms has generally progressed to the near maximal extent represented by Casablanca *d*; in Marrakech *d ~ t*, the form would appear to have undergone an additional sporadic devoicing.

Strikingly given the periphrastic origin just explicated at length, two of the **aððī li-* forms presented above have developed morphological agreement properties. Marrakech *dyāl* variably inflects for gender and number with forms *dyālt* (FSG) and

dyāwəl (CPL), and Dellys *dyal* agrees in number with plural *dyawəl*. Marrakech *dyālt* is formed by a straightforward application of the feminine singular marker *-t*, while the broken plural pattern utilized in both cases mirrors that used with reflexes of **matāʕ* in the same dialects: Marrakech *ntāwəʕ*, Dellys *ntawəʕ*.

It does not pass unnoticed that the route of development sketched here is reminiscent of other relative-derived genitive constructions encountered across the broader Semitic language family. This point is taken up below in §4.3.1.3 for full discussion.

**allī li-*

Anatolia *lēl ~ lē*	Aswan *līl*	Shukrīyah *allīl* [+]
ʻAbābdah *allīl*		

Table 46: Genitive exponents from **allī li-*

Genitive exponents deriving from **allī li-* occur in the upper Nile Valley and southeastern Anatolia. Parallel to that of the **aððī li-* forms just described, the origin of these items is a periphrastic construction consisting of a relative pronoun **allī* followed by the dative preposition **li-* 'to, for', with a compositional meaning of 'which is for, belonging to'. Relatives from **allī* characterize the great majority of modern Arabic varieties as against Classical Arabic and serve as the sole relative marker in all but peripheral regions of North Africa, the Sudan, Yemen, and the northern Syro-Mesopotamia area, generally occurring as more or less frequent variants in those zones as well (Retsö 2004; Versteegh 2014). I reconstruct the initial vowel of this element to /a/ on the basis of the specific forms presented in Table 46, while recognizing that given a

broader scope of inquiry other options may prove preferable (for discussion, see Stokes forthcoming).

The Shukrīyah and ʻAbābdah forms do not require phonological interpretation. The forms from Aswan and Anatolia seem to have undergone an irregular deletion of initial */al/. In the case of Anatolia, this appears to have occurred with the original **allī* component prior to the genitive exponent stage, as evidenced by the variant realization of the relative pronoun itself as *lē*. This would also indicate that the unexplained change in vowel quality also preceded the development of the genitive exponent, though the conditions underlying this change, like that involved in coterritorial Anatolia *ðēla ~ ðēl < *aððī li-*, remain to be understood (though whatever the ultimate source the possibility of analogical influence between the **allī (li-)* and **aððī li-* forms looms large). In light of the presence of this long vowel, whatever its quality, I break from Eksell Harning (1984) in deriving the Anatolian forms from an original **allī* rather than the variant relative pronoun form *il-* encountered in neighboring Mesopotamia (Eksell Harning's primary motivation for selecting the latter seems to be her conviction that the Anatolia *lēl ~ lē* forms need necessarily emerge contemporaneously with *ðīl*, etc., and thus predate the arrival of **allī* to the area). In Upper Egypt, the irregular loss of the initial syllable of Aswan *līl* seems to be specific to the genitive exponent, as the relative pronoun is realized as the expected *illī*. Loss of final /l/ in the shorter Anatolian variant is ascribable to the same processes described at length for the "short" **aððī li-* forms discussed above, perhaps driven by dissimilation from a following definite article *əl-*.

Shukrīyah *allīl* is notable for having developed morphosyntactic agreement properties following lexicalization of the periphrasis, indicating feminine gender of the modified noun the addition of the regular feminine singular suffix *-at*: *allīlat* (F). This is

comparable to the emergence of agreement phenomena observed for Marrakech *dyāl* and Dellys *dyal*, though in this case restricted to the category of gender.

**hī li-*

B-Kadugli *hīl* [+]	Rubāṭāb *hīl*	Shukrīyah *hīl* [+]
Nigeria *hīl*		

Table 47: Genitive exponents from **hī li-*

In the greater Sudanese region, several Arabic varieties present genitive exponents deriving from **hī li-*. This periphrastic source construction consists of the third person singular feminine subject pronoun **hī* and the previously encountered dative preposition *li-* 'to, for', with a compositional meaning of 'it (F) is belonging to me, for me'. The glideless form of the pronoun **hī* is characteristic of the greater Sudanic area to which all varieties listed in Table 47 belong. While by no means unique in a broad view of the Arabic-speaking world, they do serve as a regionally distinctive isogloss that sets these dialects off from their immediate neighbors and from historically attested forms of the language: compare Mecca *hiyya*, Cairo *hiyya*, Tripoli *hīya*, Nouakchott *hiyya ~ hīya*, Classical Arabic *hiya*.

I have selected the originally feminine inflected **hī li-* as the most appropriate citation form as this is the etymology directly underlying the forms in all four dialects mentioned above. In the cases of B-Kadugli *hīl* and Shukrīyah *hīl*, these items exist as part of a more elaborated paradigm of etymologically related forms which inflect for number and gender. Nigeria *hīl* represents a feminine singular agreement value as a suppletive member of a paradigm of genitive exponent forms otherwise based on

derivatives of **hana*. Though documentation is scarcely existent, Rubāṭāb *hīl* would seem to have generalized to lose the gender specification of the etymological source construction as the sole example recorded (the only genitive exponent found in the extremely limited Rubāṭāb texts) involves the use of *hīl* to modify a masculine singular referent (*ħagg* 'property'):

2) Rubāṭāb (Hillelson 1935: 76)
   al- ħagg … mā -ho hīl-ī
   DEF-property NEG-OBJ.3MSG *hīl*-GEN.1SG
   'The property isn't mine.'

The agreement inflection presented for Shukrīyah *hīl* fundamentally differs from those examined for other genitive exponents to this point, in that the alternations in form seem clearly to date from a point prior to the development of the source construction as a genitive exponent. The forms in question are *hūl* (MSG), *hīl* (FSG) and (variably obligatory) *hīll* (CPL), which would appear to rather transparently reflect related but distinct etymologies involving the personal subject pronouns of corresponding value. Respectively, these are **hū* (MSG) *li-*, **hī* (FSG) *li-* and **hin* (FPL) *li-*, in the last case involving assimilation of /n/ to following /l/ and the lengthening of /i/ > /ī/, perhaps under influence from the feminine singular form (**hun* (MPL) *li-* could also represent a viable source here, as fluctuation between /i/ ~ /u/ is widely noted for the dialect, like that discussed for *bitāʕ* ~ *butāʕ*). Thus, it may be more proper to discuss a complete set of etymological source constructions rather than a single one secondarily elaborated, as has appeared to be the case up to this point. This characterization obtains also for the singular B-Kadugli forms *hūl* (MSG) and *hīl* (FSG); corresponding *hilēl* (CPL), however,

seems to represent a secondary augmentation, perhaps under influence from a plural demonstrative resembling neighboring Khartoum *dēl* (CPL), Shukrīyah *dēl* (FPL), though no plural demonstratives are included in the limited B-Kadugli material available to enable a direct comparison.

**ʃõr*

Benghazi *ʃõr*		

Table 48: Genitive exponents from **ʃõr*

In Benghazi in eastern Libya, a genitive exponent is used which is derivable from *ʃõr*, utilized in the same dialect as a preposition meaning 'toward'.

As only a single reflex exists, homophonous with the apparent etymological source, no comparative phonological reconstruction is warranted or even strictly possible. Though prepositional *ʃõr* is reported to be restricted to the Libyan area (Benkato 2014), in terms of etymological links it can presumably be related to the cross-dialectally viable root *ʃ-w-r* with meanings surrounding 'to indicate, to point' (cf. Lane 1968). As a genitive exponent, *ʃõr* is reported to be invariant, showing no morphosyntactic agreement phenomena.

**gayy*

Soukhne *gayy* [+]	Khawaytnah *gī* [+]	

Table 49: Genitive exponents from **gayy*

In interior Syria, two reflexes occur of a genitive exponent deriving from proto-form tentatively reconstructed as **gayy*. More so than any other exponent surveyed in the present study, these forms deny straightforward etymologization, with no obvious source, lexical, grammatical or periphrastic, presenting itself for evaluation. Both Eksell Harning's survey and the descriptive dialectological literature specific to the region are mute as to the forms' postulated origin; as put by Peter Behnstedt, preeminent scholar of Syrian dialectology, in his description of Soukhne *gayy*: "Die Etymologie ist dunkel" (Behnstedt 1994: 122). In light of this state of knowledge, I present two possible etymological sources for the **gayy* exponents, though I caution that each must remain highly speculative until further substantiating evidence is brought to light.

The lack of an evident Arabic etymology for **gayy*, combined with marginal status of the phoneme /g/ in the two varieties concerned (entering the inventory primarily through loanwords and isolated dialect borrowings from local Bedouin varieties, in which it reflects inherited */q/), motivates the expansion of the search's scope beyond the boundaries of the Arabic to include the region's historically multilingual setting.[7] While no phonetically similar genitive particles approximating the function of the Arabic exponents present themselves in other local languages, two such items may be identified which align with cross-linguistically common diachronic *sources* of genitive operators (each, in fact, with available precedent in this very sample). The first of these consists of the Kurdish relative marker in its dialectal variant *ga* (Khan 2004), and the second is the Northeastern Neo-Aramaic preposition *qa-* 'to, for', attested at least in Barwar (Khan 2008) and Diyana-Zariwaw (Napiorkowska 2015) nearby in northern Iraq. In adding to the tentativeness of this account, it should be noted that in neither Kurdish nor Neo-

[7] My appreciation to Drs. Eleanor Coghill and Na'ama Pat-El for their willingness to comment on proposals in this regard.

Aramaic have these items developed genitive functions, nor are any non-genitive functions of the Arabic items described which might indicate a given origin.

Kurdish *ga* may initially seem preferable as a source on phonological grounds, though problems begin almost right away in recognizing that contemporary varieties of Kurdish closest to the interior Syrian area attest relative markers not in *ga* but in *go*, *ko*, *ki* and *ku* (Matras et al. 2016). Moreover, as demonstrated by the diachronic accounts of the **aððī li-* and **allī li-*derived genitive exponents above, the syntactic means by which a relative marker would come to play direct host to an Arabic enclitic possessive pronoun with an intervening supporting element such as a preposition are opaque at best, thus rendering this selection less than optimal from multiple perspectives.

Northeastern Neo-Aramaic *qa-* allays at least some of these concerns, and may in the end represent a preferable option if a non-Arabic source of **gayy* is to be considered. Though it requires a phonological transformation of */q/ > /g/ (not without precedent in these dialects), the status of Aramaic as a genetic relative of Arabic boasting a similar typological profile with regard to both prepositions and personal pronouns eases some of the questions with regard to the continuity of source constructions leading to an eventual genitive exponent, and along the way may provide a plausible account of the anomalous final /yy/. Aramaic *qa-*, like equivalent Arabic prepositions, may attach directly to the dato-genitive pronoun series to result in combined forms such as Barwar *qa-diyi* (*qa-*GEN.1SG) 'to me', *qa-diyux* (*qa-*GEN.2MSG) 'to you' (Khan 2008) structurally analogous to Arabic combinations like Soukhne *gayy-i* (*gayy-*GEN.1SG) 'my, mine', *gayy-na* (*gayy-*1PL) 'our, ours'. The geminate glide present in Arabic **gayy* but absent from a proposed Aramaic *qa-* original may plausibly represent the result of analogy with native Arabic /a/-final prepositions followed by reanalysis of the prepositional base. Native Arabic prepositions ending in /a/ (<*/ā/) generally display specific morphological

variations involving /yy/ when combined with the first person singular dato-genitive pronoun *-ī* ~ *-i*. If Aramaic *qa-* > *ga-* were borrowed into Arabic, it could easily be adopted into this broader pattern: compare, for example, Soukhne *ʕala* 'on' + *-i* GEN.1SG > *ʕalayyi* 'on me', **ga-* + *-i* > *gayyi* 'my, mine'. The resultant *gayyi* could subsequently have been reanalyzed as a base *gayy* combined with a regular pronoun *-i*, and the reinterpreted form generalized across all contexts. From this point, it would be available to develop as a marker of genitive relationship in a manner comparable to that attested for Benghazi *ʃōr*, though no contemporary traces of a prepositional, pre-genitive function remain to support this hypothesis.

The multiple layers of interpretation required in adopting any of the proposals tendered above leave the preceding discussion extremely tentative and the ultimate etymological source of reconstructed **gayy* far from clear. Whatever their origin, I will conclude by noting that both Soukhne *gayy* and Khawaytnah *gī* display morphological agreement for number and gender via an array of highly productive regular adjectival suffixes suggestive of secondary development: Soukhne *gayy* (MSG), *gayyit* (FSG), *gayyīn* (MPL), *gayyāt* (FPL), Khawaytnah *gī* (MSG), *gīt* (FSG), *gīyīn* (MPL), *giyāt* (FPL).

4.3 Analysis

In light of the attested etymological sources genitive exponents across the modern Arabic dialects, I now turn to the evaluation of the derivations presented above as potential examples of the products of CIG. This evaluation comes in two phases. In the first, I simultaneously assess the dialect data as presenting evidence of grammaticalization processes and as classifiable into multiply attested pathways of development, in line with conditions (i) and (ii) of the heuristic described in §1.4. Following this, I proceed to

consider the geographic distribution of the forms described in order to determine the suitability of areal dialect contact as a historical account for their modern incidence, thereby addressing the heuristic's condition (iii).

4.3.1 Evaluation of Grammaticalization Status and Multiply Attested Pathways

In the following subsection, I evaluate the development of genitive exponents from each etymological source recorded above as displaying evidence of four processes considered diagnostic of grammaticalization, namely desemanticization, extension, decategorialization and erosion; reference is also made here to cross-linguistic precedent in determining the status of these developments as representative recognized processes of grammaticalizing change. At the same time, I work to group these individual innovations into higher-level evolutionary pathways, and identify any which encompass multiple distinct etymological sources in a manner consistent with an account of historical replication.

4.3.1.1 Genitive exponents from 'property, possession' (GEN < PROPERTY)

The largest single source of genitive exponents in the Arabic data consists of historically lexical nouns with a meaning of 'property, possession, belonging'. The derivatives of five distinct etyma are grouped under this heading: **matāʕ*, **tabaʕ*, **ħaqq*, **māl* and **jənā*. Descriptions of grammaticalization phenomena worldwide offer ample attestation of this developmental trajectory as a cross-linguistically viable grammaticalization pathway (Heine & Kuteva 2002). Uniquely among the pathways considered in the present study, Arabic developments in this case take a central position in such cross-linguistic examination and theorization, due to the prominent analyses of Maltese *ta* <

**matāʕ* by Haspelmath (1994) and the genitive particle *ta* (< general Sudanese *bitāʕ*) < **matāʕ* of the Arabic-lexifier creole Ki-Nubi by Heine (1982). While the suitability of the latter case may vary with stances regarding the nature of creole genesis and genetic relation, examples like these have been widely considered alongside others of diverse linguistic origin in establishing GEN < PROPERTY as a frequently recurring grammaticalization path globally.

This status in the literature, however, does not alleviate the need to validate the specific Arabic reflexes attested by this sample as individually representing the products of grammaticalization. To this end, Table 50 summarizes the occurrence of desemanticization, extension, decategorialization and diffusion across the reflexes of the five relevant etyma.

Source	***Desemanticization?***	***Extension?***	***Decategorialization?***	***Erosion?***
**matāʕ*	47/47	47/47	47/47	47/47
**tabaʕ*	9/9	9/9	9/9	0/9
**ħaqq*	15/15	15/15	15/15	0/15
**māl*	12/12	12/12	12/12	0/12
**jənā*	2/2	2/2	2/2	1/2

Table 50: Occurrence of Diagnostic Grammaticalization Processes, GEN < PROPERTY

All reflexes of all five etyma subsumed under the pathway GEN < PROPERTY show evidence of desemanticization. This takes the form of a bleaching of the concrete lexical reference of the nominal source and the corresponding shift to the expression of a more abstract, grammatical relationship of possession or genitivity; thus, the material, referential aspect of the original significance is weakened and the relational, grammatical

component is strengthened and expanded. This process of desemanticization is accompanied and amply demonstrated by corresponding processes of extension to new pragmatic usage contexts which do not involve the equation of the modified noun with an identifiable, materially possessed referent. Such extension is present for all reflexes examined, though the precise means by which it is realized vary. One obvious means by which it is realized involves the application of the genitive exponent to modify a human noun clearly not pragmatically construed as a material possession (though it should be noted that this usage appears inherent to the source form for reflexes of **tabaʕ*, which show others signs of extension as seen below); examples of this include Dhofar *itˁ-tˁulāb ħaqqōt-ū* (DEF-students *ħaqq*.PL-GEN.3MSG) 'his students' (Davey 2016: 184). Similarly, exponents may be used to indicate a possessor not capable of owning or possessing in the concrete sense implied by the original semantics of the source lexeme, as in Mateur *ər-ruxsˁa mtāʕ ət-taksi* (DEF-permit *mtāʕ* DEF-taxi) 'the taxi's permit' (Mion 2014: 70). These two processes of extension would appear to be ubiquitous cross-dialectally, at least on the basis of the varieties surveyed in the current sample. Less frequent, but still attested, are examples of the extension of the genitive exponent to circumstances of inalienable possession or classification, as in Kuwait *ʃaʕar māl sibiʕ* (hair *māl* lion) 'a hair of a lion' (Brustad 2000: 78), or to situations in which the relationship implied by the exponent, while certainly genitive in nature, cannot be validly described as possession at all: Damascus *fanajīn tabaʕ ʔahwe* (cups *tabaʕ* coffee) 'coffee cups' (Lentin 2006: 552).

In addition to these processes of desemanticization and extension, decategorialization is also typical of all reflexes described for the GEN < PROPERTY pathway. This is evidenced across the board by the permissibility of the use of these exponents in attributive position to modify a noun specified by the definite article, as in

Jerusalem *ir-rādyo tabaʕ jārt-ī* (DEF-radio *tabaʕ* neighbor-GEN.1SG) ‘my neighbor’s radio’ (Rosenhouse 2006: 489), Azru *l-maʃya nta3-om* (DEF-livestock *nta3*-GEN.3MSG) ‘their livestock’ (Singer 1980: 282). In the vast majority of Arabic dialects, including all those considered here, this is not a syntactic position that can be occupied by nouns; as such, the genitive exponents used in this way demonstrate the gradual loss of their nominal status. Moreover, those exponents that have developed dependent morphosyntactic agreement properties (which include reflexes of each of the five etyma representing this developmental path, as described in §4.2) further display their departure from their original nominal status, as such phenomena are not attested for nouns in any Arabic variety examined; examples include Fezzan *jənt* (FSG) as in *rīʃa jən-t əd-djāj* (feather *jən*-FSG DEF-chicken) ‘a chicken feather’ (Caubet 2004: 88), Khartoum *ħaggīn* (MPL) as in *nās ħagg-īn kalām* (people *ħagg*-PL talk) ‘talkative people’ (Dickins 2006: 570), Basra *mālāt* (FPL) as in *il-muwaðʿðʿaf-āt māl-āt-ā* (DEF-employee-FPL *māl*-FPL-GEN.3FSG) ‘her employees’ (Mahdi 1985: 171).

In contrast to desemanticization, extension and decategorialization, which are consistently in evidence across the sample, processes of phonetic erosion are less evenly distributed. Erosion is relatively frequent among the reflexes of **matāʕ*, attested in one form or another in 35 of 47 instances (e.g. Cherchell *ntāʕ*, Beirut *tāʕ*); similarly, an eroded reflex of **jənā* observed in Fezzan *jən*. Eroded forms of the three remaining etyma **tabaʕ*, **ħaqq* and **māl* are noticeably absent. Reasons behind these observed disparities in the frequency of erosion might be made clear through an examination of its phonological details in cases when it does occur. The lexical source **matāʕ*, the reflexes of which show some degree of phonological erosion in all dialects which use the form, differs from the other four considered in that it is more materially substantial, both in a phonological sense and a morphological one. Significantly, **matāʕ* contains four

potentially valid consonantal roots: the /m/, /t/ and /ʕ/ identifiable as the etymological root of the item through internal reconstruction and comparison with lexically related derived forms, but also /ā/ which may be legitimately analyzed as representing underlying /y/ or /w/ (as demonstrated by the common broken plurals of **matāʕ* reflexes in the pattern CC*ā*CəC: Marrakech *ntāwəʕ*). When erosion occurs in the reflexes of **matāʕ*, it invariably affects the first of these available root consonants, */m/, but in all cases save those of Tozeur *t ~ aʕ* stops short of affecting a second (recall that the loss of /ʕ/ in Malta *ta* is a product of regular sound change). Thus, it would appear that pressure exists from a morphological point of view to maintain a viable triconsonantal root structure in genitive exponents. This pressure has acted as a brake on the progression of erosion processes beyond a certain point in 45 of the 47 eroded reflexes of **matāʕ* above, and in the cases of **tabaʕ*, **ħaqq* and **māl* has impeded its occurrence altogether in light of the fact that these etyma, so to speak, do not have a root to spare, each consisting of three and only three serviceable root consonants – any substantial erosion of the forms as they stand would challenge this status. That Fezzan *jən* is excepted from this tendency may be due to the presumed existence of a readily available biconsonantal analog in phonetically similar **bən* 'son', as discussed in §4.2 above. This consistent pressure toward the conservation of word-level morphological integrity could be hypothesized to relate to these exponents' frequent role as hosts for the clitic possessive pronoun series, their legacy as historical nominals, or both.

As has been seen, reflexes of the five etyma representing the pathway GEN < PROPERTY display ample evidence of the component grammaticalization processes of desemanticization, extension and decategorialization, and erosion has been seen to be prevalent when not disallowed by the presence of transparent, consistent morphological constraints. Therefore, it is secure to conclude that the developments sketched here

represent theoretically supported and cross-linguistically validated examples of grammaticalization and GEN < PROPERTY a verified grammaticalization pathway. As the latter is multiply attested by the reflexes of five distinct but synonymous etymological sources, conditions (i) and (ii) of the study's heuristic are both met and the products of GEN < PROPERTY will continue to be considered as potential examples of CIG in the geographic analysis to follow.

4.3.1.2 Genitive Exponents from 'thing' (GEN < THING)

The sample of Arabic genitive exponents comprises three distinct etymologies representing a developmental pathway which can be summarized as GEN < THING. This evolutionary trajectory is commonly attested among languages of the world, figuring prominently in the survey of grammaticalized possessive constructions presented in Heine and Kuteva (2002). The three etyma included in this grouping are **hana*, **ʃayyit* and *ħājit*, all with a lexical source meaning 'thing'. The reflexes of these etyma are evaluated here for evidence of desemanticization, extension, decategorialization and erosion, with results as displayed in Table 51 below.

Source	***Desemanticization?***	***Extension?***	***Decategorialization?***	***Erosion?***
**hana*	8/8	8/8	8/8	0/8
**ʃayyit*	4/4	4/4	3/4	0/4
**ʃuɣl*	3/3	3/3	2/3	0/3
**ħājit*	1/1	1/1	0/1	0/1

Table 51: Occurrence of Diagnostic Grammaticalization Processes, GEN < THING

Desemanticization is attested across all reflexes of all etyma, as the original lexical significance of the items weakens to the point that they, for all intents and purposes, become vehicles for the semantics of a following possessive pronoun or noun in what would historically have been an adpositional genitive construct. This desemanticization is accompanied by extension of these forms' use to novel semantic and pragmatic contexts. Some recorded usages remain entirely consist with the contexts amenable to the source construction: Damascus *lā tā-xod ha-l-əɣrād^{ʕ}, ʃyāt-i hadōl* (NEG 2-take.IPFV DEM-DEF-things, *ʃīt*.PL-GEN.1SG DEM.PL) 'don't take these things [*əɣrād^{ʕ}*], they're mine' (Cowell 1964: 490). Others, though, extend to include those involving animate possessions, human associations and inalienable derivative relationships in which the modified nouns are clearly incompatible with the original semantics of 'thing, object': Cyprus *pagra ʃayt-i* (cow *ʃayt*-GEN.1SG) 'my cow' (Tsiapera 1969: 65), Sinai *il-iwlād ʃuɣl-īn il-madrasih* (DEF-boys *ʃuɣl*-PL DEF-school) 'the boys of the school' (de Jong 2011: 156), B-Kadugli *al-elmi hān ar-ruwāba* (DEF-water *hān* DEF-curdled.milk) 'the water of the curdled milk' (Manfredi 2013: 36). While the attestations of **ħājit* are extremely limited, they do include one case which seems to show extension in that it would be highly pragmatically marked to refer to the modified noun *il-ʕarabiyya* 'the car' as a 'thing' in such a context: Aswan *il-ʕarabiyya ħājt-ī* (DEF-car *ħājt*-GEN.1SG) 'the car is mine' (Alrawy p.c.). Abstract associative functions are also possible, as in Negev *mā-hum ʃuɣl-īn ħis^{ʕ}īdih* (NEG-OBJ.3PL *ʃuɣl*-PL reaping) 'they are not of the reaping type' (lit. 'they are not of reaping') (Henkin 2006: 365). In Abéché and Nigeria of the West Sudanic area, particularly, the genitive function of these items has expanded to cover very nearly the entire range of genitive relations expressed by language, including even such highly grammatical functions as the indication of argument structure for deverbal

nouns: Abéché *nuzūl hana al-matar* (descent *hana* DEF-rain) ‘the falling of the rain’, *ʃoxol hana akil* (thing *hana* eating) ‘something to eat’ (Roth 1979: 144, 191).

Reflexes of the pathway GEN < THING also show ample evidence of decategorialization. This is demonstrated through their employment in attributive position with a definite antecedent, not generally allowable of nominals in Arabic but encountered ubiquitously in structures like Nigeria *al-bēt adˁ-dˁayīl hana ar-rājil* (DEF-house DEF-tall *hana* DEF-man) ‘the tall house of the man’ (Owens 1993: 65). An important exception here would seem to be the lone reflex of **ħājit*, which was explicitly rejected by the informing consultant when elicited in attributive position and only allowed as a predicate, a distributional detail which sets it apart from other reflexes attested here: Aswan *il-ʕarabiyya ħājt-ī* (DEF-car *ħājt*-GEN.1SG) ‘the car is mine’, never ‘my car’ (Alrawy p.c.). Ten of the sixteen reflexes considered here additionally display decategorialization through the development of dependent morphosyntactic agreement patterning, also not characteristic of Arabic nouns. Examples include Soukhne *hnayyt* (FSG) in *ha-s-siyyāra hā hnayy-t-i* (DEM-DEF-car DEM.FSG *hnayy*-FSG-GEN.1SG) ‘this car belongs to me’ (Behnstedt 1994: 177), Cyprus *ʃat* (CPL) in *gbaz ʃat-ak* (breads *ʃayt*.PL-GEN.2MSG) ‘your breads’ (Tsiapera 1969: 65), and Sinai *ʃuɣlāt* (FPL) in *iθ-θalāθah jinēh-āt ðillih ʃuɣl-āt-uk* (DEF-three pound-PL dem.PL *ʃuɣl*-PL-GEN.2MSG) ‘these three pounds [currency] are yours’ (de Jong 2011: 156).

Given the broad incidence of desemanticization, extension and decategorialization among the representatives of the GEN < THING pathway, evidence of the fourth component process grammaticalization, phonetic erosion, is conspicuously absent. This is reminiscent of the state of affairs encountered for those members of the GEN < PROPERTY path which consisted of three or fewer valid lexical root consonants, where it was argued that pressure to maintain a viable triradical base has served as a brake on the

erosion of phonological material beyond a certain point. An analogous interpretation is perhaps possible here, attributing the strong resistance to phonetic erosion across the four etymological realizations of the developmental pathway GEN < THING to the fact that none of the individual etyma involved contain more than three root consonants and thus face the same pressure to maintain morphological integrity as observed in the previous case. In fact, this pressure toward a triliteral structure is so strong that two representatives of the GEN < THING pathway, **fayyit* and **hana,* are actually seen to undergo phonological augmentation to supplement a historically biliteral root through strategies such as vowel lengthening (Cameroon *hanā*), glide insertion (Soukhne *hnayyi*) and reduplication (B'ērāt *ihnīn*). While such changes are certainly irregular in nature, they are not consistent with the automization and reduction associated with erosion in the grammaticalization context and therefore not considered such here.

To summarize, the Arabic reflexes of the path GEN < THING show ample evidence of the component grammaticalization processes of desemanticization, extension, and decategorialization. Phonetic erosion is lacking, likely due to structural pressures, but this fact is not inherently problematic for the identification of grammaticalization processes at work in the Arabic data, as erosion is typically viewed as both a sequentially and a logically secondary effect of the progress of grammaticalization. The clear occurrence of the first three processes combines the existence of plentiful cross-linguistic precedent to render an unambiguous verdict in the identification of GEN < THING as a verified grammaticalization pathway. Having so met condition (i) of the study's heuristic, the GEN < THING forms proceed to meet condition (ii) as well, as the specific developmental path they represent is multiply attested by the reflexes of four distinct source etyma. Details of these items' geographic distribution await consideration below as the next phase in the evaluation of these items as potential products of CIG.

4.3.1.3 Genitive Exponents from Relative and Preposition (GEN < REL + DAT)

The genitive exponents surveyed in this sample attest two distinct etymologies originating in periphrastic constructions which consist of a relative pronoun followed by a dative preposition (a more specific description of the semantics of the relevant preposition is provided below). The particular trajectory of grammatical evolution is, to the best of my research, not commonly attested cross-linguistically; this fact is not surprising given the fairly exacting set of structural preconditions necessary for such a development to occur as it has in Arabic, including 1) the use of externally headed, head initial relative clauses, 2) a zero copula, and 3) prepositionally-marked possession. Though such typological specificity may be seen to rule out widespread cross-linguistic occurrence, this same diachronic pathway is well attested in other, non-Arabic members of the Semitic language family (Rubin 2005), the structural features of which so align to meet the conditions enumerated above. The significance of this fact does not pass unnoticed and will be returned to in the discussion presented in §4.4. More immediately, however, I must first proceed to evaluate of the reflexes of GEN < REL + DAT vis-à-vis the established component processes of grammaticalization, following from the characterizations provided in Table 52 below.

Source	***Desemanticization?***	***Extension?***	***Decategorialization?***	***Erosion?***
**addī li-*	0/31	0/31	31/31	25/31
**allī li-*	0/5	0/5	5/5	2/5

Table 52: Occurrence of Diagnostic Grammaticalization Processes, GEN < REL + DAT

As demonstrated by the figures in Table 52, the reflexes of **addī li-* and **allī li-* differ from all other reflexes considered to this point in that they cannot be convincingly

shown to display effects of desemanticization and extension in comparison to the semantic details and usage contexts of their lexical sources. Each of **addī li-* and **addī li-* certainly represents an innovative lexicalization of the combination of relative pronoun and particle and thus displays the consequent changes to form expected of such a development, but in terms of functionality neither can be said to measurably differ from the use and compositional meaning of the original periphrastic structure from which it arises. To demonstrate this, it is necessary to consider the observed semantic and functional properties of each of the component elements outside the context of the reconstructed source construction, beginning with those of the dative preposition **li-* 'to, for'.

A consistent polyfunctionality of **li-* is widely observed across modern Arabic dialects and is in fact known since the time of the Ancient North Arabian inscriptional record, being well documented in Classical Arabic as well. The facets of this polysemy are many and include dative, directional, benefactive, temporal, purposive and genitive meanings, though it is primarily the last of these which is pertinent to the present discussion. Already among modern Arabic's earliest distinguishable relatives, we find **li-* used as an indicator of genitive relationship, as for example in the frequently occurring Safaitic formula *l-*[personal name] *h-rgm* (*l-*[personal name] DEM-cairn] 'this funerary cairn belongs to [personal name]' (Al-Jallad 2015: 145). In Classical Arabic, *li-* is used to express an array of genitive and possessive relationships. These include alienable, material possession and impermanent personal associations, as in examples (3) and (4), and range also to instances of more abstract and even inalienable possession, as in (5) and (6):

3) Classical Arabic (Wright 1898: 149)
al- māl -u li-zayd-in
DEF-property-NOM *li*-Zayd-GEN
'The property is Zaid's.'

4) *la-hu bi-baɣdād -a sittumiʔati sˁāħib -i xabar-in*
li-GEN.3MSG in-Baghdad-GEN 600 master-GEN news -GEN
'He had in Baghdad six hundred secret police.'

5) *ar- rajul-u man la-hu raʔiy -un sˁāʔib-un*
DEF-man -NOM who *li* -GEN.3MSG opinion-NOM right -NOM
'The man is he who has a right opinion.'

6) *mā l -ī ab -un wa -lā bn -un*
NEG *li*-GEN.1SG father-NOM and-NEG son-NOM
'I have neither father nor son.'

These usages persist into the majority of modern Arabic dialects: Jisr az-Zarqa *ma kān-iʃ il-u masˁārī* (NEG be.PFV-NEG *il*-GEN.3MSG money) 'he had no money' (Belinkov 2014: 57), Damascus *sˁāħəb ħamīm ʔəl-i* (friend intimate *ʔəl*-GEN.1SG) 'a close friend of mine' (Cowell 1964: 480), Abha *mā l-ah daxal fi-yya* (NEG *l*-GEN.3MSG business in-GEN.1SG) 'he has no business with me' (Al-Azraqi 1998: 136), Cairo *ana li-yya ʔumm-i w ixwāt* (SBJ.1SG *li*-GEN.1SG mother-GEN.1SG and sisters) 'I have a mother and sisters' (Woidich 2006: 143). Thus, significant potential exists for overlap of these apparently inherited functions of **li-* with diachronically later innovations like the genitive exponents considered in this chapter, as clearly shown by examples like the

following from Tripoli, in which the possessive relationship entailed is expressed in the first sentence by the genitive exponent *mtāʕ* but paraphrased in the immediate response by a reflex of **li*-:

7) Tripoli (Pereira 2008: 320)
hāda mtāʕ mənu? lē-ya āne!
DEM.MSG *mtāʕ* who? *lē*-GEN.1SG SBJ.1SG!
'Whose is this? It's mine!'

These semantics of **li*- extend to areas in which it has played a role in the formation of genitive exponents. Thus, in the Sudanese and Upper Egyptian zone, home to reflexes of **allī li*-, we find Shukrīyah *bagart-an la-yy* (cow-INDF *la*-GEN.1SG) 'a cow of mine' alongside *al-jamal allīl-ī* (DEF-camel *allīl*-GEN.1SG) 'my camel' (Reichmuth 1983: 188, 113), and similarly in the North Levantine area including reflexes of **allī li*- and **addī li*- we encounter Cyprus *li-ni xops* (*li*-OBJ.1SG bread) 'I have bread' and *ʃ-ʃaya te l-ʕarus* (DEF-things *te* DEF-bride) 'the things of the bride' (Tsiapera 1969: 68; Borg 1985: 165). The primary distinction between pairs like these seems to rest in two syntactic factors, in the first case, the definiteness of the modified noun and, in the second, the predicative vs. attributive nature of the possessive structure. These distinctions are precisely those reflected by the use of the relative particle in each instance. In the dialect of the Shukrīyah, as in the majority of Arabic varieties, relative clauses following an indefinite antecedent are zero-marked; as soon as the antecedent is marked for definiteness, the occurrence of the relative marker **allī* preceding **li*- is necessitated, thereby generating the source context from which later *allīl* arises. In the Cyprus example – where, unusually for Arabic dialects, such a distinction between

indefinite and definite relative heads does not obtain (Borg 1985: 145) – the primary function of the relativizer is to move the possessive relationship described by **li-* from a predicative to an attributive position (recalling that *te* < *tel* < **ta li-*). Thus, in each instance the resulting functions of derivatives of **allī li-* and **addī li-* are very much the sum of their parts, and no desemanticization and extension of the source constructions need be claimed.

The situation of the western North African area, the third and largest zone showing genitive exponents from GEN < REL + DAT, is somewhat more complex. Seemingly unique among major dialect areas of the modern Arabic speaking world, examples of stand-alone **li-* marking genitive or possessive relationship are extremely rare in Morocco and Algeria. The only set of exceptions seem to be occasional constructions denoting the passage of time via a metaphorical, abstract possessive structure, as Djidjelli *li-yya mədda twīla …* (*li*-GEN.1SG period long …) 'It's been a long time since I …' (Marçais 1956: 456), a usage known from other dialects though most often supported by an inchoative copula: Cairo *baʔā-l-ī sāʕa b-a-xabbatˁ ʕa l-bāb* (become.PFV-*li*-GEN.1SG hour CNT-1SG-pound.IPFV on DEF-door) 'I've been pounding on the door for an hour' (Hinds & Badawi 1986: 91). The limited diachronic record of dialectal Arabic in the region strongly suggests, though, that more general possessive and genitive uses of **li-* existed historically. We have, for example, the non-lexicalized sixteenth century Moroccan attestation of *addi yal-i* (REL *yal*-GEN.1SG) 'that which is mine' described in §4.2 above (Colin 1945), in addition to frequent usage in the closely related Arabic varieties of medieval Al-Andalus: *xubz-an liss-u l-ak* (bread-INDF NEG-OBJ.3MSG *l*-GEN.2MSG) 'bread which is not yours' (Corriente 1977: 144). Where, then, are the reflexes of possessive/genitive **li-* in contemporary Morocco and Algeria?

The response may lie in the proposition that modern reflexes of **addī li-* are all that remain of original genitive **li-*, following a set of local North African developments which have conspired to eliminate the majority of genitive and possessive uses of the preposition in syntactic contexts other than those modified by the relative marker. First, far-reaching innovations in the definiteness marking systems of the western North Africa have significantly altered the system of overtly marked relativization for definite heads vs. zero-marked relativization for indefinite heads earlier described. Instead, modification by way of overtly marked relative clauses becomes licensed for referential and specific referents indicated by the indefinite articles **ʃī* and **wāḥəd əl-* (for discussion of these values, see Brustad 2000; Turner 2013; forthcoming):[8]

8) Casablanca (Youssi 1992: 146)
waḥəd l-məsʔala lli i-mkən n-tkəllm -u ʕli -ha
INDF- issue REL 3-be.possible.IPFV 1-talk.IPFV-PL about-GEN.3FSG
'an issue we can discuss'

9) Casablanca (Harrell 1962: 165)
ʃi tˁumumbil lli tə- mʃi məzyan
INDF car REL 3FSG-go.IPFV good
'a car that will run well'

This type of specified indefinite value is precisely that marked by the suffix *-an* in the two examples of zero-marked **li-* attributive relatives given above in Shukrīyah *bagartan*

[8] Use of the overtly marked relative clause with specified indefinite heads is similarly noted in limited cases for at least Syrian, Egyptian and Kuwaiti dialects as well; the unique development of Moroccan lies more precisely in the evolution of innovative definiteness categories which circumvent many of the pragmatic and discourse factors which constrain the use of this structure in other varieties (for discussion, see Brustad 2000).

la-yy ‘a cow of mine’ and Andalus *xubzan liss-u l-ak* ‘bread which is not yours’, and thus in western North Africa it is a completely predicted development to encounter in such instances such as the following, rather than a zero marked structure consisting of the noun followed by **li-* alone:

10) Fez (Caubet 1993: 269)
wāḥed əd-dār dyāl əs- soltˁān < **wāḥed əd-dār addī li-s- soltˁān*
INDF- house *dyāl* DEF-sultan INDF- house REL *li*-DEF-sultan
‘a house of the Sultan’s’

Secondly, the extension of competing preposition **ʕind* to the expression of inalienable possession as opposed to solely alienable – a process which may be seen to be underway in numerous modern dialects – seems to have excluded plain **li-* from its major remaining domain of predicative, new information ‘have’-possession: compare Cairo *ana li-yya ʔummi w ixwāt* ‘I have a mother and sisters’ from above with Tetouan *ʕnd-u bbāh mʃhūr* (*ʕnd*-GEN.3MSG father famous) ‘he has a famous father’ (Brustad 2000: 40). Thus, the sole remaining preserve of possessive/genitive **li-* is that which falls within the syntactic context of the overtly marked relative structure **addī li-*, thereby rendering the lack of observed **li-* in this function not problematic to the development of the GEN < REL + DAT exponents in this region but directly consequent to it.

This understanding allows us to account for the significantly broader functionality of North African **addī li-* exponents compared cross-dialectal counterparts by way of more general, preexisting evolutions affecting the individual inputs to the source construction rather than positing processes of desemanticization or extension affecting

the subsequent evolution of the structure as a unit. The cross-dialectally notable aspects of this expanded functionality are observed by Brustad (2000) to include classification, quantification, and regular use for the expression of inalienable possession. The last has been seen to be consistent with the inherited semantics of **li-*, and is not restricted to North Africa but also obtains for other products of GEN < REL + PRO discussed here, as Marrakech *əd-dəmm dyāl-ek* (DEF-blood *dyāl*-GEN.2SG) 'your blood' (Sánchez 2014: 216), Anatolia *əbən lē ʕamm-i* (son *lē* paternal.uncle-GEN.1SG) 'my uncle's son' (Jastrow 1973: 94) and Berber (neighboring Shukrīyah territory in East Sudan) *ad-dimūʕ allāt wadd an-namīri* (DEF-tears *allī*.FPL Wadd al-Namīrī) 'the tears of Wadd al-Namīrī' (Reichmuth 1983: 113).

The uniquely North African functions of classification and quantification follow directly from syntactic transformations triggered by regionally specific developments to the definiteness system, namely the changes to relative clause marking described above and the bleaching of inherited definite article **al-* to a marker of default, undetermined state (Turner 2013; forthcoming). To understand this, we must turn first to the cross-dialectally attested functions of attributive **li-* as an unmarked alternative to the inherited adpositional synthetic genitive structure involving a head noun in construct state, well known from Classical Arabic (Ryding & Versteegh 2006) and active also in many modern dialects, as Damascus *xārtət tʿəroʔ* (map.CNST roads) ~ *xārta lə-tʿ-tʿəroʔ* (map *lə*-DEF-roads) 'a road map' (Cowell 1964: 460); in cases where the second term of the construct must necessarily be construed as definite but the antecedent is to remain syntactically indefinite, use of the **li-* alternative becomes obligatory, as in Damascus *xārta la-tʿəroʔ ləbnān* (map *lə*-roads Lebanon) 'a road map of Lebanon' (lit. 'a map of the roads of Lebanon'). Classificatory and quantitative relations in Arabic are generally conveyed by means of the synthetic genitive with an indefinite second term (Eksell

Harning 1980): Soukhne *sˁaħn nħās* (dish copper) ‘a cooper dish’, *xams əbyūt* (5 house.PL) ‘five houses’ (Behnstedt 1994: 174, 154). In western North Africa, however, where the unmarked, generic state of the noun includes a (functionally empty) formal marking of definiteness (Turner 2013), this indefinite synthetic construction is disallowed and the alternative periphrasis involving **li-* is forced, which is then subsequently subject to modification by an overtly marked relative clause in the manner described earlier for specified/referential possessed nouns. Thus, classificatory usages like Sefrou *waħəd l-ʔinaʔ dyal tˁ-tˁin* (INDF-pot *dyal* DEF-clay) ‘a clay pot’ (Turner 2013: 118) can be derived from a cross-dialectally typical **ʔinaʔ tˁin* through a series of diachronic developments and synchronic transformations (phonological evolution momentarily set aside for the sake of clarity):

11) a. **ʔinaʔ tˁin*
pot clay

b. **#ʔinaʔ tˁ- tˁin* >
pot DEF-clay

c. **ʔinaʔ li- tˁ- tˁin* >
pot DAT-DEF-clay

d. **waħəd l-ʔinaʔ addī li- tˁ- tˁin* >
INDF- pot REL DAT-DEF-clay

e. *waħəd l-ʔinaʔ dyal tˁ- tˁin*
INDF- pot *dyal* DEF-clay
‘a clay pot’

The same interpretation may be equally applied to cases of quantification historically expressed by indefinite construct (which may be seen to be inherently specified by virtue

of enumeration), relating examples such as Casablanca *xemsa d əl-ktūb* (5 *d* DEF-books) 'five books' (Caubet 2006: 280) to an earlier **xems ktūb* (5 books) in a manner which does not require one to posit desemanticization or extension of the original **addī li-* construction.

In contrast, reflexes of the GEN < REL + PRO pathway do show clear evidence of decategorialization and erosion. All undergo a process of lexicalization by which the morphosyntactic autonomy of the source construction's component elements is lost; this is demonstrated by their phonological and syntactic fusion and resulting inseparability, a consequence of which is that the element reflecting original **li-* is no longer free to occupy a non-initial position of the relative clause (as was the case, for example, in Andalus *xubz-an* [*Ø liss-u la-k* $_{RC}$] 'bread which is not yours' cited above). The widespread phonetic erosion of the forms also serves as strong evidence of the loss of their distinct lexical status, as forms such as Cyprus *te* and Marrakech *d ~ t* convey the semantic functions of the original preposition **li-* despite no longer containing any actual etymological material attributable to that form. Further, Shukrīyah *allīl*, Marrakech *dyāl* and Dellys *dyal* have evolved morphosyntactic agreement properties typical of the adjectival domain, thereby incontrovertibly demonstrating departure from their original categorial status. Viewed as a process in its own right, phonetic erosion is frequent, though more so among products of **addī li-* than those of **allī li-*, affecting 25/31 and 2/5 reflexes respectively.

In sum, the reflexes of etyma representing the path GEN < REL + DAT differ from those of other pathways considered thus far in that they do not display decisive evidence of desemanticization and extension. While results of two other recognized component processes of grammaticalization, decategorialization and erosion, are displayed in significant quantity, the consistency of these forms' semantic details and usage contexts

with those reconstructable for their original source construction complicates their acceptance as examples of grammaticalization on the basis of the analytical methods adopted for this investigation. The lack of desemanticization in particular is perhaps the most noteworthy in this context, as the latter is regarded by many authors (e.g., Haspelmath 1999; Heine 2007) as the cornerstone process on which other component processes of grammaticalization are based, providing a context to decategorialization and erosion which qualifies their interpretation among more general examples of lexicalization/cliticization and irregular sound change which my reflect no broader course of evolution. Without these semantic and pragmatic underpinnings, the morphosyntactic and phonetic changes affecting the products of the GEN < REL + DAT path are not in and of themselves sufficient to definitively clear the heuristic's condition (i) and identify the relevant forms as unambiguous examples of grammaticalization, at least within the confines of the methodology – at times admittedly restrictive (see §1.4) – adopted for the present work. The lack of clear cross-linguistic parallels of the GEN < REL + DAT path worldwide serves as at least partial corroboration for this conclusion: were the developments examined here attributable to universally applicable processes of metaphor extension and automization, one would expect the not infrequent recurrence of this particular evolutionary trajectory outside the context of Arabic and its close genetic relatives. The multiple attestation of this path across Arabic dialects and other Semitic languages does, of course, deserve explanation, and this topic is returned to in §4.4 below once pertinent points of geographic distribution have additionally been addressed.

4.3.1.4 Genitive exponents from Subject Pronoun and Preposition (GEN < PRO + DAT)

The set of forms apparently arising from a structure GEN < PRO + DAT are in many ways developmentally similar to the GEN < REL + DAT derivations just discussed. They consist of an original subject pronoun combined with a dative preposition **li-*, the semantics of which have been described in detail in the preceding section. The subject pronoun in this case may be viewed as originally filling the role of a supporting element in a topic-fronted structure of a zero-copula equational sentence, a structure well documented both within the specific dialects showing reflexes of GEN < PRO + DAT – Rubāṭāb *al-ʕomda hū al kātil al-jidāda* (DEF-mayor SBJ.3MSG REL kill.PTCP DEF-chicken) 'the mayor it is who killed the chicken' (Hillelson 1935: 71) – and more broadly in the Arabic-speaking world; the transformation is commonly syntactically triggered to avoid attributive/predicative ambiguity in a zero-copula context, and need not imply a specific pragmatic value (Eid 1983). Cross-linguistic parallels of this genitive exponent origin are not in evidence, perhaps limited by the highly specific typological profile necessary for the source construction to arise, consisting in this case of 1) the availability of topic-fronted structures for equational sentences with resumptive subject pronoun support, 2) a zero copula, and 3) prepositionally-marked possession. Though a similar state of affairs was proposed to constrain the cross-linguistic occurrence of the pathway GEN < REL + DAT, the pathway GEN < PRO + DAT would seem to be even further isolated by the lack of analogous developments even within the Semitic family.

Processes of desemanticization, extension, decategorialization and erosion among the reflexes of GEN < REL + DAT are discussed below, summarized in Table 53.

Source	*Desemanticization?*	*Extension?*	*Decategorialization?*	*Erosion?*
**hī li-*	0/4	1/4	4/4	1/4

Table 53: Occurrence of Diagnostic Grammaticalization Processes, GEN < PRO + DAT

Desemanticization is absent from the reflexes of **hī li-* for the same reason it is absent from those of **addī li-* and *allī li-* discussed above, namely that their respective semantic values are not observed to differ from those inherent to the original **li-* of the source construction. The significance and functions of the latter have been described in detail in §4.3.1.3 and pertain in this case as well, with the inclusion of the subject pronoun in the source construction simply directing a predicative reading in the presence of a definite antecedent, as seen in the comparison of the previously considered Shukrīyah attributive structure with the following B-Kadugli predicative one:

12) Shukrīyah (Reichmuth 1983: 188)
bagart-an la-yy
cow -INDF *la*-GEN.1SG
'a cow of mine'

13) a. B-Kadugli (reconstruction mine, (b) from Manfredi 2013: 36)
**al- bagara di hī l-i ana >*
DEF-cow DEM.FSG SBJ.3FSG *l*-GEN.1SG SBJ.1SG
b. *al- bagara di hīl -i ana*
DEF-cow DEM.FSG *hīl*-GEN.1SG SBJ.1SG
'This cow is mine.'

Thus, no extension must be posited to account for typologically notable functions of the **hī li-* forms among genitive exponents of other origins, such as their use to express the possession of non-material or human possessa in Nigeria *ādit-tum hīl barnu* (custom-GEN.3MPL *hīl* Bornu) 'their customs, of the Kanuri [Bornu]' (Owens 1993: 64), Shukrīyah *al-butʿāna biga-t hūl-hun* (DEF-Buṭānah become.PFV-3FSG *hīl*.MSG-GEN.3MPL) 'the Buṭānah [tribe] became their property/beholden to them' (Reichmuth 1983: 112). The one probable case of extension observed is that of Nigeria *hīl* to express such abstract genitive relationships as material classification, as in *maraba hīl lēs* (cloth *hīl* lace) 'a lace cloth'. Such usage cannot be ascribed to original **li-* via the same series of definiteness permutations affecting classificatory usages of North African **addī li-* forms, as these did not take place in Nigerian Arabic and the relevant structural criteria are not met. Instead, it is most likely that Nigeria *hīl* was granted these functions via its suppletive incorporation into a paradigm otherwise based on forms deriving from **hana*, the contextual extension of which has been well described and documented in both Nigerian and surrounding dialects, as described in §4.3.1.2.

Decategorialization is universally evidenced by representatives of the GEN < PRO + DAT path, shown most obviously by the lexicalization of originally periphrastic **hī li-* into a single word-level unit with the corresponding loss of morphosyntactic autonomy of its etymological constituent parts. In a departure from the developments observed for GEN < PROPERTY and GEN < THING, the reflexes of GEN < PRO + DAT remain consistent with their originally predicative nature and do not occur in attributive position. Any such apparent usages are more properly analyzed as being incorporated into a zero-marked relative clause modifying an indefinite noun (see §4.3.1.3), with the relativizer overtly marked as the noun is made definite: compare zero-marked, indefinite Shukrīyah *ʕabb-an* [*Ø hūl-u* $_{RC}$] (slave-INDF [*Ø hīl*.MSG-GEN.3MSG $_{RC}$]) 'one of his slaves' to

definite *al-bēt* [*al hūn-na* $_{RC}$] (DEF-house [REL *hīl*.MSG-GEN.1PL $_{RC}$]) ‘our house’ introduced by relative marker *al* (Reichmuth 1983: 112). However, loss of categorial status of original **hī* as a gender/number and (nominative) case marked pronominal is still apparent in uses where **hī li-* occurs as an accusative predicate and where expected gender/number agreement is controverted, both of which may be observed in Shukrīyah *al-ʕabīd big-u hūl-i* (DEF-slaves become.PFV-PL *hīl*.MSG-GEN.1SG) ‘the slaves became my property’ where *hūl* is an original masculine singular form and the entire genitive phrase serves as the second argument of accusative-assigning *bigu* ‘became’ (Reichmuth 1983: 112). Contrary to most exponents examined here, the presence of agreement patterning in reflexes of **hī li-* is generally traceable to the properties of the source form and thus not to be taken as evidence of decategorialization, save in the case of B-Kadugli *hilēl* (CPL), which appears on formal grounds to be a secondary development.

Reflexes of **hī li-* do not generally evidence phonetic erosion, the sole exception a phonologized assimilation of */n/ > /l/ in Shukrīyah *hīll* (CPL) < **hin li-*, applicable on a phonetic level across the dialect (cf. *bagartan la-yy* [bagartal_laj:] ‘a cow of mine’ (Reichmuth 1983: 188)). The loss of final /i/ across all forms is perhaps best attributed to exaptation of the eventual lexicalized form from a pronominal context in which clitic pronouns attach to a vowelless base **l-*, an inherited pattern observed across the dialect sample. This contrasts with the cross-dialectal treatment of reflexes of **addī li-* and **allī li-*, which represent structurally similar periphrastic origins but are regularly subject to phonetic degradation (compare J-Fez *di*, Anatolia *lē*). It bears noting, however, that for the **allī li-* forms overlapping with products of **hī li-* among the Shukrīyah no such erosion is described.

Despite the widespread occurrence of decategorialization, little to no evidence of desemanticization, extension or erosion exists to support an identification of the reflexes

of **hī li-* as products of grammaticalization or of GEN < PRO + DAT as a verifiable grammaticalization path; moreover, this state of affairs is mirrored by the lack of identifiable parallel developments cross-linguistically, which would be expected were the changes observed attributable to universally applicable cognitive processes and principles. The developmental trajectory GEN < PRO + DAT thus fails to meet condition (i) of the study's heuristic as well as condition (ii), the requirement for realization involving multiple distinct etymological sources. In this light, the products of the GEN < PRO + DAT pathway do not appear strong candidates for the results of CIG.

4.3.1.5 Genitive exponents from Allative Prepositions (GEN < ALL)

The Arabic sample attests a single genitive exponent deriving from an allative preposition **ʃõr*, with a concrete spatial significance of 'toward'. Though the preposition in both its allative and genitive senses would appear to be a local Libyan innovation (Benkato 2014), the lexical root *ʃ-w-r* 'point, indicate' from which it ultimately derives is a well-known feature of the common Arabic lexicon. In terms of cross-linguistic corollaries, the evolution directly from allative to genitive or possessive function would not seem to be well documented among the world's languages as a direct shift, as most such developments appear to involve an intermediate dative function linking these two meanings. Heine and Kuteva (2002), for example, cite numerous attestations of allative to dative grammaticalization and many further cases of datives evolving a possessive or genitive function, but no changes directly from allative to genitive meaning; just such a two stage development is present in the evolution of pan-Arabic **li-* discussed in §4.3.1.3 above, which may be seen to have expanded its functionality from the expression of directional to dative and finally to possessive relation. Arabic **ʃõr*, however, does not

show any evidence of such a phased progression, at least at the present state of documentation.

The occurrence of desemanticization, extension, decategorialization and erosion in the sole representative of GEN < ALL is provided in Table 54, below.

Source	***Desemanticization?***	***Extension?***	***Decategorialization?***	***Erosion?***
**ʃōr*	1/1	1/1	0/1	0/1

Table 54: Occurrence of Diagnostic Grammaticalization Processes, GEN < ALL

The lone reflex considered here, Benghazi *ʃōr*, shows evidence of desemanticization in its departure from the spatial, directional semantics of its allative source 'toward' to include notions of possession and even more abstract associative relation. This is accompanied by extension of the form to novel pragmatic usage contexts. Unfortunately, contextualized examples are extremely limited in the data provided by Benkato (2014), but what little is available is sufficient to prompt such a claim. Beyond demonstrating the original allative usage of the source form, Benghazi *ʃōr el-blād* (*ʃōr* DEF-downtown) 'towards downtown' (Benkato 2014: 89), the author simply notes that the item is also employed as a morphologically invariable genitive exponent and cites a form *ʃōr-ī* (*ʃōr*-GEN.1SG) 'mine'. The one example provided elsewhere in the text would seem to imply a high degree of semantic abstraction and pragmatic extension involving no traces of the physical directionality characteristic of the source: Benghazi *it-ʿtʿīħ miʃ ʃōr-ī* (2-fall.IPFV NEG *ʃōr*-GEN.1SG) 'if you fall, it's not my problem' (lit. 'it's not of/pertaining to me') (Benkato 2014: 89).

No evidence of decategorialization exists for Benghazi */ōr*, which is explicitly described as inflectionally invariant and seems from what data is available to behave syntactically in a manner consistent with the distributional properties of its prepositional source. Neither is any erosion attested, the form of the genitive marker being homophonous with that of the original allative preposition.

In spite of the lack of decategorialization and erosion, the clear occurrence of desemanticization and extension described here are sufficient to support a tentative identification of the Benghazi genitive exponent ***/ōr* as a product of grammaticalization and of GEN < ALL as a legitimate grammaticalization path. The lack of precise parallels among world languages may be viewed as problematic to such a conclusion; however, given the close semantic relationships demonstrated cross-linguistically between allative and dative functions on the one hand and dative and genitive functions on the other (cf. Heine & Kuteva 2002) the semantic and functional gap involved does not seem an insurmountable one, and the directionality of change would remain appropriate to the notion of progression along a universal grammaticalization cline. The lack of documentary description for Benghazi */ōr* is also relevant here, as even a limited amount of additional evidence relating to the item's broader functionality (for example, the existence of any degree of dative usage) could serve to dramatically qualify this stance. Regardless of the final verdict, the success of the GEN < ALL in meeting our heuristic's condition (i) is met in turn by its failure in satisfying condition (ii), as the path is represented by only a single etymology in the Arabic data and thus provides no opportunity for replication via CIG.

It is worth noting that, were a source for the Arabic ***gayy* exponents to be definitively identified in the Northeastern Neo-Aramaic preposition *qa* 'to, for', then its path of development would be in many ways comparable to the that described for ***/ōr*; in

one respect, it would actually be sufficiently more in line with cross-linguistic precedent, as the evolution would involve a clear dative stage – Diyana-Zariwaw NENA *yuwəl-lan qa-nafe* (give.PFV-1PL *qa*-people) 'we gave to people' (Napiorkowska 2015: 358) – bridging allative and genitive functions. However, the etymological identification involved is far from clear, and even if it were, it is not evident whether such development should be interpreted as having taken place prior or subsequent to a purported borrowing into Arabic. Given this lack of clarity on multiple fronts, the reflexes of **gayy* will not be considered in the continued evaluation of the role of CIG in the development of Arabic genitive exponents, as rigorous application of the theoretical principals underlying the study's heuristic is not possible at the present state of knowledge.

4.3.2 Evaluation of Geographic Distribution

The preceding analysis has succeeded in identifying two verified grammaticalization pathways for genitive exponents attested by multiple distinct etymological realizations in the Arabic data: GEN < PROPERTY and GEN < THING. These are represented by five and four realizations each, comprising eighty-five and sixteen individual reflexes, respectively. In this section, the geographic distribution of these forms is evaluated for consistency with a history of propagation via areal diffusion, as per the study heuristic's condition (iii).

In addition, a third multiply attested pathway was uncovered above, that of GEN < REL + DAT, though its products failed to meet condition (i) of the study's heuristic as defensible examples of grammaticalization due to the absence from their development of the essential component processes of desemanticization and extension. In the geographic evaluation provided here, the areal distribution of the thirty-six reflexes representing this

pathway will also be considered in the interest of shedding additional light on the nature of the qualitative differences already noted between this path on the one hand and GEN < PROPERTY and GEN < THING on the other.

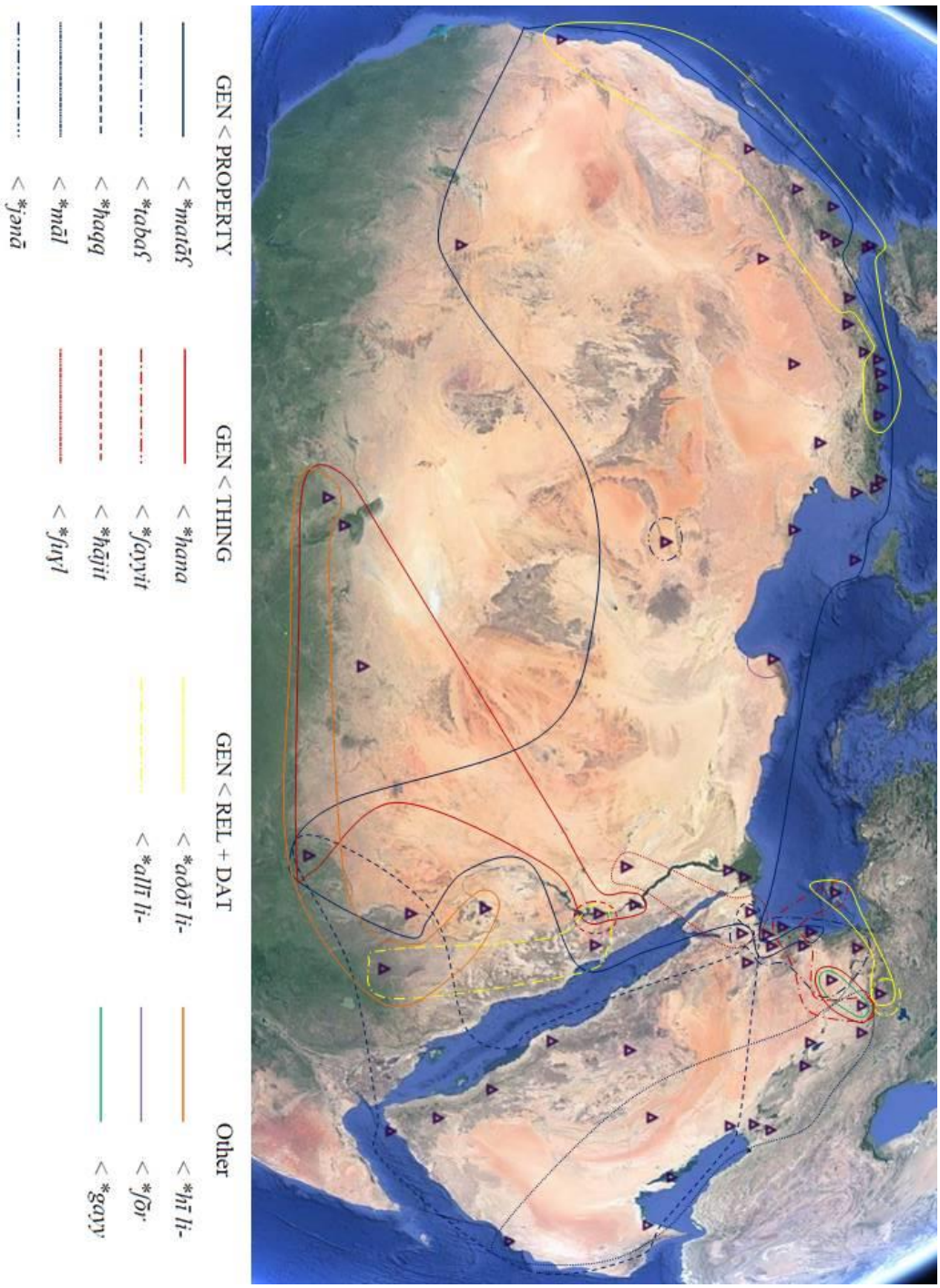

Figure 3: Geographic Distribution of Genitive Exponents (Map data: Google, S.O., NOAA, U. S. Navy, NGA, GEBCO; Image: Landsat/Copernicus)

Figure 2 demonstrates that the reflexes representing each of the two grammaticalization pathways GEN < PROPERTY and GEN < THING show a high degree of geographic cohesion. Of the etyma representing the GEN < PROPERTY path, reflexes of **matāʕ* show by far the widest distribution. We encounter them in the far west of the Arabic-speaking world in modern Morocco and the Saharan region of Mauritania and Mali, and subsequently in an uninterrupted band comprising the remainder of North Africa, the Nile Valley as far south as the central Sudan, and southern and coastal portions of the Levant. Adjoining and overlapping with these forms in the broader Levantine region are reflexes of another representative of the same pathway, **tabaʕ*. Abutting **tabaʕ*-derived forms at the intersection of the Levantine and Mesopotamian areas are those deriving from **māl*, which continue southward and westward to cover the entirety of Iraq and the Persian Gulf coast. Reflexes of **ħaqq* occur alongside those of **māl* throughout the latter zone and dominate throughout the remainder of the Arabian Peninsula, further bridging the Red Sea to occur on Sudanese soil as well. They share territory with reflexes of **matāʕ* in this area and additionally border those of **tabaʕ* in Northwest Arabia, thereby further reinforcing the geographic contiguity of the entire GEN < PROPERTY complex. We find the fifth and final realization of this pathway in interior Libya, where the reflexes of **jənā* are co-territorial with those of semantically and functionally parallel **matāʕ*.

Genitive exponents deriving from the grammaticalization pathway GEN < THING occupy a significantly smaller but not inconsiderable geographic area. Reflexes of **hana* are identified in two distinct pockets, the first and largest of which binds virtually the entire West Sudanic area to the southern portion of Upper Egypt, a linkage – though distant – borne out by a number of additional dialectological isoglosses as well as received tribal history and traditional genealogy (Owens 2003). The second, significantly

more restricted, comprises a zone of the eastern Syrian interior. Joining the first set of **hana* forms in Upper Egypt are two additional etyma representing the GEN < THING path, **ħāja* in Aswan and **ʃuɣl* at the Kharga oasis. Forms deriving from **ʃuɣl* similarly occur throughout the Sinai and the Negev; though a gap in attestation concerning the intervening area is observable on the map, that the Kharga reflex is directly linked to these more northerly cognates is supported by Behnstedt and Woidich's (2012: 168) observation that **ʃuɣl*-based genitive exponents are known throughout the Middle Egyptian zone. The incidences of these forms somewhat tenuously connect to the southernmost extent of **ʃayyit*, a connection perhaps strengthened by Eksell Harning's (1980) finding that **ʃayyit*-derived forms were until recently more robustly attested across the Palestinian area. The reflexes of **ʃayyit* in turn overlap with the northern bloc of **hana* forms in central Syria, expanding also to encompass nearby Cyprus.

The constituent realizations of the GEN < THING grammaticalization path thus form a contiguous band spanning from the West Sudanic zone at one extreme to the northern Levant in the other. Given the sometimes thin attestation and fine scope required, this determination is perhaps not so unassailable as that made for the realizations of the larger GEN < PROPERTY pathway above. However, even if one were to take the most conservative approach possible to areal grouping the results would appear divided into two regional blocs, each of which would still meet the requirement for the contiguous geographic incidence of synonymous source etyma with at least **hana* and **ʃayyit* placed together in the north and **hana, *ʃuɣl* and **ħājit* together in the south; our heuristic's condition (iii) is thus met for forms of this path under either interpretation.

As earlier promised, we now turn to the products of the developmental trajectory GEN < REL + DAT. Despite their failure to meet the heuristic's condition (i) as valid examples of grammaticalization, their geographic incidence still deserves attention in the

consideration of alternate accounts of their development beyond that possibly provided by CIG. As seen in Figure 2, the geographic distribution of forms representing the etyma **addī li-* and **allī li-* is highly fractured: reflexes of **addī li-* inhabit a broad swath of western North Africa, those of **allī li-* a territory along the central Red Coast, and derivatives of both a smaller zone on the northern cusp of the Syro-Mesopotamian area. The three areas in question could thus almost not be farther from one another within the bounds of the modern Arabic-speaking world, and additionally in two of the three zones (the two largest), reflexes of only a single distinct etymon are encountered at time. These distributional characteristics stand in complete contrast to those described previously for the members of the GEN < PROPERTY and GEN < THING grammaticalization pathways, and in roundly failing the study heuristic's condition (iii) call strongly for the identification of a developmental mechanism behind their widespread occurrence not based in processes of areal diffusion via dialect contact.

In sum, upon mapping the genitive exponent data described in the previous sections we find that each of the two multiply attested grammaticalization pathways identified by the analysis of internal structural criteria is distributed across the contemporary Arabic-speaking world in a geographically contiguous manner. In this light, the products of both GEN < PROPERTY and GEN < THING are seen to successfully meet condition (iii) of the study's heuristic in addition to conditions (i) and (ii), the implications of which are discussed below. The patterning of the reflexes of GEN < REL + DAT, however – a major family of forms in and of itself – emerges from the geographic analysis as qualitatively different from that previously observed in relation to the other two pathways. This fact informs the search for alternative account for the widespread incidence of these forms, as described in §4.4 immediately below.

4.4 Discussion and Conclusions

This chapter examined the development of the Arabic genitive exponents, a consistently cited example of pluriform development, and evaluated their modern forms as consistent with a historical account of development via CIG. In the first stage of the analysis, each of the 144 individual genitive exponent forms encountered in the sample was assigned to an appropriate etymological source, in this case determined to include 14 distinct etyma. Once this operation was complete, the specifics of development from source to target form were compared with general theoretical understandings of the nature of grammaticalization; this included a review of said developments for evidence of the recognized grammaticalization component processes of desemanticization, extension, decategorialization and erosion, and was supplemented by a search for obvious correlates which might support or detract from an evolutionary account based on purportedly universal human cognitive processes and principles. Simultaneously, the individual developments attested were evaluated with an eye to which might be grouped together under the higher level heading of grammaticalization pathway, representing a set of etymologically distinct but semantically and functionally analogous innovations which might plausibly be linked to one another by means of historical replication.

Following these steps of the analysis, two major grammaticalization pathways were identified among the developments leading to the modern Arabic genitive exponent forms which both met the theoretical criteria established for grammaticalizing change and attested multiple synonymous but etymologically distinct realizations, thereby satisfying both conditions (i) and (ii) of the study's heuristic as laid out in §1.4. These are the pathway GEN < PROPERTY, comprising products of the individual etyma **matāʕ*, **tabaʕ*, **ħaqq*, **māl* and **jənā*, and the pathway GEN < THING, represented by reflexes of the etyma **hana*, **ʃayyit*, **ħājit* and **ʃuɣl*. An additional numerically significant and

multiply attested path of development revealed was that of GEN < REL + DAT, which included the thirty-six extant reflexes of the distinct etyma **addī li-* and **allī li-*. These were shown, however, not to display evidence of desemanticization or extension, and thereby to differ from the products of the other pathways considered by lacking two widely recognized characteristics of grammaticalizing change. Rather than showing a loss of original source semantics and subsequent development of novel pragmatic usage contexts, the full functional range of these forms was found to be comparable to that of the inherited dative preposition **li-*, in many cases interacting with wider reforms to the relative marking and definiteness systems of the relevant dialects. This in no way negates the significant evolution these items have undergone in form and structure, including lexicalization, phonetic erosion, and in a few cases the innovation of morphological agreement marking, nor does it negate the importance of these derivations to any comprehensive account of the diachrony of the Arabic genitive exponents. At the same time, however, these facts do argue strongly for an explanation outside of CIG in the manner it has empirically presented for other items investigated here.

The next phase of the analysis involved the evaluation of the heuristic's condition (iii), the requirement that any forms to be recognized as the products of CIG must display a contiguous geographic distribution consistent with a history of diffusion via areal contact. The realizations of the grammaticalization pathway GEN < PROPERTY easily met this threshold, displaying multiple, robust intersections and overlapping distributions across their entire incidence. The representatives of the path GEN < THING similarly showed geographic cohesion, though over a smaller area than those of GEN < PROPERTY and with less substantial documentation over the full course of their range; it was shown, however, that even a conservative interpretation of their geographic incidence would still succeed in meet the requirements of condition (iii). The results for GEN < REL + DAT

were significantly more scattered, spread widely and occupying disparate areas of the far west, southeast and northeast of the Arabic-speaking region, with reflexes of its two constituent etymologies sharing territory in only one of these three zones. Thus, while contact-based influence remains possible on a small scale in the Anatolian area, it is an implausible account for the modern incidence of the GEN < REL + DAT pathway as a whole.

The collected products of GEN < PROPERTY and GEN < THING thus show every indication of representing the results of repeated processes of CIG. Under such a model, an initial, innovative grammaticalization would have taken place whereby a substantive meaning property or thing, respectively, was adapted to serve as a grammatical marker of genitive or possessive relationship, consistent with such processes commonly observed worldwide. Following this first development or – equally plausibly but perhaps less parsimoniously – developments, the novel usage was encountered and comprehended by speakers of neighboring dialects who proceeded to assimilate the semantic/functional underpinning of the development and replicate its progression using distinct but synonymous etymological material. This process may be presumed to have repeated, combined with more traditional forms of matter-based feature borrowing and spread, until the GEN < PROPERTY and GEN < THING reflexes reached their contemporary distribution. The continued propagation of the GEN < THING path in particular shows an interesting turn whereby the same lexical item, **hana*, serves as a grammaticalization source at two opposite ends of the pathway's geographic incidence. This apparently shows that in cases of CIG between genetically related varieties the same source construction may be independently identified by speakers multiple times as a viable means of expression and replication; this perhaps provides a more elegant account of the

observed similarity in this case than does a claim of shared inheritance between two dialect areas which otherwise share no diagnostic isoglosses in common.

The evolutionary details of the GEN < REL + DAT trajectory are qualitatively different, and open the possibility of explication via an entirely distinct mechanism. As has been seen, the developments witnessed by the representatives of this path lack the specific semantic and pragmatic dimensions discussed above as typical of grammaticalization, and instead constitute a predictable continuation of inherited source functionality and syntax down to a detailed and highly specific structural level. Moreover, their strikingly separated geographic incidence all but precludes a history of propagation through areal contact, and rather supports an account based on internal rather than external actuators.

Taken collectively, these findings provide more support to the traditional Arabist's hypothesis of linguistic drift as an explanation for the GEN < REL + DAT forms than they do one relying on CIG. In such a formulation, the inherited structural specificities of the Arabic dialects showing products of this path, both semantic and syntactic, would have in a sense "prefigured" the trajectory of development observed to have occurred at least three independent times in three distinct geographic areas. This explanation is entirely system internal, and triggered structural pressures stemming from the cumulative set of prior linguistic developments shared by the varieties. To borrow a metaphor from Law, "[like] parallel sets of dominoes, knocking over the first piece in each set will have similar consequences – the other pieces fall in a logical pattern, according to the preexisting organization" (2014: 155). In each of the Arabic cases, the common possessive/genitive functionality of inherited **li-* interacted with the particular specifying status of the frequently co-occurring, adjacent relative particle to precipitate a lexical coalescence of the two into a unitary marker of possessive/genitive relationship

which shows a specific syntactic distribution consistent with its origin. Thus, that the resultant developments are similar "is due to the preexisting similarity among the [language varieties], and not because of contact or universal tendencies in language change" (Law 2014: 155).

This conclusion is supported by the cross-linguistic distribution of the GEN < REL + DAT path, which as previously noted does not prominently appear outside the context of the Semitic language family. Within Semitic, however, its recurrence is striking and frequent: alongside the Arabic forms already discussed, Rubin (2005: 55-56) cites at least Syriac *dīl*, Bax'a (Western Neo-Aramaic) *ci l-*, Modern Hebrew *ʃel* and Tigre *əntəl*, all of which derive from a parallel lexicalization of a relative clause marker and the cognate dative preposition **li-* and are united solely on the basis of common structural inheritance, display no common factors of geography, chronology or even genetic subclassification.[9] Rubin perhaps hastily identifies these along with the Arabic forms (of which he notes only the North African derivatives of **addī li-*) as examples of grammaticalization, without reference to any supporting semantic, pragmatic or formal evidence or mention of cross-linguistic precedent. On the basis of the facts laid out above, I depart from this analysis in concluding that the Arabic products of GEN < REL + DAT, and very likely their cross-Semitic counterparts, represent the results of linguistic drift attributable to independent, internally motivated developments shaped by shared particularities of inherited linguistic structure.

In conclusion, a prominent role has been identified for CIG in the evolution of the Arabic genitive exponents, resulting in two widespread and multiply attested families of

[9] As Rubin (2005) indicates, the role of relative particles in the evolution of Semitic genitive constructions is in fact even more widely attested than this; here, though, I prefer to restrict myself to those examples which specifically mirror the cross-linguistically unique trajectory GEN < REL + DAT reflected by the Arabic data (of which there are plenty).

forms arising from the grammaticalization pathways GEN < PROPERTY and GEN < THING. A third multiply attested developmental pathway, GEN < REL + DAT (accounting for approximately one quarter of the genitive exponent reflexes recorded in the sample), did not offer strong evidence for an account of CIG but rather displayed details favoring an interpretation of linguistic drift, thereby demonstrating how multiple lines of explication may be drawn upon without contradiction to interpret the varied and complex data presented by the Arabic pluriform developments.

Chapter 5: Conclusions, Reflections and Future Steps

Following the presentation of results from each of the three individual case studies included above, I now transition to synthesize the overall findings of this investigation in an integrated manner, up to and including an evaluation of the strengths and weaknesses of the methodology employed as applied to the distinct sets data of examined here. I then proceed to consider implications of these conclusions as they relate both to the study of Arabic diachrony and to the understanding of CIG as a global phenomenon. In the final section of this chapter, I propose directions for future research, both my own and with good fortune that of others, with an eye to corroborating and clarifying the proposals I advance in the current work.

5.1 Summary and Discussion of Results

In Chapters 2, 3 and 4, respectively, we have considered the role of CIG as a formative mechanism underlying the evolution of the modern Arabic future tense markers, temporal adverbs signifying 'now', and analytic genitive exponents, all members of the oft-discussed class of Arabic pluriform developments. In Chapter 2, two major pathways of future tense marker development were identified which met all three conditions established by the study's heuristic as indicative of an origin in a process of CIG: FUT < GO and FUT < WANT. Products of the four unique etymologies representing the FUT < GO pathway, **rāyiḥ, *ɣādī, *māʃī* and **sāyir*, were found across a large northwesterly region of the Arabic-speaking speaking world, stretching across the North African littoral to the Levant and Mesopotamia in an unbroken, geographically contiguous manner. Reflexes of **yabɣā ~ yabɣī, *biddu ~ widdu, *yiʃā, *ydawr* and **bɣā*, the five unique etymologies comprising the path FUT < WANT, occupied a similarly expansive stretch

anchored to the east and south, including Saharan and Sub-Saharan African locations as well as the Arabian Peninsula and overlapping with deallative counterparts across a significant portion of Levantine territory. Together, the products of these two trajectories of development account for 128 of 139 total future tense markers attested in the sample, indicating that the vast majority of modern Arabic future tense forms are the products of developments demonstrably aligned with the predicted results of CIG.

Chapter 3 addressed the evolution of temporal adverbs meaning 'now' in contemporary Arabic varieties. In the case of these items, two important developmental tracks, NOW < THIS TIME and NOW < THE TIME, were described as meeting the three conditions of the study's heuristic and consistent with a history of CIG alongside a third, NOW < IMMEDIATELY, which was not strictly identifiable as such but potentially compatible with the relevant parameters following secondary loss of territory to the Algerian arm of NOW < THIS TIME forms. Products of NOW < THIS TIME, comprising reflexes of the reconstructable source constructions **ðī l-waqt*, **ðā l-waqt*, **hā l-waqt*, **ðā l-ħīn*, **hā l-ħīn*, **hā s-sāʕa*, **hā l-ħazza* and **ðī l-ʔawān*, spanned a contiguous geographic bloc including the Saharan interior, the Nile Valley, the Levant, Mesopotamia and the greater part of the Arabian Peninsula. Reflexes of the individual source etyma **al-ħīn*, **al-ħazza*, representing the NOW < THE TIME grammaticalization pathway, covered a comparatively smaller but still geographically cohesive zone including eastern Syria, Northwest Arabia, and parts of Persian Gulf region. Together, the products of these two pathways constitute 120 of the sample's 143 total forms for 'now', once again displaying a significant potential role for CIG as an explanatory mechanism underlying these pluriform developments. If further evidence clarifying the spread of the Algerian **ðā l-waqt*-based forms were to come to light, it may be possible to incorporate the

remaining 23 items representing NOW < IMMEDIATELY into this fold, though for the moment any such proposal must remain purely speculative.

Lastly, in Chapter 4 we evaluated the modern exponents of a third Arabic pluriform development, that of the genitive exponents, for evidence pointing to a history of replication and spread through CIG. Two primary grammaticalization pathways were identified which met all three conditions of the study's heuristic: these were GEN < PROPERTY and GEN < THING. The former included reflexes of the source etyma **matāʕ*, **tabaʕ*, **ħaqq*, **māl* and **jənā*, and covered a territory ranging across nearly the entirety of the modern Arabic-speaking world, excepting only Anatolia, parts of coastal Morocco, and the West Sudanic area. The latter, composed of the individual etyma **ħana*, **ʃayyit*, **ħājit* and **ʃuɣl*, displayed a much more restricted but still unified geographic distribution, its reflexes occurring in a band reaching from the West Sudanic zone through Upper Egypt to the Northwest Arabian area and the Levant. Between these two pathways, CIG arises as a viable account for 101 of the 144 genitive exponent forms encountered in the sample. Uniquely among the three sets of pluriform items examined, investigation of the genitive exponents reveals a numerically substantial family of forms showing a common path of development not attributable to CIG, or for that matter grammaticalization in general or any contact-induced process. The pathway was that described as GEN < REL + DAT, and included reflexes of the two distinct etyma **addī li-* and **allī li-*. These forms were determined not to meet the conditions of the study's heuristic for the identification of the effects of CIG, and to perhaps be better ascribed to processes of linguistic drift, thereby representing the parallel results of internally motivated changes guided by a highly specific set of common structural inputs; the lack of contact as a driving factor in these forms' development was further evidenced by the highly fractured nature of their geographic distribution. The existence of the GEN < REL

+ DAT, then, makes the important point that CIG may not be (and shouldn't be expected to be) the sole force at work in shaping the Arabic pluriform data, and serves as a critical reminder of the Saussurean maxim that the headings under which "genitive exponents," "future markers," etc., are classed are purely synchronic descriptors of grammatical function, and can and do comprise operators stemming from completely distinct diachronic developmental trajectories.

In sum, then, this investigation concludes that a significant majority of the forms representing each of the three sets of items examined above show evidence of a development indicative of CIG, according to the criteria laid out in the study's heuristic. It is appropriate now to return briefly this heuristic itself to recall why it is that the specific confluence of factors in question has been selected as essential to identifying the empirical products of CIG, and why any consequent positive results may be taken with a high degree of confidence to indicate a history of evolution and promulgation through that process.

First, the heuristic's condition (i) requires that a rigorously defined identification of grammaticalization be made for the items considered at the level of the individual etymon, based on the examination of the data for evidence of four recognized component processes of grammaticalization – desemanticization, extension, decategorialization and erosion – and theoretically appropriate patterns of co-occurrence or primacy among these interrelated processes. Though perhaps basic in nature, care and attention during this step of the evaluation is fundamental and indispensable to any further argumentation, due simply to the fact that if a certain development is to be labeled a product of contact-induced grammaticalization it must first be established to be a bona fide output of grammaticalization processes more generally. The latter two conditions of the heuristic allow for the distinction between instances of CIG on the one hand and "run-of-the-mill,"

internally motivated cases of grammaticalization on the other. Condition (ii) establishes the presence of multiple, distinct realizations of a given pathway which are simultaneously synonymous or functionally analogous yet etymologically unrelated, thereby providing grounds for a claim that the grammaticalization pathway in question has undergone replication over the course of its history. The corroboration of a multiply attested pathway is a necessary but not a sufficient condition for claiming the occurrence of CIG, as grammaticalization paths by definition result from human cognitive universals and thus commonly recur across world languages. It is this state of affairs which necessitates the implementation of the heuristic's condition (iii). The requirement for a contiguous areal distribution seriously mitigates the likelihood that the distinct realizations of a given grammaticalization pathway may be attributed to the chance recurrence of an independent, internally-motivated change or series of changes.

In the latter circumstance, one would expect the geographic incidence of these realizations to mirror that encountered for the products of GEN < REL + DAT, argued here to be an internal development motivated by drift from a common set of structural precursors. The representatives of GEN < REL + DAT, accounting for a full quarter of the genitive exponents examined, occur in three distinct geographic pockets not relatable to one another by any obvious means of geographic adjacency, societal connectivity, or population movement. The observation, however, that the pathways passing condition (iii) cluster on the map in a decidedly non-random fashion, despite the fact that all dialects in the sample might be viewed as equally likely to have independently developed a realization of any of these cross-linguistically common grammaticalization paths, is strong evidence for the role of areal diffusion leading to their modern distribution, thereby confirming the "contact-induced" element of the finding that they are best described as the products of CIG.

As an organizing methodology, the application of this heuristic brings certain advances and certain limitations in comparison with previous approaches espoused in the study of both the Arabic pluriform developments and CIG. From the Arabist's perspective, its primary advance lies in its inherent focus not only on the diachronic sources of pluriform items but also on the longitudinal inference of diachronic processes linking these sources to their modern exponents. This priority is embodied both in the comprehensive tracing of phonological and morphological changes leading from historical sources to modern forms and, perhaps more prominently, in the adoption of a process-based definition of grammaticalization which hinges on the identification of specific diachronic changes in the domains of semantics, usage/pragmatics, morphosyntax, and phonology. Such should presumably represent the ultimate object of all historical linguistic inquiry: not solely to uncover evidence for an earlier state of a language at a previous point in time, but also to trace and describe how the language has progressed from that state to come to be what it is today (cf. Watkins 1973). However, these research priorities (or at least the successful execution of these priorities) remain regrettably rare in investigations of Arabic functional elements. The preoccupation with "whence" and lack of concern for "how" is perhaps excusable in an account like Eksell Harning's (1980) study of genitive exponents, which is positioned as an essentially synchronic description of these items' functional and usage details in comparison with alternative genitive marking strategies available in the dialects concerned; her reference, then, to etymological source acts more as an organizing principle than as a core research objective. It is more damaging, however, to examinations and surveys such as those presented by Rubin (2005) and Watson (2006), which take as their primary focus the grammaticalization of Arabic functional items yet pay insufficient attention to the investigation of actual processes of dynamic evolution evidenced in the data,

concentrating instead on the identification of seemingly static startpoints and endpoints. This is not to claim that process-based accounts of the evolution of Arabic pluriform data are entirely absent – Stewart (1998) is a welcome counter-example – but to date they have not been the norm in the field, and the present investigation thus represents an advance in that arena.

Among general studies of CIG, a specific contribution of the account presented here lies in its willingness to step beyond Heine and Kuteva's (2005: 185) preference to avoid linking multiple instances of contact-induced replica grammaticalization together to be considered as a single process of diffusion, as earlier advocated by Dahl (2001). Though Heine and Kuteva's position is a reasoned one in requiring that any proposed example of CIG stand on its own two theoretical feet, and its conservatism is perhaps a desideratum in any initial effort at rigorously establishing the existence of the phenomenon, it overlooks the significant potential of multiple, linked replications of the same grammaticalization pathway to stand as powerful evidence in distinguishing cases of CIG from instances of grammaticalization lacking an external impetus. Grammaticalization paths are cross-linguistically frequent by definition, and as such it is not beyond the realm of reasonable possibility that equivalent instantiations of a given pathway might independently co-occur in any two neighboring languages or dialects with no causal relationship linking them together. However, with each additional instance of replication in the same geographic area, the odds of chance similarity decrease dramatically. When this state of affairs has advanced to the degree evidenced, for example, by the five etymologically distinct realizations of FUT < WANT occurring within a single, continuous geographic bloc and not outside it, it may be concluded that the likelihood of those facts being the result of random occurrence is vanishingly small. Thus, while respecting Heine and Kuteva's warning that individual cases of proposed

replication must stand on their own merits, the present study's identification of five separate grammaticalization paths attested by three or more geographically contiguous realizations adds significant empirical weight to the assertion of CIG as distinguishable from the occurrence of grammaticalization as a broader phenomenon, in a manner which would not so convincingly provided by a more methodologically conservative investigation.

In discussing limitations of the methodology utilized across the previous chapters, two major factors may be mentioned, namely the heuristic's blindness to social aspects involved in the actuation of the CIG replication processes identified and to the absolute or relative chronology of the individual developments involved. The former is taken up in greater detail in §5.3 below, along with a few potential ameliorations, and thus will not be discussed at length here – for the moment, suffice it to note that any sociolinguistic characteristics of the CIG processes identified by these methods must be inferred indirectly from the evidence gathered or gleaned from external sources of nonlinguistic information pertaining to the sociological setting of a relevant time and place. In the second case, we must remark that the methodology employed in this investigation is inherently agnostic of the age or temporal ordering of the individual instances of replication identified to represent a given family of CIG-resultant forms. As Heine and Kuteva argue, each process of grammaticalization stemming from a unique replication event progresses once begun as a stand-alone piece of linguistic evolution (2003: 555-559): presumably, this entails the possibility that it may either outpace the development of the grammaticalization that inspired it or lag behind it, proceed to a logical completion or stall mid-process. As the rates of such progression are far from constant, whether within or across different languages or source constructions (Bybee 2011), it is not necessarily valid to simply identify the most advanced case of the grammaticalization

representing a given pathway as its oldest (and thus original) member, of which all others should be interpreted as replications or replications of replications. While in a fortunate few cases the capture of a grammaticalized usage in the sparse written record may provide a *terminus ante quem* for the emergence of a particular form, or the first attestation of a locally specific, innovative source a (somewhat weaker) logical bound from the opposite direction, values of absolute chronology similarly remain dark on the basis of the linguistic data alone. Thus, the methodology as it stands is unable to comment on the question of when the body of attested developments occurred, save within the extremely broad range of a) after the Arabic language area reached something approaching its contemporary geographic extent (though a smaller subset may have been present prior to this point in regions where Arabic has a longer history) and b) before the present. It is similarly agnostic regarding the related issue of directionality and the relative age of apparently replicated grammaticalizations, offering no clues as to a localizable origin of each interconnected family of forms within its observed contemporary distribution.

Having now recapitulated the study's major findings, advancements and limitations, I proceed to discuss in turn the implications of these conclusions and results for the broader study of historical Arabic dialectology on the one hand and the phenomenon of CIG on the other.

5.2 Implications for the Study of Arabic Diachrony

The present investigation proposes and defends a new account for the long-standing question of the Arabic pluriform developments, as it initially set out to do. It differs from previous accounts not only in method and its reliance on the novel theoretical construct of

CIG, but additionally and perhaps more fundamentally in the view it takes of the arc of Arabic dialect development over time. As described in §1.2, earlier approaches to the analysis of the Arabic pluriform elements have almost universally adduced that the mechanisms underlying the items' evolution and spread are traceable to the period preceding or immediately concurrent with the explosive expansion of the Arabic language beyond its traditional borders as a result of sudden Arabo-Islamic geopolitical and societal dominance in the seventh through tenth centuries. Such mechanisms are proposed to take multiple forms, ranging from latent structural predispositions (as Abboud-Haggar 2006) to reshuffled inherited traits (as Wilmsen 2014) to instantaneously effective creole universals (as Versteegh 1984), but all have in common the search for roots that predate the "birth" of the Arabophone region in the shape it is known today. The present account, however, provides powerful evidence that the pluriform developments continued to occur *following* the establishment of this new geographic and demographic domain of the Arabic language, if the areal diffusion of grammaticalization pathways observed is to be interpreted as such. The CIG-based approach, then, views the Arabic pluriform developments as an emergent phenomenon, arising gradually over an extended (and ongoing) period of *in situ* evolution. This stands in contrast not only to previous studies of pluriform developments specifically but also to more general, global approaches to historical Arabic dialectology, in which researchers have tended to attribute cross-dialectal difference and similarity to patterns of initial Arab expansion and settlement history (e.g., Kaye & Rosenhouse 1997; Palva 2006; Versteegh 2014) and to downplay the occurrence of significant convergent or divergent changes since that epoch. In so doing, they impute a degree of linguistic constancy or equilibrium to those intervening centuries which the evidence presented here indicates is not warranted.

The findings of this investigation similarly do not conform to conventional understandings of Arabic dialect subgrouping, with isoglosses at the level of both the etymon and the grammaticalization path cross-cutting varieties which, though geographically adjacent, are not otherwise understood to be closely related in a genetic sense. Of the five primary dialect groups traditionally distinguished – North African, Egypto-Sudanese, Levantine, Mesopotamian and Peninsular (cf. Versteegh 2014) – only the Levantine group may be seen to be defined by the occurrence of unique diagnostic features in the current sample: in this case, future tense markers derived from **biddu* ~ *widdu*, 'now' from **hā l-waqt*, and genitive exponents from **taba*ʕ. However, these features are neither fully ubiquitous in this area nor without competing reflexes derived from other sources; moreover, once the level of analysis is switched from the study of particular source etyma to that of broader patterns of replicated grammaticalization paths, the area's distinctiveness disappears entirely. Even the far more basic classificatory divides proffered by Kaye and Rosenhouse (1997), those based on a primary Eastern Arabic/Western Arabic split on the one hand and a fundamental distinction between Bedouin and sedentary varieties on the other, do not seem to be reflected by any discernable signal in the CIG data. The cases examined here reveal not a single grammaticalization pathway coextensive with an Eastern or a Western dialect bloc, as defined by the East/West divide postulated to run approximately between Egypt and Lybia, extending westward around the West Sudanic region (Kaye & Rosenhouse 1997: 265). Instead, the existence of major cross-cutting isoglosses like FUT < GO, FUT < WANT, NOW < THIS TIME, and GEN < PROPERTY emphasize continuity across this proposed split in a manner indicative of vigorous and extended contact between the two zones. Similarly, no consistent distinctions between the dialects of neighboring Bedouin

and sedentary populations are evidenced by the results of the CIG analysis, on either a micro- or a macro-scale.

This lack of correspondence between the data presented here and the traditional results of dialect classification might be explained on the basis that the data on which the prior classifications rest differs qualitatively from that examined in the current evaluation of CIG, the former having been specifically selected to ignore instances of later innovation and prefer features reconstructable to a greater time depth, thereby reflecting distinct branches of genetic relationship to which the developments realized by way of CIG are chronologically secondary. This interpretation might be readily accepted were it not for the fact that the traditional subgroupings (North African vs. Egypto-Sudanese vs. Levantine vs. Mesopotamian vs. Peninsular, Eastern vs. Western, Bedouin vs. sedentary) are in all cases presented as constituting descriptive groupings based on synchronically viable isoglosses, and in many cases rely on lines of argumentation – such as negative evidence and shared inheritance – that are highly problematic when evaluated as reflecting common genetic descent (e.g., Kaye & Rosenhouse 1997; Palva 2006; Versteegh 2014). Thus, the inability of the previously proposed classification schemes to accommodate developments such as those documented here as occurring via CIG poses a significant challenge to their broader acceptability, at least as currently formulated. In this light, the findings of the present investigation may add additional empirical backing to the position that conventional Arabic dialectological subgroupings are the product of a "pseudo-diachronic" approach, which attends primarily or solely to feature types deemed to be suitable for deep genetic reconstruction (e.g., regular sound correspondences, inflectional morphology, core lexicon) while not simultaneously carrying through on the application of rigorous comparative methodology which might justify the exclusion of other forms of evidence relevant to a professed goal of descriptive, synchronic

classification. This disconnect between evidence and argumentation in Arabic dialectological inquiry has been noted before in relation to the domains of morphology (Magidow 2017) and phonology (Taine-Cheikh 1998), and deserves attention in relation to the aforementioned tendency to ascribe pluriform developments to the pre-Islamic or early Islamic eras, with the general, unmotivated exclusion of manifestly innovative variables rather cyclically supporting the self-fulfilling assertion that "[by] the 10th century, dialectal areas were already shaped" (Abboud-Haggar 2006: 620).

While it is valuable to reconsider the dialectological subgroupings of Arabic as they stand in light of new analyses, I believe the most important contribution of the data presented here is rather to call into serious question the suitability of any single classificatory model to express a unified account of Arabic language diversity. The question has already been raised in the broader Semitic context by Huehnergard and Rubin (2011), who note the occurrence of both tree-like and wave-like diachronic development across Semitic languages (Arabic included) which can at times lead to concurrent states of convergence and divergence between languages depending on which particular feature or set of features is considered. The instances of CIG considered here complicate this picture even further, in that they present scenarios in which both convergent and divergent directionality may be imputed to a single linguistic change. For example, the areally diffused innovation of a genitive exponent via the pathway GEN < PROPERTY in geographically adjacent areas of Egypt, the Levant and the Arabian Peninsula should certainly be viewed as an example of convergence linking varieties spoken across these three zones. Simultaneously, however, the etymologically distinct surface forms of **matāʕ*, **tabaʕ* and **ħaqq* which reflect this common change may equally be seen to increase the degree of divergence between the three areas, resulting as they do in sociolinguistically significant variables differentiating the usage of each region

from that of its neighbors. This collapse of a clear convergence/divergence dichotomy is typical of, but not limited to, the results of CIG, and perhaps prompts the question of whether a subgrouping framework based on a metaphorical view of "language/dialect-as-unit" is appropriate in this case at all (for additional perspective on such a position in language contact studies, see Enfield 2003; in the study of Arabic in particular, Magidow 2017). Whatever the future of the analysis presented here, it is clear that this CIG data is not compatible with traditional paradigms underlying the description of Arabic language diversity and change, and thus adds to the call for far-reaching reevaluation in this area.

5.3 Implications for the Study of Contact-Induced Grammaticalization

The most immediately evident contribution of the present work to the broader study of CIG lies in its empirical corroboration of the phenomenon in the context of genetically related varieties. Though an important fixture in early formulations of CIG theory, scenarios involving the diffusion of grammaticalization pathways across related languages have not figured prominently in subsequent descriptions and studies due to the methodological difficulties inherent in unambiguously establishing the occurrence of CIG in instances when other potential mechanisms to account for similarity, such as drift and shared inheritance, are readily available (cf. Heine & Kuteva 2003; 2005). While exceptions to this tendency do exist, they are generally presented in short form (e.g., Dahl 2001) or as partial components of broader accounts (as Law 2014), on the whole studies of CIG between related language varieties remain sparse and the current investigation thus presents a timely step toward filling that critical lacuna.

I deem such an expansion of knowledge critical based on its ability to address two pressing questions (or suites of questions) relevant to a thorough understanding of CIG as

a diachronic process. The first question is whether the factor of genetic relatedness serves in any way to encourage or facilitate the occurrence of CIG between two language varieties in contact. Such a hypothesis is certainly plausible on a theoretical level, in that a similar arrangement of grammatical categories and values between linguistic relatives (close or distant) might allow for systemically viable replica grammaticalizations on a more frequent basis than otherwise; additionally and perhaps even more pertinently, one might posit that speakers of closely related varieties within the threshold of mutual intelligibility would require significantly less prior exposure and bilingual ability in order to successfully make the interlingual/interlectal identifications necessary to infer the lexical source of a model grammaticalized construction encountered, a prerequisite to any subsequent replication. Studies like this one are well positioned to evaluate such a hypothesis, if placed in comparison with additional investigations of contact between unrelated languages in a controlled manner (a proposal which will be further elaborated in the discussion of future research prospects included in §5.3, below).

The second question on which studies of CIG between closely related varieties might shed light is the ever-present issue of actuation, namely how the linguistic processes of replication apparently observed are executed on the sociolinguistic level of the speech community and the psycholinguistic level of the individual. This topic is one which Heine and Kuteva (2003; 2005) do not treat in detail in their foundational account of CIG, and they have been legitimately criticized for this omission – see, for example, Ross (2007). Matras (2009; 2011) however, provides a more elaborated proposal for the actuation of replication at the level of the individual bilingual, with a basis in the creative impulse theorized to underlie both grammaticalization and general patterns of structural borrowing. Matras suggests that the bilingual speaker, aware of active grammatical metaphors and their corresponding source forms in the model language, is driven to

recreate them in the replica language due to "the need to meet certain expectations of the interlocutor in respect of the choice of word forms" (2011: 288); essentially, that CIG arises through the interaction of contact-inspired, creative instances of language use with sociolinguistic norms or communicative needs governing which language's surface forms are used with whom. The prediction, then, is that replication of grammaticalization pathways is most likely to occur in those circumstances under which 1) societal bilingual knowledge is insufficient to support widespread use of the model code or 2) there exists significant sociolinguistic pressure against utilizing multiple linguistic forms with the same interlocutor or class of interlocutors (for precedent, see Epps 2005).

In the case of CIG between varieties as closely related as the modern Arabic dialects, the former condition appears unlikely, and it seems that the primary available motivator for CIG actuation – at least as described by Matras – is a sociolinguistic constraint disfavoring the use of forms external to the speech community's conventionalized linguistic repertoire. Such a state of affairs is, of course, typical of "focused" language varieties that display low internal variability and external permeability in scenarios of dialect contact (Le Page & Tabouret-Keller 1985) and is fully conceivable in the Arabic case, in which lexical distinctions consistently arise as highly salient (Behnstedt & Woidich 2005). If focused sociolinguistic norms of this type were operable, the cartographic outcome of replication in a feature mapping project like this one would be a series of relatively "clean" internal borders separating distinct etymological realizations of the same grammaticalization pathway, representing the constraining influence of speech community divisions through which the underlying metaphorical framework of a given grammaticalization process is able to diffuse but its accompanying etymological trappings are not. This type of distribution is indeed well evidenced among the Arabic data presented here, perhaps best demonstrated by the

multifarious etyma representing the future tense markers of each of the FUT < GO and FUT < WANT paths, which are largely arrayed in sequential, adjacent fashion in broad east-west stretches of the Arabic-speaking world.

A second type of geographic distribution is also attested, however, more prominently in the data for the genitive exponents and temporal adverbs 'now'. This distribution is not characterized by neat internal divisions within grammaticalization paths but rather by significant areas of overlapping territory in which multiple etymological realizations of the same pathway co-occur. Examples include the Sudanese reflexes of **hā s-sāʕa* 'now', which exist alongside reflexes of other realizations of the NOW < THIS TIME pathway over nearly their entire range (Rubāṭāb *hassaʕ*, *daħīn*, Shukrīyah *hassaʕ ~ hassiʕ*, *daħīn* and Kordofan *hassaʔ, dahīn* < **hā s-sāʕa*, **ðā l-ħīn*; Nigeria *hassa*, *duggut* < **hā s-sāʕa*, **ðī l-waqt*), and the Peninsular attestations of **māl*, which invariably co-occur with reflexes of fellow GEN < PROPERTY product **haqq* (Kuwait *māl, ħagg*; Sudayr *māl, ħagg*; Bahrain *māl, ħagg*; Abu Dhabi *māl, ħagg*; Dhofar *māl, ħaqq*). Areal dispersions like these seemingly conflict with the proposed model for CIG actuation based on patterns of lexical avoidance and focused boundaries between speech communities (which, we recall, are also very much in evidence in other subsets of the Arabic CIG data). They may thus be interpreted in at least three ways.

This first interpretation is that both the discrete and overlapping patterns are reflective of the same actuation process viewed at different points in its progression. In this scenario, one would presume the discrete etymological isogloss boundaries of, for example, FUT < WANT to represent the earlier stage, still more or less directly transmitting the focused speech community boundaries in place at the time of replication in the initial phase of CIG. Under this model, the frequent overlap and interpenetration of distinct etymological realizations characteristic of pathways like GEN < PROPERTY

would represent a later stage of development following that of the original actuation, wherein the once focused sociolinguistic norms which resulted in replication have become more diffuse and allowed for the subsequent occurrence of matter-based dialect borrowings. These secondary sociolinguistic developments would allow for the "spill-over" of once geographically adjacent forms into territory occupied by etymologically distinct products of the same grammaticalization path, resulting in the overlapping distributions observed in much of the Arabic CIG data today. Such vacillation between diffuse and focused norms is common fare in situations of dialect contact over time (Trudgill 1986), and given the aforementioned blindness of the current method to questions of relative time depth there is nothing in the data presented here which would inherently disqualify such an account.

The second interpretation, however, positions the overlapping incidences of distinct realizations of the same grammaticalization path as original, and the emergence of clear-cut, impermeable isogloss boundaries – as observed for multiple pathways in the Arabic case – as a secondary phenomenon. This would mirror, for example, the scenario proposed to have governed the development of distinct discontinuous negation markers in the Romance languages of France and northern Italy. In historically documented forms of these languages, a number of substantives with a lexical meaning of something like 'a bit, a small amount' were grammaticalized to more or less advanced degrees, first as minimizing adverbs and subsequently as negative polarity items that eventually become obligatory components of verbal negation. In its earliest stages, this grammaticalization pathway was represented within participating dialects by multiple distinct but semantically related lexical items which all came to perform analogous grammatical functions – Old French negators, for example, included reflexes of *pas* 'step', *mie* 'crumb', *point* 'point', *goutte* 'drop' and *grain* 'grain' (Zeijlstra 2010). Over time, this

field was thinned in each of the individual subregions constituting this larger area due to linguistic focusing and possibly structural forces of obligatorification and paradigmaticization pertaining to the ongoing grammaticalization process itself (for discussion, see Hansen & Visconti 2009). This progressed to the point that the contemporary French and Gallo-Italian varieties typically show a single grammaticalized relic of this former diversity which presents a clear isogloss boundary with forms used in other dialects: compare diffuse Old French *pas*, *mie*, *point* with focused (modern) French *pas*, North Italian *miga*, Florentine *punto*, etc. (Poletto 2016).

Specifics of the data, however, render such an explanation for the Arabic developments unlikely on several counts. Were the full range of etymological variants for each path (or even a large subsection of those variants) inherent to a given dialect at an earlier stage of its history, then one would not necessarily expect any remaining multiplicity of forms to occur in specific transitional or border zones. Put another way, had the entirety of the Arabian Peninsula and Mesopotamia once contained active forms of both **māl* and **ħaqq*, for what reason would variation between the forms continue to persist only in a "Venn diagram" pattern straddling a homogeneous **māl* and a homogeneous **ħaqq* area and not elsewhere? Further, details of phonetic form strongly suggest that many of the Arabic items in overlapping areas do not represent developments indigenous to those regions, such as the decidedly non-local occurrence of /d/ and /iy/ in Sinai *dilwagtiy* 'now' (occurring alongside fellow NOW < THIS TIME reflex *ħalħīn*) or the highly unusual sporadic /b/ of Khartoum genitive exponent *bitāʕ*, otherwise typical of the Egyptian area (found alongside *ħagg*, both of the path GEN < PROPERTY). This stands in stark contrast to the Romance negators mentioned just above, in which the cognate reflexes *mie* and *miga*, *point* and *punto* clearly display all outward signs of local etymological development.

These facts in turn support a third interpretation of overlapping isoglosses in the Arabic CIG data. In this version of events, matter-based borrowing of a grammaticalized functional item from Dialect A into Dialect B would precede the replication of its perceived grammaticalization pathway using etymological material inherent to Dialect B. In order for such borrowing to occur, the projected linguistic environment would be diffuse, without strong sociolinguistic constraints against the use of specific surface forms in interactions with a certain class of interlocutor. It is no longer possible, then, to attribute the etymologically distinct replication of a grammaticalization path to Matras' sociolinguistically driven filter of interlocutor expectation. Instead, we would seem to have returned full circle to Lehmann's (1985) "pure" manifestation of the creative impulse as prompting language internal grammaticalization processes, the urge to say the same thing in a different way (see §1.3.1). In this case, however, speakers "crib" from an alternate form already existent in their repertoire, which they have acquired through contact with an outside speech community, the end result being the contact-induced replication of a grammaticalization pathway that, until recently, was not active in the dialect in question.

Though evidence has been presented recommending against the adoption of the second interpretation, both the first and the third scenarios accounting for the existence of overlapping realizations of the same CIG pathway remain viable on the basis of the findings of the present investigation. Additional research based on the overlapping Arabic data or any similar cases identified cross-linguistically has much to add to discussions surrounding the actuation of CIG and the question of whether the filter of focused sociolinguistic norms is, as proposed by Matras, an indispensable component of the process.

5.4 Directions for Future Research

In addition to the research avenue described just above, several additional directions for future inquiry present themselves on the basis of the work presented here. Perhaps the most immediately executable of these would be an examination of the correspondence of the replicated Arabic grammaticalization paths described here to the grammaticalizations of similar features in neighboring and co-territorial non-Arabic languages, in an attempt to confirm or refute Heine and Kuteva's assertion that "genetic relationship is entirely irrelevant" to the process of CIG (2005: 184). Importantly, the Arabic case would allow for a tiered investigation of the degree to which the diffusion of grammaticalization pathways is observed to continue beyond the confines of closely related Arabic dialects, as the relevant regional languages examined would comprise multiple discrete levels of relation: other extant Semitic languages such as the Modern South Arabian languages, Tigre, numerous Neo-Aramaic varieties and Modern Hebrew, far more distantly related but in many respects typologically similar Berber and Cushitic languages, and representatives from entirely unrelated language groups like Iranian, Turkic and Nilo-Saharan. The observation of continuity (or lack thereof) across any or all of these tiers would constitute a meaningful contribution toward illuminating the influence of genetic relatedness in processes of CIG.

An additional direction for further research based on the findings presented here would involve a concentrated effort to attach relative or absolute chronology to a subset of the replicated forms identified by the present work. This would perhaps only be feasible in a scientifically rigorous way if restricted to a particular geographic area within which multiple distinct products of specific grammaticalization paths are located and for which sufficient historical attestation exists to make such an attempt. At first glance, the most viable option would likely be the intersection of Egypt and the coastal Levant, for

which substantial dialectological documentation is found from the late nineteenth century onward alongside scattered but not insignificant evidence dating from earlier periods. Increased evidence as to previous forms, functions, and incidences of the grammaticalized items in question could begin to shed light, if indirectly, on some of the questions of actuation raised in §5.3, and if nothing else would add a critically lacking chronological dimension to the descriptive findings presented by this study.

Thirdly, perhaps the single most exciting opportunity to build on the findings of the current investigation would be to capture a further instance of CIG occurring in process, i.e. at an incipient stage of grammaticalization closely following replication of an assimilated grammaticalization pathway. Such an occurrence, unfortunately, is not one that the researcher can simply walk out and find; it would require patience, perceptivity, and luck to identify this type of change taking place at an early and perhaps outwardly unremarkable stage, but given the prevalence of CIG processes attested by the Arabic data the task is perhaps not insurmountable. A valid starting point might be any of the comparatively "shallow" cases of grammaticalization uncovered in the present sample, such as the Nouakchott *ydōr* future tense marker or the Aswan *ħājit* genitive exponent, both of which show evidence of initial desemanticization but little in the way of extension, decategorialization or erosion. Examination of speakers' use/lack of use and explicit etymological knowledge of neighboring or co-territorial products of the FUT < WANT and GEN < THING pathways, respectively, could prove extremely valuable in efforts to confirm or refute various hypotheses relating to both the actuation and the advancement of CIG more generally.

5.5 CONCLUSION

In the preceding chapters, I have provided a framework for the evaluation of the Arabic pluriform developments as the products of contact-induced grammaticalization and applied my methodology in turn to the study of future tense markers, temporal adverbs 'now', and genitive exponents drawn from across the modern Arabic-speaking world. In so doing, I have put forward and defended the proposition that CIG is a prominent mechanism underlying the historical evolution of the forms examined. I have not claimed that CIG carries explanatory force for each and every piece of data considered in this investigation – when other diachronic processes have presented themselves as more suitable lenses for the interpretation of specific items, they have been preferred – but this fact does not detract from the considerable and pervasive apparent role of CIG in shaping the developments examined here. The expectation that a successful account be a universal and exclusive one is perhaps what has left the debate over the Arabic pluriform developments unresolved after nearly a century of concerted inquiry. It is my sincere hope that these findings and the spirit in which they are made contribute to a more nuanced understanding of Arabic diachrony and to an advanced recognition of the diverse forms and directions contact-induced grammaticalization may take as it operates among the languages of the world.

Appendix: Summary of Individual Dialect Forms and Sources

Dialect	Future Tense Marker	Temporal Adverb 'now'	Genitive Exponent	Source
'Abābdah	*rāħ ~ ħa-*	*dilwagti, dalwagti*	*bitāʕ*	(de Jong 2002)
Abéché	*b-*	*hassa*	*hana*	(Roth 1979)
Abha	*b-*	*ðalħīn*[10], *alħīn*[11]	*ħagg*	(Al-Azraqi 1998)
Abu Dhabi	*b-*	*halħīn ~ alħīn, halħazza ~ alħazza, lħīn, lħazza*	*ħagg, māl*	(Qafisheh 1977)
Aden	*bā-*	*daħīn*	*ħaqq*	(Feghali 1991)
Āl Wahībah	*bi-*	*halħīn ~ alħīn*	---	(Webster 1991)
Aleppo	*bəddo ~ bdo ~ b-*	*hallaq ~ hallaqtēn ~ hallaqne*	*tabaʕ*	(Sabuni 1980)
Algiers	*rāħ*	*dork, ðork*	*mtaʕ, dyal*	(Boucherit 2006)
Amman	*rāyiħ ~ rāħ ~ ħa-, biddo*	*halla*	*tāʕ, tabaʕ*	(Al-Wer 2006)
Anatolia	*tə- ~ də-*	---	*ðīla ~ ðīl ~ ðēla ~ ðēl, lēl ~ lē*	(Jastrow 2006)
Anjra	*māʃ*	*dāba*	*dyāl ~ d*	(Vicente 2000)
Antiochia	---	*ħallaq, alħaz*	---	(Arnold 2006)

[10] (Nakshabandi 1998)
[11] (Nakshabandi 1998)

Aswan	*ħa-, ʕa-*	*dilwagti, dilwakītī*	*bitāʕ, ihnīt*[12], *ħājt*[13], *līl*	(Schroepfer forthcoming)
ʻAwāmrah	*laħ ~ ħa-*	*dilwagt, halwagit*	*bitāʕ*	(Behnstedt & Woidich 1985)
Azru	---	---	*ntaʕ, dyāl ~ d*	(Singer 1980)
Baghdad	*raħ*	*hassa*	*māl*	(Abu-Haidar 2006)
J-Baghdad	*ɣaħ*	---	*māl*	(Mansour 2006)
Bahrain	*b-*	*alħīn, halħazza, ilħīn*	*ħagg, māl*	(Holes 2006a)
Banī Ṣakhr	---	*halħīn, hassāʕ ~ hassaʕ, ðilwān*	*ħagg*	(Palva 1980)
Basra	*rāyiħ*	*hassa*	*māl*	(Mahdi 1985)
Bdūl	*b-*	*halħīn, alħīn*	---	(Bani Yasin & Owens 1984)
Beirut	*raħ*	*hallaʔ*	*tāʕ ~ tāʕūl, tabaʕ*	(Naïm 2006)
Benghazi	*yibbi ~ yib- ~ ibi-*	*towwa*	*imtāʕ, ʃõr*	(Benkato 2014)
Bʻērāt	*rāħ ~ raħa ~ ħa-*	*dilgē, dilgēti*	*ibtāʕ, ihnīn*	(Woidich 2006b)
Cairo	*ħa- ~ ha-*	*dilwaʔti*	*bitāʕ*	(Woidich 2006a)
Cameroon	---	---	*hanā*	(Echu & Aminou 2006)
Casablanca	*ɣādi ~ ɣa-*	*dāba*	*mtāʕ, dyal ~ d*	(Caubet 2006)
Cherchell	*rāyiħ*	*ðərwaq*	*ntāʕ, dyāl*	(Grand'Henry

[12] (Alrawy, Slah, p.c.)
[13] (Alrawy, Slah, p.c.)

				1972)
Cilicia	*baddu ~ baddi- ~ bad-*	*hallaq*	---	(Procházka 2006)
Cyprus	*tta-*[14] *~ ta-*	*ʔalok*	*ʃayt, tel ~ te*[15]	(Tsiapera 1969)
Damascus	*raħ ~ laħ ~ raħa ~ laħa ~ ħa-, bəddo ~ b-*	*hallaʔ*	*tabaʕ ~ tabaʕīt, ʃīt*	(Lentin 2006)
Dellys	*rayəħ ~ ħa-*	*ðurwək ~ ðurk ~ ðˤukk ~ ðˤukka*	*ntaʕ ~ taʕ, dyal*	(Souag 2005)
Dhofar	*bā-*	*ðalħīn ~ ðaħħīn, ilħīn*	*ħaqq, māl*	(Davey 2016)
Djidjelli	*rāyiħ, māʃi*	*delwoq ~ derwoq*	*mtāʕ, əddil ~ əddi ~ dyāl*	(Marçais 1956)
J-Fez	*ɣadi, masi ~ mas*	*daba*	*ntāʕ, dyal ~ di*	(Heath 2002)
Fezzan	*bī ~ b-*	*taww ~ taw*	*mtāʕ, jnā ~ jən*	(Caubet 2004)
Goulimine	*ɣadi, lahi ~ la*	*druk, daba*	*dyal ~ d*	(Heath 2002)
Hadhramaut	*bā-*	*ðalħīn ~ ðaħħīn*	---	(Al-Saqqaf 2006)
Ḥarb	*yabɣa ~ yaba ~ ba-*	*ðaħīn, halħīn, ðulwān ~ ðuwān*	*ħagg*	(Il-Hazmy 1975)
Hit	*raħ*	*hassaʕ*	*māl*	(Khan 1997)
Jebel Ansariye	*baddo ~ bado ~ b-*	*hallaq*	---	(Lewin 1969)

[14] (Borg 1985)
[15] (Eksell Harning 1984)

Jerusalem	*rāyiħ ~ rāħ ~ ħā-, biddo ~ b-*	*halʔēt*	*tabaʕ*	(Rosenhouse 2006)
Jisr az-Zarqa	*rāyiħ, b-*	*hassa*	*tabaʕ, ʃīt*	(Belinkov 2014)
Kabābīsh	*bi-*	*taħīn*	---	(Hillelson 1935)
Kadugli	*ha-, bi-*	---	*bitā, hagg*	(Manfredi 2013)
B-Kadugli	---	*towwa*	*hān, hīl*	(Manfredi 2013)
Al-Khaburah	*b-*	*taww*	---	(Brockett 1985)
Kharga	*ha-*	---	*bitāʕ, ʃaɣl*	(Behnstedt & Woidich 1985)
Khartoum	*ħa-, bi-*	*hassaʕ ~ hassi*	*bitāʕ, ħagg*	(Dickins 2006)
Khawaytnah	*bəddu, ta-*	*hassaʕ ~ hassaʕēn*	*māl, hnīt, ʃīt, gī*	(Talay 1999)
Khuzestan	*raħ*	*hassa*[16]	*māl*	(Ingham 2006a)
Kordofan	*b-*	*dahīn, hassaʔ*	---	(Hillelson 1935)
Kuwait	*b-*	*alħīn, halħazza*	*ħagg, māl*	(Holes 2006b)
Larbaâ	---	*darwak ~ dark ~ darka*	*ntāʕ*	(Dhina 1938)
Mali	*ydawr, lāhi ~ lā*	*ðark, ðrayk*	*ntāʕ*	(Heath 2003)
Malta	*ħa-*[17]*, sɛyyɛr ~ sɛr- ~ sɛ-*[18]	*issa*	*ta*	(Mifsud 2006)
Marrakech	*ɣādi ~ ɣa-, bɣa ~ ba-*	*drūk, drūka, dāba*	*ntāʕ ~ tāʕ, dyāl ~ d ~ t*	(Sánchez 2014)
Mateur	*bāʃ*	*tawwa ~ taw*	*mtāʕ*	(Mion 2014)
Mecca	*rāħ ~ ħa-*	*daħħīn*	*ħagg*	(Abu-Mansour 2006)

[16] (Shabibi 2006)
[17] (Vanhove 1993)
[18] (Vanhove 1993)

Misīrīyah	*bi-*	*tauwa*	---	(Hillelson 1935)
Mosul	*də-*	*hassaʕ ~ hassaʕta*	*māl*	(Jastrow 1979)
M'zab	---	*dˁarwek ~ dˁrūk ~ ðrūk ~ dˁarka*	*mtāʕ*	(Grand'Henry 1976)
Negev	*rāyiħ, widdih ~ d-*	*halħīn ~ halħīniy, alħīn ~ alħīniy*	*tabaʕ, ʃuɣl*	(Henkin 2006)
Nigeria	*b-*	*duggut, hassa*	*hana, hīl*	(Owens 1993)
Nouakchott	*ydōr, lāhi*	*ðˁark*	*ntaʕ, dyal*	(Taine-Cheikh 2006)
Rubāṭāb	*be-*	*daħīn, hassaʕ*	*hīl*	(Hillelson 1935)
Sana'a	*ʃa-, ʕā-*	*ðalħīn*	*ħagg*	(Watson 2006b)
Saoura	*ɣādi, ba-*	---	*mtāʕ ~ ntāʕ*	(Grand'Henry 1979)
Saïda	---	*darwək ~ ðˁorwok*	*ntāʕ*	(Marçais 1908)
Shukrīyah	*bi-*	*daħīn, hassaʕ ~ hassiʕ*	*bitāʕ ~ butāʕ, ħagg, allīl, hīl*	(Reichmuth 1983)
Sinai	*rāħ ~ raħ ~ ħa- ~ ha-, widdih ~ biddu*	*dilwagtiy, ħalħīn ~ alħīn ~ halħīniy ~ alħīniy ~ halħīnit, ilħīn ~ ilħīnih*	*btāʕ ~ tāʕ, ħagg, ʃuɣl*	(de Jong 2006)
Soukhne	*b-*	*alħazz ~ alħaz ~ alħa ~ alħaztēn*	*tabaʕ, hanayyi, gayy*	(Behnstedt 1994)

Sousse	*māʃ ~ bāʃ*	*tawwa*[19]	*mtāʕ*	(Talmoudi 1980)
Sudayr	*yabi ~ ab- ~ b-*[20]	*halḥīn, alḥīn*	*ḥagg, māl*	(Ingham 1994)
Tetouan	*maʃ*	*daba*	*dyal ~ d*	(Heath 2002)
Tlemcen	---	*derwaq, dāba*	*ntāʕ, dyāl ~ dī ~ əddi*	(Marçais 1902)
Tozeur	*ḥa-, ʕa-*	*towwa ~ toww ~ tow ~ tū*	*mtāʕ ~ tāʕ ~ t ~ aʕ ~ ntīʕ*	(Saada 1984)
Tripoli	*ḥā-, bə- ~ b-*	*tawwa*	*mtāʕ*	(Pereira 2006)
Tunis	*bāʃ ~ biʃ*	*təwwa*	*ntāʕ*	(Gibson 2006)
J-Tunis	*māʃ*	*dəlḥīn, tawwa*	*ntāʕ ~ tāʕ*	(Cohen 1975)
Ẓafīr	---	*hālḥīn ~ hāḥīn, alḥīn ~ aḥīn*	---	(Ingham 1982)

[19] (Talmoudi 1981)
[20] (Ingham 2006b)

References

Abboud-Haggar, Soha. 2006. Dialects: Genesis. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Abdel-Massih, Ernest, Zaki Abdel-Malek, El-Said Badawi & Ernest McCarus. 2009. *A reference grammar of Egyptian Arabic*. Washington, D.C.: Georgetown University Press.

Abu-Haidar, Farida. 2006. Baghdad Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Abu-Mansour, Mahasen Hasan. 2006. Meccan Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Al-Azraqi, Munira. 1998. Aspects of the syntax of the dialect of Abha (southwest Saudi Arabia). University of Durham: Durham Doctoral dissertation.

Al-Jallad, Ahmad. 2015. *An outline of the grammar of the Safaitic inscriptions*. Leiden, Boston: Brill.

Al-Saqqaf, Abdullah Hassan. 2006. Wādī Ḥaḍramawt Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Al-Wer, Enam. 2006. Jordanian Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Andersen, Henning. 2008. Grammaticalization in a speaker-oriented theory of change. In Thórhallur Eythórsson (ed.), *Grammatical change and linguistic theory: The Rosenthal Papers*, 11–44. Amsterdam, Philadelphia: John Benjamins.

Arnold, Werner. 2006. Antiochia Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Bani Yasin, Raslan & Jonathan Owens. 1984. The Bduul dialect of Jordan. *Anthropological Linguistics* 26(2). 202–232.

Bauer, Leonhard. 1926. *Das palästinische Arabisch: Die Dialecte des Städters und des Fellachen*. Leipzig: J. C. Hinrichs'sche Buchhandlung.

Behnstedt, Peter. 1994. *Die arabische Dialekt von Soukhne (Syrien): Teil 2. Phonologie, Morphologie, Syntax, Teil 3. Glossar*. Wiesbaden: Harrassowitz Verlag.

Behnstedt, Peter. 1997. *Sprachatlas von Syrien*. Wiesbaden: Harrassowitz Verlag.

Behnstedt, Peter. 2016. *Dialect atlas of North Yemen and adjacent areas*. (Trans.) Gwendolin Goldbloom. Leiden, Boston: Brill.

Behnstedt, Peter & Manfred Woidich. 1985. *Die ägyptisch-arabischen Dialekte*. Wiesbaden: Dr. Ludwig Reichert Verlag.

Behnstedt, Peter & Manfred Woidich. 2005. *Arabische Dialektgeographie: Eine Einführung*. Leiden, Boston: Brill.

Behnstedt, Peter & Manfred Woidich. 2012. *Wortatlas der arabischen Dialekte: Band II. Materielle Kultur*. Leiden, Boston: Brill.

Behnstedt, Peter & Manfred Woidich. 2014. *Wortatlas der arabischen Dialekte: Band III. Verben, Adjektive, Zeit und Zahlen*. Leiden, Boston: Brill.

Belinkov, Yonatan. 2014. The Arabic dialect of Ǧisr izZarga: Linguistic description and a preliminary classification, with sample texts. Tel Aviv: Tel Aviv University Master's thesis.

Ben Abdelkader, Rached, Abdeljelil Ayed & Aziza Naouar. 1977. *Peace Corps English-Tunisian Arabic Dictionary*. Washington, D.C.: Peace Corps.

Benkato, Adam. 2014. The Arabic dialect of Benghazi, Libya: Historical and comparative notes. *Zeitschrift für Arabische Linguistik* 59. 57–102.

Bin Sharīfah, Muḥammad. 1987. *Mal'abat al-Kafīf al-Zahrūnī*. Rabat: al-Maṭba'ah al-Malakīyah.

Bisang, Walter. 1998. Grammaticalization and language contact, constructions and positions. In Ana Giacalone-Ramat & Paul Hopper (eds.), *The limits of grammaticalization*, 13–58. Amsterdam, Philadelphia: John Benjamins.

Borg, Albert & Marie Azzopardi-Alexander. 1997. *Maltese*. London, New York: Routledge.

Borg, Alexander. 1985. *Cypriot Arabic: A historical and comparative investigation into the phonology and morphology of the Arabic vernacular spoken by the Maronites*

of Kormakiti village in the Kyrenia district of North-Western Cyprus. Stuttgart: Kommissionsverlag Franz Steiner Wiesbaden GMBH.

Boucherit, Aziza. 2006. Algiers Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Britain, David (ed.). 2007. *Language in the British Isles*. Cambridge: Cambridge University Press.

Britain, David. 2014. Geographical dialectology. In Janet Holmes & Kirk Hazen (eds.), *Research methods in sociolinguistics: A practical guide*, 246–261. Malden: Wiley Blackwell.

Brockett, A. A. 1985. *The spoken Arabic of Khābūra on the Bāṭina of Oman*. Manchester: Journal of Semitic Studies.

Brustad, Kristen. 2000. *The syntax of spoken Arabic: A comparative study of Moroccan, Egyptian, Syrian, and Kuwaiti dialects*. Washington, D.C.: Georgetown University Press.

Burr, J. Millard & Robert Collins. 2008. *Darfur: The long road to disaster*. Princeton: Markus Weiner Publishers.

Bybee, Joan. 2003. Cognitive processes in grammaticalization. In Michael Tomasello (ed.), *The new psychology of language*, vol. II, 145–167. Mahwah, New Jersey and London: Lawrence Erlbaum Associates, Inc.

Bybee, Joan. 2011. Usage-based theory and grammaticalization. In Heiko Narrog & Bernd Heine (eds.), *The Oxford handbook of grammaticalization*, 69–78. Oxford: Oxford University Press.

Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago, London: Chicago University Press.

Camilleri, Antoinette. 1997. *Merħba bik: Welcome to a course in Maltese for foreigners*. Mġarr, Malta: Colour Image.

Caubet, Dominique. 1993. *L'arabe marocain: Tome II. Syntaxe et catégories grammaticales, textes*. Paris, Louvain: Éditions Peeters.

Caubet, Dominique. 2004. Les parlers arabes nomades et sédentaires du Fezzàn, d'après William et Philippe Marçais. In Martine Haak, Rudolf de Jong & Kees Versteegh (eds.), *Approaches to Arabic dialects: A collection of articles presented to Manfred Woidich on the occasion of his sixtieth birthday*, 67–96. Leiden, Boston: Brill.

Caubet, Dominique. 2006. Moroccan Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Chambers, J. K. & Peter Trudgill. 1998. *Dialectology*. 2nd ed. Cambridge: Cambridge University Press.

Claudi, Ulrike & Bernd Heine. 1986. On the metaphorical basis of grammar. *Studies in Language* 10(2). 297–335.

Coetsem, Frans van. 1988. *Loan phonology and the two transfer types in language contact*. Dordrecht: Foris.

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

Cohen, David. 1970. Le système des voyelles brèves dans les dialectes maghribins. *Études de linguistique sémitique et arabe*. The Hague, Paris: Mouton.

Cohen, David. 1975. *Le parler arabe des Juifs de Tunis: Tome II. Étude linguistique*. The Hague, Paris: Mouton.

Colin, G. S. 1945. De Juifs nomades retrouvés dans le Sahara marocain au XVIe siècle. *Mélanges d'études luso-marocaines dédiés à la mémoire de D. Lopès et P. de Cénival*, 53–66. Lisbon.

Corriente, Federico. 1971. On the functional load of some synthetic devices in Arabic and Semitic morphology. *Jewish Quarterly Review* 52. 20–50.

Corriente, Federico. 1977. *A grammatical sketch of the Spanish Arabic dialect bundle*. Madrid: Instituto Hispano-Arabe de Cultura.

Cowell, Mark W. 1964. *A reference grammar of Syrian Arabic (based on the dialect of Damascus)*. Washington, D.C.: Georgetown University Press.

Dahl, Östen. 2001. Principles of areal typology. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals: An international handbook*, 1456–1470. Berlin, New York: Walter de Gruyter.

Davey, Richard J. 2016. *Coastal Dhofari Arabic: A sketch grammar*. Leiden, Boston: Brill.

Davies, Humphrey. 1981. 17th-century Egyptian Arabic: A profile of the colloquial material in Yūsuf al-Širbīnī's Hazz al-Quhūf Fī Šarḥ Qaṣīd Abī Ṣādūf. Berkeley: University of California, Berkeley Doctoral dissertation.

Davies, Humphrey. 2016. *Yusuf al-Shirbini: Brains confounded by the ode of Abu Shaduf expounded*. . Vol. 1. New York: New York University Press.

Davydova, Julia. 2014. Morphosyntactic analysis in sociolinguistics. In Janet Holmes & Kirk Hazen (eds.), *Research methods in sociolinguistics: A practical guide*. Hoboken: Wiley Blackwell.

DeGraff, Michel. 2005. Linguists' most dangerous myth: The fallacy of Creole Exceptionalism. *Language in Society* 34(4). 533–591.

Dhina, Amar. 1938. Notes sur la phonétique et la morphologie du parler des 'Arbâɛ. *Revue Africaine* 82. 313–353.

Dickins, James. 2006. Khartoum Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Echu, George & Mohamadou Aminou. 2006. Cameroon Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Eid, Mushira. 1983. The copula function of pronouns. *Lingua* 59(2–3). 197–207.

Eksell Harning, Kerstin. 1980. *The analytic genitive in the modern Arabic dialects*. Göteborg: Acta Universitatis Gothoburgensis.

Eksell Harning, Kerstin. 1984. On the development of d-particles as genitive exponents in Arabic dialects. *Acta Orientalia* 45. 21–42.

Elias, Elias & Edward Elias. 1981. *Egyptian Arabic manual for self study*. 19th ed. Cairo: Elias Modern Publishing House.

Enfield, N.J. 2003. *Linguistic epidemiology: Semantics and grammar of language contact in mainland Southeast Asia*. London, New York: RoutledgeCurzon.

Epps, Patience. 2005. Areal diffusion and the development of evidentiality: Evidence from Hup. *Studies in Language* 29(3). 617–650.

Farāhīdī, al-Khalīl ibn Aḥmad al-. 1980. *Kitāb al-'Ayn*. (Ed.) Mahdī al-Makhzūmī & Ibrāhīm al-Sāmarrā'ī. Baghdad: al-Jumhūrīyah al-'Irāqīyah, Wizārat al-Thaqāfah wa-al-I'lām, Dār al-Rashīd.

Feghali, Habaka J. 1991. *Arabic Adeni textbook (Yemen)*. Wheaton, Maryland: Dunwoody Press.

Gensler, Orin. 2011. Morphological typology of Semitic. *The Semitic languages: An international handbook*, 279–302. Berlin, Boston: De Gruyter Mouton.

Gibson, Maik. 2006. Tunis Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Grand'Henry, Jacques. 1972. *Le parler arabe de Cherchell (Algérie)*. Louvain-la-Neuve: Université Catholique de Louvain, Institut Orientaliste.

Grand'Henry, Jacques. 1976. *Les parlers arabes de la région du Mzāb (sahara algérien)*. Leiden: E. J. Brill.

Grand'Henry, Jacques. 1979. Le parler arabe de la Saoura (sud-ouest algérien). *Arabica* 26(3). 213–228.

Gumperz, John J. & Robert Wilson. 1971. Convergence and creolization: A case from the Indo-Aryan/Dravidian border in India. In Dell Hymes (ed.), *Pidginization and creolization of languages: Proceedings of a conference held at the University of the West Indies, Mona, Jamaica, 1968*, 151–167. Cambridge: Cambridge University Press.

Hansen, Maj-Britt Mosegaard & Jacqueline Visconti. 2009. On the diachrony of "reinforced" negation in French and Italian. In Corinne Rossari, Claudia Ricci & Adriana Spiridon (eds.), *Grammaticalization and pragmatics: Facts, approaches, theoretical issues*, 137–171. Bingley: Emerald Group Publishing Limited.

Harrell, Richard S. 1962. *A short reference grammar of Moroccan Arabic*. Washington, D.C.: Georgetown University Press.

Harrell, Richard S. (ed.). 1966. *A dictionary of Moroccan Arabic: Moroccan-English*. Washington, D.C.: Georgetown University Press.

Haspelmath, Martin. 1994. The growth of affixes in morphological analysis. In Geert Booij & van Marle (eds.), *Yearbook of morphology*, 1–29. Dordrecht: Kluwer.

Haspelmath, Martin. 1997. *From space to time : temporal adverbials in the world's languages*. Munich: LINCOM Europa.

Haspelmath, Martin. 1998. Does grammaticalization need reanalysis? *Studies in Language* 22(2). 271–287.

Haspelmath, Martin. 1999. Why is grammaticalization irreversible? *Linguistics* 37(6). 1043–1068.

Heath, Jeffrey. 2002. *Jewish and Muslim dialects of Moroccan Arabic*. London, New York: Routledge.

Heath, Jeffrey. 2003. *Hassaniya Arabic (Mali): Poetic and ethnographic texts*. Wiesbaden: Harrassowitz Verlag.

Heath, Jeffrey. 2004. *Hassaniya Arabic (Mali)-English-French dictionary*. Wiesbaden: Harrassowitz Verlag.

Heine, Bernd. 1982. *The Nubi language of Kibera - an Arabic creole: grammatical sketch and vocabulary*. Berlin: Reimer.

Heine, Bernd. 2007. Grammaticalization. In Brian Joseph & Richard Janda (eds.), *The handbook of historical linguistics*, 575–601. Oxford: John Wiley & Sons, Ltd.

Heine, Bernd, Ulrike Claudi & Friederike Hünnemeyer. 1991. *Grammaticalization: A conceptual framework*. Chicago: University of Chicago Press.

Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge, New York: Cambridge University Press.

Heine, Bernd & Tania Kuteva. 2003. On contact-induced grammaticalization. *Studies in Language* 27(3). 529–572.

Heine, Bernd & Tania Kuteva. 2005. *Language contact and grammatical change.* New York: Cambridge University Press.

Henkin, Roni. 2006. Negev Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Henkin, Roni. 2016. Functional codeswitching and register in educated Negev Arabic interview style. *Bulletin of the School of Oriental and African Studies, University of London* 79(2). 279–304.

Hillelson, S. 1935. *Sudan Arabic texts with translation and glossary*. Cambridge: Cambridge University Press.

Hinds, Martin & El-Said Badawi. 1986. *A dictionary of Egyptian Arabic: Arabic-English*. Beirut: Librairie du Liban.

Holes, Clive. 2006a. Bahraini Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Holes, Clive. 2006b. Kuwaiti Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Hopper, Paul. 1991. On some principles of grammaticization. In Elizabeth Traugott & Bernd Heine (eds.), *Approaches to grammaticalization*, vol. I, 17–35. Amsterdam, Philadelphia: John Benjamins.

Hopper, Paul & Elizabeth Traugott. 2003. *Grammaticalization*. 2nd ed. Cambridge: Cambridge University Press.

Hourani, Albert. 1991. *A history of the Arab peoples*. New York: Warner Books.

Hudson, R. A. 1999. *Sociolinguistics*. 2nd ed. Cambridge: Cambridge University Press.

Huehnergard, John. 2017. Arabic in its Semitic context. In Ahmad Al-Jallad (ed.), *Arabic in context: Celebrating 400 years of Arabic at Leiden University*, 3–34. Leiden, Boston: Brill.

Huehnergard, John & Aaron D. Rubin. 2011. Phyla and waves: Models of classification of the Semitic languages. *The Semitic languages: An international handbook*, 259–278. Berlin: De Gruyter Mouton.

Ibn Manẓūr, Muḥammad ibn Mukarram. 1988. *Lisān al-ʻArab / lil-Imām ibn Manẓūr ; nassaqahu wa-ʻallaqa ʻalayhi wa-waḍaʻa fahārisahu ʻAlī Shīrī*. (Ed.) ʻAlī Shīrī. 1st ed. Beirut: Dār Iḥyāʾ al-Turāth al-ʻArabī.

Il-Hazmy, Alayan Mohammed. 1975. A critical and comparative study of the spoken dialect of the Ḥarb tribe in Saudi Arabia. Leeds: University of Leeds Doctoral dissertation.

Ingham, Bruce. 1971. Some characteristics of Meccan speech. *Bulletin of the School of Oriental and African Studies, University of London* 34(2). 273–297.

Ingham, Bruce. 1982. Notes on the dialect of the Ḍhafīr of north-eastern Arabia. *Bulletin of the School of Oriental and African Studies, University of London* 45(2). 245–259.

Ingham, Bruce. 1994. *Najdi Arabic: Central Arabian*. Amsterdam, Philadelphia: John Benjamins.

Ingham, Bruce. 2006a. Khuzestan Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Ingham, Bruce. 2006b. Najdi Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Jastrow, Otto. 1973. *Daragözü: eine arabische Mundart der Kozluk-Sason-Gruppe (Südostanatolien): Grammatik und Texte*. Nuremberg: Verlag Hans Carl.

Jastrow, Otto. 1979. Zur arabischen Mundart von Mossul. *Zeitschrift für Arabische Linguistik* 2. 76–99.

Jastrow, Otto. 2006. Anatolian Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Jong, Rudolf de. 2002. Notes on the dialect of the ʻAbābda. In Werner Arnold & Hartmut Bobzin (eds.), *"Sprich doch mit deinen Knechten aramäisch, wir verstehen es!": 60 Beiträge zur Semitistik, Festschrift für Otto Jastrow zum 60. Geburtstag*, 337–360. Wiesbaden: Harrassowitz Verlag.

Jong, Rudolf de. 2006. Sinai Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Jong, Rudolf de. 2011. *A grammar of the Bedouin dialects of central and southern Sinai*. Leiden, Boston: Brill.

Kahle, Paul & Georg Jacob. 1930. *Das orientalische Schattentheater, Vol. 1: Der Leuchturm von Alexandria, ein arabisches Schattenspiel aus dem mittelalterlichen Ägypten*. Stuttgart: W. Kohlhammer.

Kaye, Alan S. & Judith Rosenhouse. 1997. Arabic dialects and Maltese. In Robert Hetzron (ed.), *The Semitic languages*. London, New York: Routledge.

Khan, Geoffrey. 1997. The Arabic dialect of the Karaite Jews of Hīt. *Zeitschrift für Arabische Linguistik* 34. 53–102.

Khan, Geoffrey. 2004. *The Jewish neo-Aramaic dialect of Sulemaniyya and Ḥalabja*. Leiden, Boston: Brill.

Khan, Geoffrey. 2008. *The neo-Aramaic dialect of Barwar*. Leiden, Boston: Brill.

King, Ruth. 2000. *The lexical basis of grammatical borrowing: A Prince Edward Island French case study*. Amsterdam, Philadelphia: John Benjamins.

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Lambton, Ann K. S. 1974. *Persian grammar*. Cambridge: Cambridge University Press.

Lane, Edward W. 1968. *An Arabic-English lexicon*. Beirut: Librairie du Liban.

Law, Daniel. 2013. Inherited similarity and contact-induced change in Mayan languages. *Journal of Language Contact* 6(2). 271–299.

Law, Daniel. 2014. *Language contact, inherited similarity and social difference: The story of linguistic interaction in the Maya Lowlands*. Amsterdam, Philadelphia: John Benjamins.

Le Page, R. B. & Andrée Tabouret-Keller. 1985. *Acts of identity: Creole-based approaches to language and ethnicity*. Cambridge: Cambridge University Press.

Lehmann, Christian. 1985. Grammaticalization: Synchronic variation and diachronic change. *Lingua e Stile* 20. 303–318.

Lehmann, Christian. 2015. *Thoughts on grammaticalization*. Berlin: Language Science Press.

Lentin, Jérôme. 2006. Damascus Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Lewin, Bernhard. 1969. *Notes on Cabali: The Arabic dialect spoken by the Alawis of "Jebel Ansariye."* Stockholm: Almqvist & Wiksell.

Lucas, Christopher. 2010. Negative -š in Palestinian (and Cairene) Arabic: Present and possible past. *Brill's Annual of Afro-Asiatic Languages and Linguistics* 2. 165–201.

Lydon, Ghislaine. 2009. *On trans-Saharan trails: Islamic law, trade networks, and cross-cultural exchange in nineteenth-century western Africa*. New York: Cambridge University Press.

Macdonald, Michael C. A. 2004. Ancient North Arabian. (Ed.) Roger Woodard. *The Cambridge encyclopedia of the world's ancient languages*. Cambridge: Cambridge University Press.

Magidow, Alexander. 2013. Towards a sociohistorical reconstruction of pre-Islamic Arabic dialect diversity. Austin: The University of Texas at Austin Doctoral dissertation.

Magidow, Alexander. 2017. Linguistic history and the history of Arabic: A speech communities approach. In Ahmad Al-Jallad (ed.), *Arabic in context: Celebrating 400 years of Arabic at Leiden University*, 405–440. Leiden, Boston: Brill.

Mahdi, Qasim R. 1985. The spoken Arabic of Baṣra, Iraq: A descriptive study of phonology, morphology and syntax. Exeter: The University of Exeter Doctoral dissertation.

Manfredi, Stefano. 2013. Native and non-native varieties of Arabic in an emerging urban centre of western Sudan: Evidence from Kadugli. In Mena Lafkioui (ed.), *African Arabic: Approaches to dialectology*, 13–50. Berlin, Boston: De Gruyter Mouton.

Mansour, Jacob. 2006. Baghdad Jewish Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Marçais, Philippe. 1956. *Le parler arabe de Djidjelli (Nord constantinois, Algérie)*. Paris: Librairie d'Amérique et d'Orient, Adrien-Maisonneuve.

Marçais, William. 1902. *Le dialecte arabe parlé a Tlemcen: Grammaire, textes et glossaire*. Paris: Ernest Leroux.

Marçais, William. 1908. *Le dialecte arabe des Ūläd Brāhîm de Saïda (Département d'Oran)*. Paris: Honore Champion.

Matras, Yaron. 2009. *Language contact*. Cambridge: Cambridge University Press.

Matras, Yaron. 2011. Grammaticalisation and language contact. In Bernd Heine & Heiko Narrog (eds.), *Handbook of grammaticalisation*, 279–290. Oxford: Oxford University Press.

Matras, Yaron. 2016. 4.3.1 Relative clause marker. (The Dialects of Kurdish). University of Manchester. http://kurdish.humanities.manchester.ac.uk/relative-clause-markers/ (2 March, 2018).

Matras, Yaron & Jeanette Sakel. 2007. Investigating the mechanisms of pattern replication in language convergence. *Studies in Language* 31(4). 829–865.

Meillet, Antoine. 1912. *Linguistique historique et linguistique générale*. Paris: Champion.

Mifsud, Manwel. 2006. Maltese. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Mion, Giuliano. 2014. Eléments de description de l'arabe parlé à Mateur (Tunisie). *al-Andalus Magreb* 21. 57–77.

Murad, Noura. 2000. Ba'd kill hā l-wa't. Paper presented at the Mahrajān al-Qāhira al-dawlī lil-masraḥ al-tajrībī, Cairo.

Naïm, Samia. 2006. Beirut Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Nakshabandi, Anwar Mohammad H. 1998. A descriptive study of the phonology and morphology of the Abha dialect. Washington, D.C.: Georgetown University doctoral dissertation.

Napiorkowska, Lidia. 2015. *A grammar of the Christian Neo-Aramaic dialect of Diyana-Zariwaw*. Leiden, Boston: Brill.

Owens, Jonathan. 1984. *A short reference grammar of Eastern Libyan Arabic*. Wiesbaden: Otto Harrassowitz.

Owens, Jonathan. 1993. *A grammar of Nigerian Arabic*. Wiesbaden: Harrassowitz Verlag.

Owens, Jonathan. 2003. Arabic dialect history and historical linguistic mythology. *Journal of the American Oriental Society* 123(4). 715–740.

Owens, Jonathan. 2006. *A linguistic history of Arabic*. Oxford: Oxford University Press.

Palva, Heikki. 1980. Characteristics of the Arabic dialect of the Bani Ṣaxar tribe. *Orientalia Suecana* 29. 112–139.

Palva, Heikki. 2006. Dialects: Classification. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Pat-El, Na'ama. 2009. The development of Semitic definite article: A syntactic approach. *Journal of Semitic Studies* 54(1). 19–50.

Pat-El, Na'ama. 2017. Digging up archaic features: "Neo-Arabic" and comparative Semitic in the quest for Proto Arabic. In Ahmad Al-Jallad (ed.), *Arabic in context: Celebrating 400 years of Arabic at Leiden University*, 441–475. Leiden, Boston: Brill.

Pereira, Christophe. 2006. Tripoli Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Pereira, Christophe. 2008. Le parler arabe de Tripoli (Libye): phonologie, morphosyntaxe et catégories grammaticales. Paris: INALCO Doctoral dissertation.

Poletto, Cecilia. 2016. Negation. In Adam Ledgeway & Martin Maiden (eds.), *The Oxford guide to the Romance languages*, 833–846. Oxford: Oxford University Press.

Power, Timothy. 2012. *The Red Sea from Byzantium to the Caliphate, AD 500-1000*. Cairo, New York: The American University in Cairo Press.

Procházka, Stephan. 2000. Über einige "rätselhafte" Formen auf -(ē)n in den arabischen Dialekten. *Oriente Moderno* 19 (80)(1). 99–109.

Procházka, Stephan. 2006. Cilician Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Qafisheh, Hamdi. 1977. *A short reference grammar of Gulf Arabic*. Tucson, Arizona: University of Arizona Press.

Qafisheh, Hamdi. 1997. *NTC's Gulf Arabic-English dictionary*. Chicago: NTC Publishing Group.

Rabin, Chaim. 1955. The beginnings of Classical Arabic. *Studia Islamica* 4. 19–37.

Reichmuth, Stefan. 1983. *Der arabische Dialekt der Šukriyya im Ostsudan*. Hildesheim, Zurich, New York: Georg Olms Verlag.

Retsö, Jan. 2004. Relative-clause marking in Arabic dialects: A preliminary survey. In Martine Haak, Rudolf de Jong & Kees Versteegh (eds.), *Approaches to Arabic dialects: A collection of articles presented to Manfred Woidich on the occasion of his sixtieth birthday*, 263–274. Leiden, Boston: Brill.

Robinson, Francis. 1996. *The Cambridge illustrated history of the Islamic world*. Cambridge: Cambridge University Press.

Rosenhouse, Judith. 2006. Jerusalem Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Ross, Malcolm. 2006. Metatypy. (Ed.) Keith Brown. *Encyclopedia of language and linguistics*. Oxford: Elsevier.

Ross, Malcolm. 2007. Calquing and metatypy. *Journal of Language Contact*(1). 1.

Roth, Arlette. 1979. *Esquisse grammaticale du parler arabe d'Abbéché (Tchad)*. Paris: Librairie Orientaliste Paul Geuthner.

Rubin, Aaron D. 2005. *Studies in Semitic grammaticalization*. Winona Lake, Indiana: Eisenbrauns.

Ryding, Karin C. & Kees Versteegh. 2006. ʾIḍāfa. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Saada, Lucienne. 1984. *Éléments de description du parler Arabe de Tozeur (Tunisie): Phonologie, morphologie, syntaxe*. Paris: Librairie Orientaliste Paul Geuthner.

Ṣabbāgh, Mīkhā'īl. 1886. *Mîḫâ'îl Ṣabbâg's Grammatik der arabischen Umgangssprache in Syrien und Aegypten: Nack der Münchener Handschrift*. (Ed.) H. Thorbecke. Strasbourg: Verlag von Karl J. Trübner.

Sabuni, Abdulghafur. 1980. *Laut- und Formenlehre des Arabischen Dialekts von Aleppo*. Frankfurt, Bern, Cirenchester: Verlag Peter D. Lang.

Sánchez, Pablo. 2014. *El árabe vernáculo de Marrakech: Análisis lingüístico de un corpus representativo*. Zaragoza: Prensas de la Universidad de Zaragoza.

Sapir, Edward. 1949. *Language: An introduction to the study of speech*. 2nd ed. San Diego, California: Harcourt Brace Jovanovich.

Schreiber, Giselher. 1970. Der arabische Dialekt von Mekka: Abriss der Grammatik mit Texten und Glossar. Münster: Westfälischen Wilhelms-Universität zu Münster Doctoral dissertation.

Schroepfer, Jason. Forthcoming. Notes on Aswan Arabic.

Shabibi, Maryam. 2006. Contact-induced grammatical changes in Khuzestani Arabic. Manchester: University of Manchester doctoral dissertation.

Siegel, Jeff. 2010. *Second dialect acquisition*. Cambridge: Cambridge University Press.

Silverstein, Michael. 1981. *The limits of awareness*. (Working Papers in Sociolinguistics 84). Austin: Southwest Education Development Laboratory.

Singer, H.-R. 1980. Text aus Azru, Marokko. In Peter Behnstedt, Wolfdietrich Fischer & Otto Jastrow (eds.), *Handbuch der arabischen Dialekte*, 281–285. Wiesbaden: Harrassowitz Verlag.

Souag, Lameen. 2005. Notes on the Algerian Arabic dialect of Dellys. *Estudios de Dialectología Norteafricana y Andalusí* 9. 151–180.

Stewart, Devin. 1998. Clitic reduction in the formation of modal prefixes in the post-classical Arabic dialects and Classicial Arabic "sa-/sawfa." *Arabica* 45(1). 104–128.

Stokes, Phillip. Forthcoming. The plural demonstratives and relatives based on *ʾVl in Arabic and the origin of dialectal illī. *Harrassowitz Verlag*.

Stowasser, Karl & Moukhtar Ani (eds.). 2004. *A dictionary of Syrian Arabic: English-Arabic*. Washington, D.C.: Georgetown University Press.

Szmrecsanyi, Benedikt. 2014. Forests, trees, corpora and dialect grammars. In Benedikt Szmrecsanyi & Bernhard Wälchli (eds.), *Aggregating dialectology, typology, and register analysis: Linguistic variation in text and speech*, 89–112. Berlin, Boston: Walter de Gruyter.

Taine-Cheikh, Catherine. 1998. Deux macro-discriminants de la dialectologie arabe (la réalisation du qâf et des interdentales). *Atériaux Arabes et Sudarabiques (GELLAS)* 9. 11–51.

Taine-Cheikh, Catherine. 2006. Ḥassānīyya Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Talay, Shabo. 1999. *Der arabische Dialekt der Khawētna: I. Grammatik*. Wiesbaden: Harrassowitz Verlag.

Talmoudi, Fathi. 1980. *The Arabic dialect of Sūsa (Tunisia)*. Göteborg: Acta Universitatis Gothoburgensis.

Talmoudi, Fathi. 1981. *Texts in the Arabic dialect of Sūsa (Tunisia): Transcription, translation, notes and glossary*. Göteborg: Acta Universitatis Gothoburgensis.

Tamis, Rianne & Janet Persson. 2013. *Sudanese Arabic-English, English-Sudanese Arabic: A concise dictionary*. Dallas: SIL International Publications.

Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley, Los Angeles, Oxford: University of California Press.

Trudgill, Peter. 1986. *Dialects in contact*. New York: Basil Blackwell.

Tsiapera, Maria. 1969. *A descriptive analysis of Cypriot Maronite Arabic*. The Hague, Paris: Mouton.

Turner, Michael L. Forthcoming. Moroccan Arabic. In John Huehnergard & Na'ama Pat-El (eds.), *The Semitic languages*. London, New York: Routledge.

Turner, Michael L. 2013. Definiteness marking in Moroccan Arabic : Contact, divergence, and semantic change. Austin: The University of Texas at Austin Master's thesis.

Vanhove, Martine. 1993. *La langue maltaise: études syntaxiques d'un dialecte arabe "périphérique."* Wiesbaden: Harrassowitz Verlag.

Versteegh, Kees. 1984. *Pidginization and creolization: The case of Arabic*. Amsterdam, Philadelphia: John Benjamins.

Versteegh, Kees. 2001. *The Arabic language*. 1st ed. Edinburgh: Edinburgh University Press.

Versteegh, Kees (ed.). 2006. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Versteegh, Kees. 2014. *The Arabic language*. 2nd ed. Edinburgh: Edinburgh University Press.

Vicente, Ángeles. 2000. *El Dialecto árabe de Anjra (norte de Marruecos): Estudio lingüistico y textos*. Zaragoza: Universidad de Zaragoza.

Vicente, Ángeles. 2006. Demonstratives. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Vollers, Karl. 1895. *The modern Egyptian dialect of Arabic: A grammar with exercises, reading lessons and glossaries*. (Trans.) Francis Burkitt. Cambridge: Cambridge University Press.

Watkins, Calvert. 1973. Language and its history. *Daedalus* 102(3). 99–111.

Watson, Janet. 1993. *A syntax of Ṣanʿānī Arabic*. Wiesbaden: Harrassowitz Verlag.

Watson, Janet. 2006a. Adverbs. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Watson, Janet. 2006b. Ṣanʿānī Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Webster, Roger. 1991. Notes on the dialect and way of life of the Āl Wahība Bedouin of Oman. *Bulletin of the School of Oriental and African Studies, University of London* 54(3). 473–485.

Weinreich, Uriel. 1953. *Languages in contact: Findings and problems*. New York: Linguistic Circle of New York.

Wilmsen, David. 2014. *Arabic indefinites, interrogatives, and negators: A linguistic history of western dialects*. Oxford: Oxford University Press.

Winford, Donald. 2003. *An introduction to contact linguistics*. Oxford: Blackwell.

Woidich, Manfred. 2006a. *Das Kairenisch-Arabische: Eine Grammatik*. Wiesbaden: Harrassowitz Verlag.

Woidich, Manfred. 2006b. Bʿēri Arabic. (Ed.) Kees Versteegh. *Encyclopedia of Arabic language and linguistics*. Leiden, Boston: Brill.

Wolfram, Walt & Natalie Schilling-Estes. 1998. *American English: Dialects and variation*. Malden: Blackwell.

Woodhead, D. R. & Wayne Beene (eds.). 1967. *A dictionary of Iraqi Arabic: Arabic-English*. Washington, D.C.: Georgetown University Press.

Wright, W. 1898. *A grammar of the Arabic language, translated from the German of Caspari, and edited with numerous additions and corrections*. 3rd ed. Cambridge: Cambridge University Press.

Youssi, Abderrahim. 1992. *Grammaire et lexique de l'arabe marocain moderne*. Casablanca: Wallada.

Zack, Liesbeth. 2009. Egyptian Arabic in the seventeenth century: a study and edition of Yūsuf al-Maġribī's Dafʿ al-iṣr ʿan kalām ahl Miṣr. Amsterdam: University of Amsterdam Doctoral dissertation.

Zeijlstra, Hedde. 2010. On French negation. In I. Kwon, H. Pritchett & J. Spence (eds.), *Proceedings of the 35th annual meeting of the Berkeley Linguistics Society*, 447–458. Berkeley: BLS.