Studies in Arabic Linguistics

Perspectives on Arabic Linguistics XXVI

Edited by Reem Khamis-Dakwar and Karen Froud

2

John Benjamins Publishing Company

Perspectives on Arabic Linguistics XXVI

Studies in Arabic Linguistics

This book series aims to publish original research in all fields of Arabic linguistics, including – but not limited to – theoretical linguistics, historical linguistics, sociolinguistics, pragmatics, typology, and language acquisition. Submissions from all current theoretical frameworks are welcome. Studies may deal with one or more varieties of Arabic, or Arabic in relation to or compared with other languages. Both monographs and thematic collections of research papers will be considered.

The series includes monographs and thematically coherent collective volumes, in English.

For an overview of all books published in this series, please see
http://benjamins.com/catalog/sal

Volume 2

Perspectives on Arabic Linguistics XXVI
Edited by Reem Khamis-Dakwar and Karen Froud

Perspectives on Arabic Linguistics XXVI

Papers from the annual symposia on Arabic Linguistics
New York, 2012

Edited by

Reem Khamis-Dakwar
Adelphi University

Karen Froud
Columbia University - Teachers College

John Benjamins Publishing Company
Amsterdam / Philadelphia

The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Library of Congress Cataloging-in-Publication Data

Annual Symposium on Arabic Linguistics (26th : 2012 : New York City)
Perspectives on Arabic Linguistics XXVI : Papers from the annual symposium on Arabic Linguistics. New York, 2012 / Edited by Reem Khamis-Dakwar and Karen Froud.
p. cm. (Studies in Arabic Linguistics, ISSN 2212-8042 ; v. 2)
Papers were selected from the Arabic Linguistics Society Annual Meeting that was held in New York City in March, 2012.
Includes bibliographical references and index.
1. Arabic language--Congresses. 2. Arabic language--Grammar--Congresses. I. Khamis-Dakwar, Reem, editor. II. Froud, Karen, editor. III. Title.
PJ6303.A56 2014
492.7--dc23 2014023415
ISBN 978 90 272 0030 3 (Hb ; alk. paper)
ISBN 978 90 272 6968 3 (Eb)

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Table of contents

Introduction

Diversity and innovation in Arabic Linguistics

Reem Khamis-Dakwar and Karen Froud

It is an ambitious mission to incorporate studies from different domains of Arabic linguistics in one volume. Since its inception, the *Perspectives on Arabic Linguistics* Series has provided a vehicle for the contributions of Arabic linguists to the study of human languages. This volume highlights the growth of Arabic linguistics and showcases the diversity of approaches that is driving forward this ever-evolving field. Fittingly, these papers were selected from the Arabic Linguistics Society Annual Meeting that was held in New York City in March, 2012 – the most diverse and innovative city in the world, hosting the foremost scholars and thinkers in this most varied and pioneering of fields.

We believe that the study of Arabic linguistics is on the cusp of exciting changes that will bring greater attention, collaboration, and recognition of the contributions made within this field to the broader study of human language. We are proud to be associated with this volume, which represents a collection of unique, varied, and wholly innovative approaches that will contribute to this emerging and evolving dynamic field of study.

The key notions woven throughout this volume are innovation and diversity. The structure of this volume in itself represents an innovation, and we have carefully selected papers that reflect the diversity of approaches, opinions and analyses. Alongside several groundbreaking papers in key aspects of Arabic morphosyntax, semantics, phonology, and sociolinguistics, we are also pleased to include representative contributions from language acquisition and neurolinguistics. The application of these disciplines to the study of Arabic is growing, and the inclusion of papers from these fields is an indication of the importance of innovation for the *Perspectives on Arabic Linguistics*. The juxtaposition of papers from emerging disciplines with those from more established approaches illustrates innovation and diversity across domains; each of the contributions to this volume also demonstrates the kind of within-domain innovation that has brought the study of Arabic linguistics to its present stage.

The volume is structured as follows. This introduction provides an overview and identifies some unifying themes between the papers. The first two papers are contributions from keynote speakers and field leaders, Jamal Ouhalla and Enam Al-Wer. These are followed by the Syntax papers, which represent a wide range of approaches and themes including perspectives from Optimality Theory and generative grammar, tackling issues including negation, prosody, agreement, and cliticization. The contributions from Phonology are next, including an examination of the vexed question of secondary stress in Arabic and a view of the interface between derivation and inflection. The next contribution is from Sociolinguistics and illustrates the application of a new kind of analysis to an examination of contact influences. This is followed by a paper on the Semantics and Pragmatics of terms of endearment and anger, which includes an analysis of these notoriously complex expressions. The penultimate paper in this volume presents some new data from child language acquisition, revealing the complexity and diversity of acquisition of geminates and consonant clusters. Finally, we are pleased to include a paper describing approaches to the neurolinguistics of Arabic, which we hope offers a new direction for the study of Arabic languages at a different level of abstraction.

In the first of our keynote papers, *Lexical change and the nature of lexicon and word derivation*, a narrowly defined linguistic phenomenon is used to focus a discussion that has implications for the very nature of human language. The article examines grammaticalization of motion participles such as Moroccan *ġadi* and Levantine *rayħ* into future tense markers on a par with English *going to* (e.g. *Bill is going to go to college after all*). Ouhalla shows how "semantic bleaching", the process whereby meaning components and arguments are diachronically removed from representations of the motion particle, provides insight into the internal structure of the human grammar. He derives a view of the language system as including *both* a Lexicon, that likely consists of abstract and morphosyntactic universal primitives, *and* a Vocabulary, that is, the repository of language-specific associations between phonological, semantic, and morphosyntactic features. On such a view, Lexical items are only (phonologically) realized to the extent that they spell out bundles of abstract features within a syntactic frame. This innovative work represents an extension of Ouhalla's (2012) proposals and has implications not only for adding to our sparse knowledge of gramaticalization and functional morpheme derivation in Arabic but also for our understanding of the processes and representations involved in human language.

In our second keynote paper, *Yod-dropping in b-imperfect verb forms in Amman*, Enam Al-Wer provides a framework for the discussion of sociolinguistic phenomena through fascinating data on yod dropping from Amman, where the spoken dialect has been examined over the past three generations. Al-Wer's groundbreaking work on dialect formation and contact has already revealed hitherto unsuspected systematicity in vowel shift and has pointed to the importance of adolescents and females in dialect individuation. Through the ongoing 'Amman Project', a large-scale sociolinguistic investigation of the formation of the Amman dialect (see Al-Wer, 2007), we are provided with a unique test case for current theories on dialect formation and contact.

Having set the stage for consideration of linguistic and sociolinguistic phenomena in Arabic with these two papers, each of which emphasizes the broader contributions of Arabic linguistics for the study of human language, we turn to papers on syntax. Hoyt provides data on negative concord, a phenomenon well known in Romance but hitherto little-documented in Arabic. In Levantine, negative concord can be a long-distance dependency, in which an "n-word" inside a subordinate clause can be licensed by negation in a higher clause. Though this long-distance licensing appears at first to be an idiosyncratic property of certain verbs, Hoyt's novel analysis – a multi-dimensional approach that incorporates structural and acoustic analyses alongside a wealth of comparative cross-linguistic data – shows that this phenomenon really derives from constrained interactions between syntactic and prosodic constituency.

Similarly, the paper by Benmamoun and Al-Asbahi presents some unusual data that appear superficially anomalous and shows that the observed pattern can really be derived through a deeper understanding of the underlying processes. They review negation data from San'ani Arabic, a process that depends on cliticization, and contrast this unusual pattern with data from Moroccan Arabic, where negation involves head movement. Through a detailed analysis of the negation structures in various Arabic varieties, they derive an account for San'ani Arabic negation as naturally arising through parameterization. In the spirit of the grammaticalization hypothesis described by Ouhalla (2012 and current volume), Benmamoun and Al-Asbahi also propose an account of cliticization as a possible mechanism for the diachronic emergence of the negative *laysa* in Classical Arabic.

As pointed out by Benmamoun and Al-Asbahi (this volume, p.xx), "Arabic varieties provide fertile grounds for testing current linguistic approaches." Soltan's paper on negation in Cairene Arabic rises to this challenge, making use of Arabic data to evaluate the utility of current formulations for relative positioning of T and Neg heads in the functional domain, deriving a new perspective in the process. Theoretical problems associated with syntactic head movement are well known (for example, the lack of a clear motivation in terms of feature checking, failure to satisfy the extension condition, violation of the condition on chain uniformity, the so-called "traffic rule" problem, and the fact that head movement does not have semantic effects at LF; all discussed in Chomsky, 2001, *inter alia*). With respect to negation, Soltan argues for an account that is morphological rather than strictly syntactic and shows how this perspective could circumvent many problems. Soltan's account places Neg higher than T and splits the Neg domain into two distinct heads that encode semantic and formal negation. Through an examination of structures involving negative disjunction and negative concord, Soltan is able to show that a split-Neg approach permits an economical account of previously disparate phenomena, such as the phenomenon of *-š* spell-out (or lack thereof) in particular syntactic contexts. Like all the best research, this paper asks more questions than it answers. Soltan lays a foundation for much future work on the cross-linguistic implications of a split-domain account for Neg.

Ouali's paper on multiple agreement facts in Arabic is similar in approach to the papers by Soltan and Benmamoun & Al-Asbahi in that it evaluates an existing theoretical framework in the light of Arabic data and derives a more empirically favorable account in the process. Biclausal complex tense sentences that exhibit multiple subject verb agreement in Arabic present a challenge to the Feature Inheritance approach put forward by Chomsky (2008, 2013), because they provide counter-evidence for a view of C as the locus of φ-features. In environments where T must carry φ-features, Chomsky has proposed that these features are *inherited* from C and cannot be inherently specified on T. Ouali shows that there exist multiple agreement contexts in Arabic where no C head is manifest and therefore proposes that T can be lexically specified for φ-features. This proposal is shown to have utility for multiple agreement facts that otherwise resist a parsimonious analysis and also speaks to parameterization distinctions between languages like Arabic (where T can be lexically specified for φ-features) and English (where T inherits φ-features from C during the course of the derivation, prior to establishing an Agree relation with the subject). Again, this proposal has implications for the study of other Arabic varieties as well as other languages.

In another demonstration of the diversity of approaches represented in this volume, the final syntax paper also focuses on the operationalization of Agreement but does so from a very different perspective and to account for a distinct set of cliticization facts. Walkow presents data on clitic restrictions in Classical Arabic and discusses how such restrictions can be derived without recourse to Last Resort interface conditions. Through a re-evaluation of the operation AGREE (as in Nevins, 2007), Walkow shows how Multiple AGREE (that permits one probe to agree with multiple goals) falls short of accounting for restrictions on third person clitics. However, these same facts are readily subsumed under an application of Cyclic AGREE (e.g., Béjar & Řezáč, 2009), whereby a probe may AGREE with one goal when first merged but can access a second or subsequent one only through expansion of the phrase marker via Merge. Similarly to the other Syntax papers in this volume, Walkow's approach permits greater parsimony in accounting for observed superficial variability, showing that alternative structures used when cliticization is blocked are not mechanisms of Last Resort but instead are independently motivated by the PF realization of syntactic relations.

The drive for parsimony and clarity in theoretical frameworks is also much in evidence in the papers on Phonology in this volume. From an Optimality Theory perspective, Aquil discusses the commonly held view that secondary stress is not extant in Arabic, an anomalous but widely held view in the literature. Aquil provides a theoretically motivated account of secondary stress that shows it is readily motivated by interactions between bimoracity and externally motivated high-ranked OT interface constraints. This is a theoretical paper that provides a framework for future empirical work, in contrast to the second of our papers in Phonology, which is an empirical paper that presents data from Lebanese Arabic. Haddad and Wiltshire also take an OT approach, grounding their account in constraints that apply specifically to derivational or inflectional forms (McCarthy, 2005). Their main evidence comes from a case of phonologically unmotivated gemination in a derivational paradigm, which, according to the authors, can only be explained by a requirement that stress be on the same syllable of all the paradigmatically related forms in the derivation. However, the authors show that, with respect to segmental phonology, derivational forms need to respect Base-Output constraints that operate over inflectional forms. This is the reasoning behind their novel approach: in Lebanese Arabic, the derivational output appears to resemble both a Base (i.e., Base-Output constraints apply) and other members of a paradigm (i.e., Optimal Paradigm constraints apply). As before, the drive is for a parsimonious and independently motivated account of an otherwise idiosyncratic set of Arabic data; the approach is once again innovative and offers a way to account for diverse facts cross-linguistically.

The offering from Sociolinguistics for this volume (in addition to the keynote paper by Al-Wer) comes from Miller, who provides a fascinating and in-depth account of data from Bukhara Arabic. Very much in keeping with the theme of innovation, Miller applies Principal Parts Analysis to the verbal system of Bukhara, in the process shedding light on contact effects between Semitic, Indo-Aryan, and Turkic varieties. The analysis reveals effects of concatenative contact on non-concatenative verbal derivative systems. Miller provides details of her analysis to an unusually generous extent, with the express aim of facilitating further application of this novel analysis to other varieties of Arabic.

Such methodological innovation is a unifying factor for the final three papers in the volume, which constitute offerings from (broadly construed) Semantics and Pragmatics, Language Acquisition, and Neurolinguistics. The first of these, by Mohammad Mohammad, is in part a response to a challenge laid down by Ferguson (1997), who claimed that the constructions under scrutiny (so-called

"God-wishes") would take years to analyze. Mohammad provides a comprehensive description of God-wishes, including data from observed social exchanges and from Eid television shows. Again exemplifying the diversity of approaches that are instantiated here, Mohammad expands the initial description of the use of God-wishes in prescribed social interactions, including a description of their pragmatic uses and interpretations, and proposes a grammatical framework for this complex, socially-determined phenomenon.

Ragheb and Davis provide data on the L1 development of final consonant clusters in Cairene Arabic, and in doing so they tackle an understudied aspect of Arabic phonological acquisition. Based on data collected from two typically developing siblings at the age of 2 years 8 months of age, these authors reveal that final consonant clusters are acquired relatively early. Errors in final consonant cluster production follow interesting patterns that shed light on the underlying knowledge of these young children about the phonology of their L1. Consonant deletion (cluster simplification) is very infrequent, for instance, since the children instead opt to lengthen or geminate one of the elements in the cluster as a strategy to maintain the bimoraic structure. The data also reveal patterns of consonant preservation in cluster production that respect sonority distinctions: lower sonority consonants are more likely to be geminated, and falling sonority clusters are acquired sooner than rising sonority clusters. Hence, this study is innovative in terms of both the data provided and in the way it addresses the distinction between cluster reduction and gemination.

The final paper in this volume, on neurocognitive approaches to the study of diglossia, was written by ourselves, the volume co-editors. It was written with the intention that Arabic linguistics continues to move forward in terms of methodological innovation. By pointing out the possibilities for linguistic investigations that are emerging as neurocognitive methodologies become more and more accessible, and by providing examples of the applications for neuroscience in examining language representation and processing in Arabic diglossia, we hope to catalyze interest in neurocognitive approaches and lead the way for their incorporation in Arabic linguistics. The emphasis on converging evidence from linguistics and neuroscience in recent years has been proving more productive, and in our paper we suggest a preliminary mental model of the cognitive inter-relationships between lexical and syntactic components associated with two language varieties in Arabic diglossia. It is our hope that this kind of approach, together with a new source of evidence about language representation and processing, will generate novel approaches, more testable hypotheses, and yet more questions in need of answers.

Diversity and innovation are infused throughout the work presented here. The many studies included in this volume offer innovative data, provide details of innovative analyses, address innovative questions, and apply innovative ideas to the field of Arabic linguistic studies. However, there is one more unifying theme that we wish to emphasize: all the contributors to the volume, and to the annual meeting of the Arabic Linguistics Society, are also united in the hope that the work on Arabic linguistics will catalyze future research, as well as drawing further attention to the potential contribution of Arabic linguistics for the broader study of human language. We have shown the utility of our work in providing test cases for specific theoretical approaches, for refining existing theories, and for laying the groundwork for hypothesis-formation and empirical testing. We have each been able to contribute, in different ways, to the sparse knowledge of language representation and processing in Arabic. No discussion of Arabic language varieties can ignore the variation, dialectal or diglossic, that is inherent; in itself this is a contribution, since linguistic variation is often not a topic of direct investigation in the traditional approaches to the study of human language. We argue that this inclusive approach, whereby diversity in Arabic dialects (and between the two language varieties of Arabic) forms part of the object of study, can ultimately inform our understanding of innate linguistic knowledge. Diversity and innovation are crucial for human languages, and Arabic linguistics leads the way in this respect.

References

Béjar, S., & Řezáč, M. (2009). Cyclic agree. *Linguistic Inquiry*, 40(1), 35–73. DOI: 10.1162/ling.2009.40.1.35

Chomsky, N. (2001). Derivation by phase. In M. Kenstowicz (Ed.), *Ken Hale: A life in language* (pp. 1–52). Cambridge, MA: MIT Press.

Chomsky, N. (2008). On phases. In R. Freidin, C. Otero, & M. L. Zubizarreta (Eds.), *Foundational issues in linguistic theory* (pp. 133–166). Cambridge, MA: MIT Press.

Chomsky, N. (2013). Problems of projection. *Lingua*, 18, 1–35.

Ferguson, C. (1997). God-wishes in Syrian Arabic. In R. K. Belnap, & N. Haeri (Eds.), *Structuralist studies in Arabic linguistics: Charles Ferguson papers, 1954–1994* (pp. 212–228). La Haye/New York/Cologne: Brill.

McCarthy, J. (2005). Optimal paradigms. In L. Downing, T. A. Hall, & R. Raffelsiefen (Eds.), *Paradigms in phonological theory* (pp. 295–371). Oxford: Oxford University Press.

Nevins, A. (2007). The representation of third person and its consequences for person-case effects. *Natural Language & Linguistic Theory*, 25, 273–313. DOI: 10.1007/s11049-006-9017-2

Ouhalla, J. (2012). Lexical change and the architecture of the lexicon. In E. Torrego (Ed.), *Of grammar, words and verses* (pp. 41–66). New York and Amsterdam: John Benjamins. DOI: 10.1075/lfab.8.06ouh

Acknowledgments

This volume was a major undertaking, and we would like to acknowledge the help and support of many people whose expertise and support made this publication possible. The first thanks goes to the authors, experts and innovators all, who without exception have been cooperative, collegial, and deeply understanding throughout the peer review and manuscript preparation process. The support and advice of the Series Editors, Abbas Benmamoun and Enam Al-Wer, has been invaluable, and we thank them deeply. We also owe an enormous thanks to the very many anonymous reviewers who graciously shared their unique expertise on each of the papers. Their invaluable reviews and suggestions contributed greatly to the final form of the papers that are presented here.

The production of *Perspectives on Arabic Linguistics* every year owes an enormous debt to the editors and publishers at John Benjamins for their support and commitment. We recognize and appreciate their commitment to the dissemination of our work to the linguistic community, and to the promotion of Arabic linguistics within the broader field of theoretical linguistics.

Special thanks and recognition are due to all the members of the NeuroCognition of Language Lab at Teachers College, Columbia University. Not only did they make the 2012 annual meeting of the ALS a huge success, but they also went far above and beyond the call of duty in their subsequent support of the co-editors. They are our village. Special thanks to David Stone for his professional assistance with the final edits.

Lastly, a special thanks to our family members, especially our children: Sam Froud, Fouad Dakwar and Nadey Dakwar. The joy they bring to our lives keeps us going, and keeps everything – even Arabic linguistics – in perspective.

The development of future participles and future tense markers from motion predicates

Semantic, morphosyntactic and structural reduction

Jamal Ouhalla
University College Dublin

In this paper I outline an analysis for the development of the Moroccan Arabic motion participles *ġadi* 'going' and *maši* 'going' into future participles and future tense markers, and explore the implications of this process for the nature of the lexicon and word-derivation. The analysis is framed in the context of the approach to grammaticalization outlined in Ouhalla (2012a & b), which assumes that words are derivational constructs with no existence outside the sentence-contexts that include them. This approach entails a principled distinction between grammaticalization and lexical change: the former targets selected word forms such as the participle in sentence-contexts, while the latter targets roots in isolation and is reflected in all the words based on the root. I will argue for a distinction between Lexicon, that may consist of abstract semantic and morphosyntactic primitives that are uniform across languages, and vocabulary, consonantal roots and functional morphemes with phonological features. Instances of grammaticalization of the type discussed here affect vocabulary rather than Lexicon.

Keywords: grammaticalization, motion participles, future tense participles, Moroccan Arabic

1. Introduction

This article outlines an analysis for the development of the Moroccan Arabic motion participles *ġadi* 'going' and *maši* 'going' into future participles and future tense markers. This development parallels the development of the motion participle *rayḥ* 'going' into a future participle and a future tense marker in some Eastern dialects of Arabic, including Palestinian (Rice and Sa'id, 1960). The analysis is framed in the context of the approach to grammaticalization outlined in

Ouhalla (2012a & b), which assumes a root-based Lexicon and the concomitant outcome that words are derivational constructs with no existence outside the sentence-contexts that include them. Among other things, the approach makes a principled distinction between lexical change, which targets roots in isolation and is reflected in all the words based on the root, and grammaticalization, which targets selected word forms such as the participle in sentence-contexts. It also offers a principled account of the transformation of word forms such as motion participles into future participles and future tense markers in terms of the common morphological processes of Feature Copying and Feature Deletion. The main focus in this article is on the implications of the deletion of lexical semantic and morphosyntactic features for sentence structure and derivation, including structure-reduction.

The article is organised as follows. Section 2 presents the basic data, situates it in a broader context and isolates the characteristic properties of the change from a motion participle to a future participle to a future tense marker. Section 3 outlines the lexical and morphosyntactic properties of Arabic participles in the context of the framework which assumes that the Arabic Lexicon is root-based and words are derivational constructs. Section 4 provides a summary of the approach to grammaticalization adopted and shows how it applies to the instances discussed here. Section 5 shows how the deletion of lexical semantic features (semantic bleaching) in the development of future participle forms turns the future participle into a semantically impoverished raising predicate. Section 6 shows how the deletion of morphosyntactic features in the development of the future tense forms results in structure-reduction from a bi-clausal to a mono-clausal sentence-structure.

2. Motion participle > future participle > future tense

Examples (1a–c) are from Moroccan Arabic. (1a) shows the use of the participle form *ġadi* 'going' as the directed motion predicate of a sentence. As is generally the case with participles, *ġadi* is specified for number and gender and agrees in these features with the subject of the sentence, and, when present, the auxiliary as well. (1b) shows the use of *ġadi* as a marker of future tense that co-occurs with a directed motion verb in the imperfective form. The form lacks the directed motion meaning GO, but retains its participle status, including its agreement features and the agreement relation with the subject and the auxiliary. (1c) shows the use of *ġadi* as a future tense marker that co-occurs with a directed motion verb in the imperfective form. In this particular use, the form lacks not only the directed motion meaning GO, but also the number-gender features characteristic of participles in Arabic. Moreover, it is often reduced to a smaller form consisting of an open syllable typical of functional vocabulary in the language. All three forms are reported and analysed in Benmamoun (2000).[1]

(1) a. l-bnat (kan-u) ġadi[y]-in (l-s-sinima).
the-girls (were.3.PL) going.PL (to-the-cinema)
'The girls are/were going (to the cinema).'

b. l-bnat (kan-u) ġadi[y]-in imši-w (l-s-sinima).
the-girls (were-3.PL) going.PL go-3.PL (to-the-cinema)
'The girls are/were going to go (to the cinema).'

c. l-bnat ġa(di) imši-w (l-s-sinima).
the-girls will go-3.PL (to-the-cinema)
'The girls will go (to the cinema).'

A similar situation involving the participle *ra[y]ḥ* 'going' is found in some Eastern dialects of Arabic, including Palestinian (Rice and Sa'id, 1960). (2a) shows *rayḥ* in its function as a directed motion predicate of a sentence, inflected for number and gender. (2b) shows *ra[y]ḥ* in its function as a future tense participle, missing the directed motion meaning GO but not the number-gender inflection. (2c) shows *rayḥ* in its function as a future tense marker, missing both the directed motion meaning GO and the number-gender inflection. (2b&c) are cited in Rice and Sa'id (1960) and (2a) was kindly provided by Reem Khamis-Dakwar (personal communication).[2]

(2) a. il-bana:t (ka:n-u) rayḥi:n (ʢa-s-si:nama).
the-girls (were.3.PL) going.PL.M. (to-the-cinema)
'The girls are/were going (to the cinema).'

b. ʔaddēš rāyḥ-īn tibʔu hōn?
how.much go.ACT.PTCP.PL stay.2PL here
'How long are you (plural) going to stay here?'

1. A reviewer has pointed out that the future participle form in (1b) should perhaps be more accurately described as denoting prospective future in the sense of Comrie (1976: 64–67) on a par with English *going to*. I take this point on board and show below that prospective future, understood as future relative to sentence tense (as opposed to utterance time), receives a formal and transparent explanation in the context of the analysis outlined here.

2. The Palestinian Arabic examples (2b&c) include a stative predicate, which contrasts with the motion, activity meaning of the source participle of *ra[y]ḥ* shown in (2a). A similar example from Moroccan Arabic is given in (i).

(i) l-bnat ġadi ibqa-/iwqf-/istənna-w təmma.
the-girls will stay-/stand-/wait-3.PL there
'The girls will stay/stand/wait there.'

c. ʔaddēš rāḥ tibʔa hōn?
how.much raḥ stay.2.M.S here
'How long are you (singular) going to stay here?'

Assuming that the future participle form and the future tense form initially developed from the motion participle form, it is an interesting question if they did so independently of each other or one form developed first and then served as input to the other. According to the widely assumed cline repeated in (3), the future participle form is supposed to develop from the motion participle form first and then serve as input to the future tense form. If so, the more specific cline in (4) can be thought of as an instance of this general cline. It describes a path of change whereby a content word (the motion participle) develops into a grammatical, function word (future participle) and later into a grammatical, function morpheme (future tense). Languages may differ as to which stage in the cline the relevant change is at. For example, English *going to* (e.g. *Bill is going to go to college*) is at the future participle stage, and the French forms based on *aller* 'go' (e.g. *Les filles vont/allaient aller au cinéma*) at the auxiliary stage (Hopper and Closs Traugott, 2003).

(3) content item > grammatical word > clitic > inflectional affix
(Hopper and Closs Traugott, 2003, p. 7)

(4) directed motion participle > future participle > future tense

The question whether the future tense form can develop directly from the motion participle form without going through the future participle form arises in situations where a given dialect/language has the future tense form but not the future participle form. One such situation is reported in Ouhalla (2012a & b) to exist in the dialect of Arabic spoken in the northwestern region of Morocco overlooking the Strait of Gibraltar. This dialect has grammaticalized the motion participle form *maši* 'going' into a future tense marker, but lacks the corresponding future participle form. It is conceivable that the future participle form existed in the past and later dropped out, in which case this situation does not necessarily amount to a counterexample to the cline shown in (4) (but see Ouhalla 2012b).[3]

3. Example (i) is a tongue twister from the northwestern dialect of Moroccan Arabic, which arguably casts doubt on the possibility that homophony and haplology serve as constraints on the development of functional vocabulary. The example includes the sentence negation markers ma and ši in addition to the future tense marker maši and the imperfective form of the motion verb mši.

(i) ma maši ši i-mši bwaḥdu.
NEG will NEG 3.M.S-go alone
'He will not go alone.'

(5) a. l-bnat (kan-u) maši[y]-in (n-s-sinima).
the-girls (were-3.PL) going.3.PL (to-the-cinema)
'The girls are/were going (to the cinema).'

b. l-bnat maši ymši-w (n-s-sinima).
the-girls will go-3.PL (to-the-cinema)
'The girls will go to the cinema.'

However the path in (4) is interpreted, it is possible to demonstrate that the two stages it shows are a function of changes that affect distinct features, which, moreover, are distributed over different constituents of the participle form. The first stage, which yields the future participle form, involves a change in the lexical semantic features of the root and amounts to the loss of the motion-meaning component GO and the arguments it selects (semantic bleaching) (6a&b).

(6) Future participle (semantic bleaching)
a. Loss of the motion meaning component GO.
b. Loss of the arguments selected by GO.

The second stage, which yields the future tense form, involves change in morphosyntactic features associated with the participle morpheme, and amounts to the loss of the properties characteristic of participles, including number-gender features (morphological reduction). In the case of *ġa(di)*, this stage also involves an optional change in phonological features, which amounts to the optional loss of the second syllable of the participle form. To these should be added a further, less visible change, which can be described as structure-reduction. As will be shown below, while the future participle form in (1b) corresponds to a clause in a bi-clausal sentence structure, the future tense form in (1c) and (5b) corresponds to a Tense node in a mono-clausal sentence structure. These changes are listed (7a–d) and discussed in more detail in Sections 4–6 below.

(7) Future tense (morphosyntactic, structural and phonological reduction)
a. Loss of number-gender features
b. Loss of participle features (category change)
c. Loss of structure (bi-clausal > mono-clausal)
d. Loss of phonological features

A proper understanding of how motion participles develop into markers of future tense requires prior understanding of the semantic and morphosyntactic properties of participles, including their derivation. This task is carried out in the next section.

The morphosyntax of participles

The analysis of Arabic participles assumed here is the one outlined in Ouhalla (2011, 2012a), which builds on previous work on Arabic participles, including Benmamoun (2000), Fassi Fehri (1993) and Shlonsky (1997). The analysis is based on data from Standard Arabic involving various lexical categories, but it generalises to parallel data from the spoken dialects such as the Moroccan Arabic paradigm in (8a–e). The latter shows the basic paradigm based on the root *√MŠ*, which has the directed motion meaning GO in Moroccan Arabic rather than the original manner of motion meaning WALK it (still) has in Standard Arabic (Ouhalla 2012a & b).[4]

(8) a. maši (active participle)
 b. mši (1st/2nd person perfective verb)
 c. mša (3rd person perfective verb)
 d. mši (imperfective verb)
 e. mši (imperative verb)
 f. mši (derived nominal)

The verb and derived nominal forms have essentially the same template and, moreover, share essentially the same melody, apart from the third person perfective verb form, the vocalic melody of which shows a 3rd versus 1st/2nd person distinction. However, the participle form has both a distinct templatic form and a distinct melody, thereby excluding the possibility that it derives from a common word (word-based derivation). Rather, the participle form is the product of a root-and-pattern derivation (McCarthy, 1979, 1981). Its constituent structure includes the root *√MŠ* with the lexical meaning shown in (9) (event structure) characteristic

4. This conclusion is based on, among other things, the fact that (verb) forms based on *√MŠ* are compatible with manner of motion PPs such as 'by car' in Moroccan Arabic, as shown in (i). Note that genuine manner of motion verbs such as English *walk* are incompatible with such PPs (e.g. *They walked (!by car)*).

(i) mša-w f-ṭ-ṭomobil.
 went-3.PL in-the-car
 'They went by car.'

In the northwestern dialect, the change in meaning undergone by *√MŠ* (*√MŠ* 'walk' > *√MŠ* 'go') exists in addition to the change undergone by the participle form based on this root into a future tense marker shown in (5a & b) (*maši* 'going' > *maši* 'future tense'). The combination of the two changes yields a common path of change widely discussed in the literature in connection to Latin and French and often expressed as in (ii). These changes and their implications for the issue of lexical change versus grammaticalization as well as the debate relating to the nature of the Arabic Lexicon are discussed in Ouhalla (2012a).

(ii) Latin *ambulare* 'walk' > French *aller* 'go' > future auxiliary
(Hopper and Closs Traugott, 2003, p. 102)

of directed motion predicates (Jackendoff, 1985, 2010; Levin & Rappaport Hovav, 1992; Rappaport Hovav & Levin, 2010). The event structure includes the primitive motion predicate MOVE, its external argument *x* and a path component with the distal value interpreted relative to a reference point understood in discourse such as the speaker's location. The distal-proximal values of the path component yield the distinction between roots with the meaning GO ([DISTAL] = 'move away from the speaker') and those with the meaning COME ([PROXIMAL] = 'move towards the speaker').

(9) √*MŠ*: [x move [path: [distal]]

In addition, the participle form includes the component *v*(oice), which encodes the active-passive distinction in terms of the feature [act(ive)] and [pass(ive)] (Fassi Fehri, 1993). This component bears an intimate selectional relation to the root, the nature of which determines the voice value of *v*. While some roots are compatible with both the active and passive values, meaning they can feed both an active and a passive participle form, others are only compatible with the active value, meaning they only correspond to an active participle form. The latter type includes the directed motion root √*MŠ*. The component *v*[act/pass] is shared with verb forms and derived nominals, and is in part intended to capture the widely known fact that participles and derived nominals incorporate a verb inside them (have internal verbal properties) (Fassi Fehri, 1993). Participles include a third component identified in Ouhalla (2011) as neutral aspect, meaning an aspect constituent that does not encode the perfective/imperfective distinction in view of the fact that Arabic participles do not show such a distinction. This constituent is also shared with verbs, except that in verbs it has the perfective-imperfective values. For current purposes, this constituent will be called simply Part[agr], and assumed to be the locus of the number-gender features characteristic of participles. Accordingly, the motion participle in (8a) has the constituent structure shown in (10), where the event structure of the directed motion root is represented as GO for brevity.

(10) [Part[agr] [*v*[act] [√[GO]]]]

Each constituent of the participle form projects as an independent category in syntactic structure. In contexts where the participle is the predicate of the sentence such as (5a), the phrase structure of the participle functions as the complement of finite T along the lines shown in (11a). The derivational steps that take place in Syntax (as opposed to PF, see below) are as outlined in (11b–d). These include an Agree relation between Part[agr] and the subject, which results in the valuation of the number-gender features of Part[agr] based on those of the subject (participle-subject agreement). This agreement relation does not result in the valuation of the Case feature of the subject so that the subject remains active and therefore free to enter a second Agree relation with T[agr]. The latter, which includes the feature [person], results in the valuation of the agreement features of T[agr] and the Case feature of the subject to nominative (auxiliary-subject agreement). Whether agreement with T/aux is what is responsible for raising of the subject to Spec, Part[agr] is unclear. For current purposes, it will be assumed that raising of the subject is motivated by an obligatory occurrence of EPP associated with Part[agr] and an optional occurrence associated with T[agr]. (11d) shows the derivation of the participle form by Head-movement of the root to *v* and of the complex head [[√]-*v*] to Part[agr]. The auxiliary form, when present (in past tense sentences), is inserted directly under T[agr].

(11) a. Initial structure

$[_{TP}$ T[agr] $[_{PartP}$ Part[agr] $[_{vP}$ $[_{DP}$ the-girls] $[_{v'}$ *v*[act] $[_{\surd P}$ √[GO]...

b. Participle-subject agreement

$[_{TP}$ T[agr] $[_{PartP}$ $[_{DP}$ the-girls] $[_{Part'}$ Part[F.PL] $[_{vP}$ ~~[DP the-girls]~~ $[_{v'}$ *v*[act] $[_{\surd P}$ √[GO]...

c. Auxiliary-subject agreement

$[_{TP}$ T[3.F.PL] $[_{PartP}$ $[_{DP}$ the-girls] $[_{Part'}$ Part[F.PL] $[_{vP}$ ~~[DP the-girls]~~ $[_{v'}$ *v*[act] $[_{\surd P}$ √[GO]...

d. Head-raising

$[_{TP}$ T[3.F.PL] $[_{PartP}$ [[[√[GO]]-[*v*[act]]]-Part[F.PL]]$[_{vP}$ ... $[_{v'}$ ... $[_{\surd P}$ ...

The output of the syntactic derivation in (11a–d) is subsequently submitted to Lexical Insertion and morphological derivation at PF. Lexical Insertion is preceded by housekeeping morphological operations that include fusion of *v*[act] and Part[agr] into a single node along lines shown in (12a). They also include fission of Agr into a separate node of its own (Embick, 2000) along the lines shown in (12b). Lexical Insertion then inserts the appropriate root under √, the appropriate morpheme (vocalic melody) under the fused *v*-Part node, and the appropriate agreement morpheme under the Agr node (12c). The output is then submitted to root-and-pattern derivation, which maps the consonantal root and the co-occurring morphemes onto the relevant template, deriving a linear form with no internal morphological structure (12d).[5]

5. The presentation here assumes the initial proposal in Halle and Marantz (1993) that roots lack phonological features in the Lexicon and are therefore subject to Late Insertion in terms of root exponents listed in Vocabulary. A different view is proposed in Embick (2000) and Embick and Noyer (2007) whereby roots have phonological features in the Lexicon and therefore are not subject to Late Insertion.

(12) a. Fusion of *v*[act] and Part[agr]
...[√[GO]-*v*[act]]-Part[agr]]... → ...[√[GO]]-*v*[act]/Part[agr]]...

b. Fission of Agr
...[√[GO]-*v*[act]/Part[agr]]... → ...[√[GO]-[*v*/Part[act]-Agr[numb,gend]..

c. Lexical Insertion
...[√[GO]-*v*/Part[act]-Agr]... → [√[GO]:*MŠ*-*v*/Part:[*a-i*] Agr:[*in]*]...

d. Root-and-pattern derivation and linearization
[√[GO]:*MŠ*-*v*/Part:[*a-i*]-Agr:[*in*]] → [*maši[y]in*]

According to the analysis outlined in (11a–d) and (12a–d), the lexical features of words are distributed over their constituent parts in line with Distributed Morphology (Halle and Marantz, 1993) and the model it assumes (13). The lexical semantic features are supplied by roots. The morphosyntactic features are supplied by the functional categories that co-occur with roots responsible for parts-of-speech categorisation, *v* and Part in the case of participles. The phonological features are supplied by the exponents (vocabulary items) of each of these constituents.

(13)
$$\begin{tikzcd} \text{Vocabulary} \arrow[r] & \text{PF/Morphology} & & \text{LF/Meaning} \\ \text{Lexicon} \arrow[r] & & \text{Syntax} \arrow[ul] \arrow[ur] & \end{tikzcd}$$

A significant consequence of this arrangement, relevant to the present concerns, is that roots lack morphosyntactic features, which are a property of words, more precisely, the functional morphemes they include. Another equally significant consequence is that the participle form (8a) and indeed all the other forms listed in (8a–f) are essentially derivational constructs that have no existence outside the sentences that include them. They are not listed in the Lexicon nor can they be said to be derived in isolation and then inserted under appropriate nodes in syntactic structures. This arrangement raises obvious questions for the type of change discussed here (grammaticalization), including the question how precisely it can target word forms such as the participle *maši* 'going' in (8a) that are not listed. This and related questions are addressed in the next section.

For the moment, it is useful to clarify that in the present context, the expression 'the Arabic Lexicon is root-based' should actually read 'the Arabic Vocabulary is root-based' or, more precisely, 'the Arabic Vocabulary includes roots that are purely consonantal.' The Lexicon in (13) is likely to consist of abstract semantic and morphosyntactic primitives that are uniform across languages, combinations, bundles of which correspond to specific predicates and functional categories. It is the Vocabulary component in (13) that includes consonantal roots and functional morphemes with phonological features in addition to semantic and morphosyntactic features. Moreover, instances of grammaticalization of the type discussed here affect Vocabulary rather than Lexicon by adding to it new items that correspond to (i.e. can spell out) appropriate bundles of abstract features in syntactic structures that previously may have been null.

Grammaticalization as feature-copying and feature-deletion

The approach to grammaticalization adopted here is based on certain key properties identified in Ouhalla (2012a & b), four of which are listed in (14a–d). First, grammaticalization applies to words in sentence contexts rather than to individual lexical items. This is more evident in instances where it targets a sequence of two contiguous words situated in different clauses such as GO(ING)+TO in English (Hopper and Closs Traugott 2003) and Berber (Ouhalla 2012b) and WANT + SUBJUNCTIVE AUX in Greek (Roberts and Roussou, 2003). Secondly, grammaticalization can be highly selective with respect to the words it targets such that only the participle form of an otherwise complete paradigm is selected, as is the case in Arabic and English. Thirdly, grammaticalization involves the creation of new functional vocabulary (partially) homophonous with the source form such as the future participle *ġadi* and the future tense markers *ġa(di)* and *maši*. Fourthly, grammaticalization is largely reductionist in nature as indicated by commonly used descriptive terminology such as 'loss' and 'reduction' of features. The first two properties are essentially a function of the organization of grammar as outlined above, in particular the conclusion that lexical words have no existence outside sentence contexts. The third and further properties indicate the involvement of the morphological processes of Feature Copying and Feature Deletion.

(14) Key properties of grammaticalization
   a. It targets words in a sentence-context.
   b. It is highly selective with the respect to the words it targets.

c. It results in the creation of new functional vocabulary (partially) homophonous with the source form.
d. It is generally reductionist in nature (feature-loss/reduction).[6]

To see how the approach works at a formal level, we will use the instance of grammaticalization that involves the participle *ġadi* for two main reasons. One is that it shows both stages of change, that is, the future participle stage and the future tense stage. The other is that it has a cognate that expresses future tense, namely the future adverbial *ġədda* 'tomorrow' and its counterparts in other dialects, including the nominal *l-ġad* 'the-morrow' in Standard Arabic. As pointed out in Ouhalla (2012a & b), the very existence of this particular nominal-adverbial form inevitably points to the root *√ĠD* shared by both words as the source of the component responsible for the future reference. Since roots only have semantic features, it follows that the component in question resides in the event structure of motion roots. The representation of the motion participle *ġadi* in a sentence context such as (1a) is shown in (15), modelled on the representation of *maši* above. The representation is supposed to be the one obtained subsequent to Lexical Insertion and root-and-pattern derivation and takes the form of bundles of semantic, morphosyntactic and phonological features. The lexical semantic features are summarised as GO and the phonological features are represented as *ġadi*.

(15) $[_{\text{TP}}$ 'the-girls' $[_{\text{T'}}$ T[agr] $[_{\text{Part}}$ [*ġadi*]:[[GO],[*v*[act]],[Part[agr]]]...]

The first step in grammaticalization involves a process that extracts the participle form from its sentence-context, which is to all intents and purposes similar to the process of root-extraction out of word-contexts assumed in the acquisition of the Lexicon of Semitic languages with root-and-pattern derivation (Berman, 2003). The process can be understood to take the form of the common and independently motivated morphological operation of Feature Copying, which by virtue of targeting all the features that make up the participle amounts to a process of word-copying. Applied to *ġadi* in a sentence context such as (15), it results in the creation of a new and homophonous vocabulary item along the lines roughly shown in (16), ignoring the phonological features. The word status of the extracted motion

6. This is also true of instances of grammaticalization that involve a combination of two words such as GO(ING)+TO in English (Hopper and Closs Traugott, 2003) and Berber (Ouhalla, 2012b) and WANT+SUBJUNCTIVE AUX in Greek (Roberts and Roussou, 2003). These instances too involve substantial reduction in the feature-content of the verb/root-component of the complex (Ouhalla, 2012b).

participle form *ġadi* is shown with the symbol ω and its representation to the right of the arrow shows the details of its event structure supplied by GO.[7]

(16) Feature Copying (word-extraction out of a sentence-context)

$[_{TP}$ 'the-girls' $[_{T'}$ T[agr] $[_{Part}$ [*ġadi*]:[[GO],[*v*[act]],[Part[agr]]]...] →

$[_{\omega}$ [*ġadi*]:[[x MOVE [DISTAL]],[*v*[act]],[Part[agr]]]

The next step takes the form of the equally common morphological process of Feature Deletion. The making of the future participle form, which, recall, lacks the directed motion meaning but retains the participle status, involves selective deletion of the motion predicate MOVE and its external argument (17). The path component of the event structure with the distal value remains and is (re)interpreted relative to utterance time as opposed to utterance place. This particular aspect of the analysis is consistent with the view that motion predicates in general have a spatio-temporal dimension, component the value of which is determined by various considerations, including context (Kratzer, 1995). It is this component that enables words based on motion predicates such as GO to develop into future tense markers of various sorts. In addition to the path component of the event structure supplied by the root, the morphosyntactic features of the participle form supplied by *v*[act] and Part[agr] also remain.

(17) *ġadi* 'motion participle' > *ġadi* 'future participle'

$[_{\omega}$ [*ġadi*]:[[x MOVE [DISTAL]], [*v*[act]], [Part[agr]]] →

$[_{\omega}$ [*ġadi*]:[[~~x MOVE~~ [DISTAL]], [*v*[act]], [Part[agr]]]

The making of the future tense form involves further deletion, this time of the morphosyntactic features associated with *v*[act] and Part[agr] and, optionally, the phonological features corresponding to the second syllable. The source form can in principle be either the directed motion participle itself or the new future participle as the mechanisms of Feature Copying and Feature Deletion do not favor one or the other. The difference is the extent of deletion involved, and ultimately boils down to the issue whether change is necessarily always gradual, as is widely

7. Agreement raises difficulties regarding Feature Copying, though not unique to the analysis here. According to Chomsky (1995), the uninterpretable agreement features of functional heads such as T[agr] and Part[agr] delete subsequent to evaluation under Agree in Syntax. With this mind, Embick (2000) and Embick and Noyer (2007) suggest that a whole new Agr node is created at the level of morphological analysis and its content obtained by Feature Copying from the agreeing subject. What is required here is a scenario whereby the agreement features of Part[agr] (participles) remain, but in an unvalued form. This would be possible in the context of an analysis where deletion erases the values of agreement features but not necessarily the agreement features themselves.

assumed in the Gramamticalization literature (Hopper and Closs Traugott, 2003), or can be catastrophic, as argued in Lightfoot (1999). The former view implies the gradual scenario motion participle > future participle > future tense, while the latter allows for the catastrophic scenario motion participle > future tense. For the purposes of presentation here, we will assume that the future tense form develops from the future participle form. The process, initially, involves creating a copy of the future participle form by Feature Copying, which accounts for the existence of the future tense form as a new and additional vocabulary item. This is then followed by deletion of all the morphosyntactic features that define participles and, optionally, the phonological features associated with the second syllable (18). The output is a vocabulary item that is specified for the feature [DISTAL] and which can correspond to the node T[FUTURE] in a sentence structure ([FUTURE] = [DISTAL] 'away/forward from utterance time').[8]

(18) *ġadi* 'future participle' > *ġadi* 'future tense'
$[_{\omega}$ [*ġadi*]:[~~x MOVE~~ [DISTAL]], [*v*[act]], [Part[agr]]] →
$[_{\omega}$ [*ġa(di)*]:[[~~x MOVE~~ [DISTAL]]], ~~[v[act]]~~, ~~[Part[agr]]~~]]

Deletion of the lexical semantic features and morphosyntactic features has consequences for the structure and derivation of the sentences that include the future participle form and the future tense form that are explored in the next two sections.

3. The future participle as a raising predicate

We start with the structural and derivational implications of the deletion of the meaning component GO and its external argument in the creation of the future participle form. The discussion here assumes the background conclusions above concerning the constituent structure and derivation of directed motion participles

8. A reviewer has inquired why the development of the future tense form necessarily involves deletion of the agreement features of the source participle in view of the fact that functional categories in general carry agreement features. What matters in the analysis outlined here is deletion of the features that make up the category Part[agr], assumed to exist over and above the agreement features, although their exact nature remains elusive. In this respect, Part[agr] is not necessarily different from *v*[agr], also assumed to include verbal features in addition to (object) agreement features. A future tense form that retains the agreement features of the source participle is consistent with the analysis, although this would require addition of the person feature characteristic of finite T[agr] in Arabic (Benmamoun, 2000). The presence and absence of the participial agreement features are treated here as overt indicators of the presence and absence of Part[agr].

in contexts such as (1a), repeated in (19)–(21). The subject 'the-girls' in (19)–(20) realises the external argument of the motion predicate MOVE in the representation of the participle form (the vocabulary item) repeated in (21). The subject originates in Spec, *v* and subsequently raises to Spec, T[agr] along the lines shown in (20).

(19) l-bnat (kan-u) ġadi[y]-in (l-s-sinima).
the-girls (BE.3.PL) going.PL (to-the-cinema)
'The girls are/were going (to the cinema).'

(20) [$_{\text{TP}}$ [$_{\text{DP}}$ the-girls] [$_{\text{T'}}$ T[3.PL] [$_{\text{PartP}}$ Part[F.PL] [$_{v\text{P}}$ ~~{t_{DP} the-girls}~~ [$_{v'}$ *v*[act] [$_{\sqrt{\text{P}}}$ √[GO] ...

(21) [$_{\omega}$ [*ġadi*]:[[x MOVE [DISTAL]], [*v*[act]], [Part[agr]]]

The future participle form in contexts such as (1b), repeated in (22), has the modified, reduced representation repeated in (23), which lacks the motion predicate MOVE and its external argument. The question that needs to be addressed is what kind of syntactic structure the representation of the future participle in (23) implies for sentences such as (22). A related question is how the future participle form contributes the future reading in (22) in view of the fact that the sentence includes an independent T[agr] node specified for present tense.

(22) l-bnat ġadi[y]-in imši-w (l-s-sinima).
the-girls going.PL go-3.PL to-the-cinema
'The girls are going to go (to the cinema).'

(23) [$_{\omega}$ *ġadi*:[[~~x MOVE~~ [DISTAL]], [*v*[act]], [Part[agr]]]

Because the future participle form in (23) retains the semantic feature [DISTAL] associated with the root and the morphosyntactic features associated with the functional heads *v* and Part, it corresponds to a predicate in sentence structure, essentially as in (20). However, since the future participle form lacks an external argument, the subject 'the-girls' of (22) cannot be its argument. Rather, it is an argument of the motion imperfective verb *mši* in the embedded clause and its presence in the subject position of the future participle must be the result of subject-raising to the root clause. These observations amount to the conclusion that, as a consequence of the loss of the motion predicate MOVE and its external argument, the future participle form becomes a raising predicate that occurs in a biclausal raising structure along the lines shown in (24), irrelevant details omitted. The word GOING in (24) is intended as shorthand for the representation in (23).

(24) [$_{\text{TP}}$ [$_{\text{DP}}$ 'the-girls'] [$_{\text{T'}}$ T[agr] [$_{\text{PartP}}$ GOING [$_{\text{TP}}$ ~~{t_{DP} the-girls}~~ [$_{\text{T'}}$ T[agr] ...
[$_{v\text{P}}$ ~~{t_{DP} the-girls}~~ [$_{v'}$ go ...

The analysis is consistent with the fact that raising predicates in general are semantically impoverished, although they may bear the same amount of morphosyntactic features as other predicates. In the case of the future participle *ġadi*, the impoverished semantic content consists of the path component with the distal value, interpreted relative to the tense value of root T[agr]. In (22), root T[agr] has the tense value [present], that is, a value that overlaps with the value of utterance time, assuming the interpretive mechanism outlined in Fassi Fehri (1993, 2012). This means that the distal path component of the participle expresses future relative to present. When root T[agr] has the tense value [past] (i.e. when the tense value is anterior relative to utterance time), as in (25), the distal path component of the participle expresses future in the past.[9]

(25) l-bnat kan-u ġadi[y]-in imši-w.
the-girls were-3.PL going.PL go-3.PL
'The girls were going to go (to the cinema).'

According to the analysis outlined, the component of the participles with the meaning GOING responsible for their ability to develop into future tense markers resides in the event structure of the root in the form of a path component with the distal value. This property has very little if anything to do with the participle form as such, which is defined by morphosyntactic features rather than semantic features. The possibility that these features may play a role comes from the view often expressed in the literature on Arabic that participles can express future tense and are widely used to express future, especially in nominal sentences. As pointed out in Ouhalla (2012b), while this is true of participles based on motion, activity roots such as those in (26), it is not true of participles based on stative roots such as those in (27). The latter do not have a future reading, as shown by the fact that they are incompatible with a future tense adverbial.

(26) l-bnat ġadi[y]-in/msafr-in (daba/ġədda).
the-girls going-PL/travelling-PL (now/tomorrow)
'The girls are going/travelling (now/tomorrow).'

9. The analysis amounts to a formal explanation of the concept of prospective future described in Comrie (1976, pp. 64–67), understood as future relative to the tense value of the sentence. The tense value of the sentence is encoded in root T[agr] and may be [present] or [past], the latter realised by the perfective form of √*KN* 'be.' Future is expressed by the feature [DISTAL] encoded in the root component of the participle form and interpreted relative to the tense value of root T[agr] to yield prospective future. In other words, prospective future is computed over two constituents, root T[present]/[past], which functions as the anchor, and the participle form with the path feature [DISTAL] 'away/forward from T[present]/[future].' As is explained below, when *ġa(di)* functions as a tense marker (i.e. a T-morpheme), its value is computed relative to utterance time in root contexts, as with T-morphemes in general.

(27) l-bnat naçs-in/fayq-in (daba/*ġədda).
the-girls sleeping-PL/awake-PL (now/*tomorrow)
'The girls are asleep/awake (now/*tomorrow).'

Morphosyntactic and structural reduction

We move on to discussion of the structural and derivational consequences of the deletion of the morphosyntatic features of the source participle form in the creation of the future tense form. The relevant example is repeated in (28) along with the representation of the future tense form arrived at above (29).

(28) l-bnat ġa(di) imši-w (l-s-sinima).
the-girls will go-3.PL to-the-cinema
'The girls are going to go (to the cinema).'

(29) [$_{\omega}$ [*ġa(di)*]:[[~~x MOVE~~ [DISTAL]]], ~~[v[act]]~~, ~~[Part[agr]~~]]

The rump feature-content of the future tense form shown in (29) implies that it can be inserted under a node specified for a temporal feature, that is, a node of the type T[future]. In structural terms, this means that the form does not correspond to a participle predicate of a clause; rather, it corresponds to a T[future] node in the clause that contains the main verb. Accordingly, sentences such as (28) have the mono-clausal structure shown in (30).

(30) [$_{TP}$ ['the-girls'] [$_{T'}$ T[future]:[*ġa(di)*] [$_{vP}$ ~~['the-girls']~~ [$_{v'}$ GO ...

Compared to the structure of sentences that include the future participle form (24), which is b-clausal, (30) can be understood to mean that the change future participle > future tense results in structure-reduction schematised in (31). However, such descriptions are at best metaphorical in nature, bearing in mind that the relevant change affects individual vocabulary items rather than syntactic structures in combination with the fact that syntactic structures are merely projections of features. The reasoning also applies to the fact that future tense *ġa(di)* ceases to be a participle and becomes a tense constituent (reanalysis). This does not amount to a change in category features or class, but, rather, to the deletion of the morphosyntactic features that define participles, namely *v*[act/pass] and Part[agr]. The ability to express (future) tense is not restricted to vocabulary items of the T-class, and can be expressed by a variety of categories such as temporal adverbs, temporal PPs and temporal participles.

(31) [the girls are/were going [the girls go to the cinema]] →
[the girls will go to the cinema]

As far as the interpretation of the future tense form *ġa(di)* is concerned, the idea is that the feature [DISTAL] receives the temporal interpretation future in root context such as (28) because it is interpreted relative to utterance time by virtue of the fact that it occupies root T[agr] as with T-morphemes in general. When *ġa(di)* co-occurs with another, higher T[agr] in contexts that involve complex tense such as (32), it is interpreted relative to higher T[agr]. In (32), *ġa(di)* occupies lower T[future] and expresses future relative to past encoded in higher T[past] (future in the past) on the plausible assumption that such sentences with complex tense involve two projection of T with different values along the lines shown in (33) rather than two clauses (see Fassi Fehri, 2012 and references cited therein).

(32) l-bnat kan-u ġa(di) imši-w (l-s-sinima).
the-girls were-3.PL will go-3.PL (to-the-cinema)
'The girls were going to go (to the cinema).'

(33) $[_{\mathrm{TP1}}$ ['the-girls'] $[_{\mathrm{T'}}$ T[past]:['were'] $[_{\mathrm{TP2}}$ $[_{\mathrm{T'}}$ T[future]:[*ġa(di)*]
$[_{v\mathrm{P}}$ ~~['the-girls']~~ $[_{v'}$ go ...

3. Conclusion

The development of future participles such as *ġadi + Agr* and future tense markers such as *ġa(di)* and *maši* from related directed motion participles in Moroccan Arabic can be accounted for in terms of the independent morphological processes of Feature Copying and Feature Deletion. Feature Copying (word-extraction out of a sentence-context) yields a new vocabulary item that is (partially) homophonous with the source participle form. Feature Deletion fashions the copy into a new functional vocabulary item by removing the semantic component responsible for the motion meaning GO and the arguments it selects (future participle) and the morphosyntactic components that define participles (future Tense). Feature Deletion can also reduce the copy to a shorter form by removing phonological features corresponding to prosodic units such as syllable as in the case of *ġa(di)*. Because grammaticalization applies to words, its effects are to some extent the reverse of those familiar from word-formation. Word-formation is essentially a process of feature-composition and structure-building that culminates into permissible output forms, words. This is a function of the fact that the objects that word-formation makes use of and applies to are primitives in the form of features and feature-bundles. In contrast, grammaticalization largely involves feature-deletion, which is a function of the fact that the objects it makes use of and applies to are words (output forms) rather than lexical primitives.

References

Benmamoun, E. (2000). *The feature structure of functional categories*. Oxford: Oxford University Press.

Berman, R. A. (2003). Children's lexical innovations: Developmental perspectives on Hebrew verb structure. In J. Shimron (Ed.), *Language processing and acquisition in languages of semitic, root-based morphology* (pp. 243–291). Amsterdam: John Benjamins. DOI: 10.1075/lald.28.13ber

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA.: MIT Press.

Comrie, B. (1976). *Aspect*. Cambridge: Cambridge University Press.

Embick, D. (2000). Features, syntax, and categories in the Latin perfect. *Linguistic Inquiry*, 31, 185–230. DOI: 10.1162/002438900554343

Embick, D., & Noyer, R. (2007). Distributed morphology and the syntax/morphology interface. In G. Ramchand, & C. Reiss (Eds.), *The Oxford handbook of linguistic interfaces* (pp. 289–324). Oxford, New York: Oxford University Press.

Fassi Fehri, A. (1993). *Issues in the structure of Arabic clauses and words*. Dordrecht: Kluwer. DOI: 10.1007/978-94-017-1986-5

Fassi Fehri, A. (2012). *Key features and parameters in Arabic grammar*. Amsterdam: John Benjamins Publishing Company. DOI: 10.1075/la.182

Halle, M., & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale, & S. J. Keyser (Eds.), *The view from building 20: Essays in linguistics in honor of Sylvain Bromberger* (pp. 111–176). Cambridge, MA.: MIT Press.

Hopper, P. J., & Closs Traugott, E. (2003). *Grammaticalization*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139165525

Jackendoff, R. (1985). Multiple subcategorization and the Theta Criterion. *Natural Language and Linguisitic Theory*, 3, 271–295. DOI: 10.1007/BF00154264

Jackendoff, R. (2010). *Meaning and the lexicon: The parallel structure*. Oxford/New York: Oxford University Press.

Kratzer, A. (1995). Stage level and individual level predicates. In G. N. Carlson, & F. J. Pelletier (Eds.), *The generic book* (pp. 125–175). Chicago/London: University of Chicago Press.

McCarthy, J. (1979). Formal problems in semitic phonology and morphology. Unpublished Ph.D. dissertation, MIT.

McCarthy, J. (1981). A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry*, 12, 373–418.

Levin, B., & Rappaport Hovav, M. (1992). The lexical semantics of verbs of motion: The perspective from unaccusativity. In I. M. Roca (Ed.), *Thematic structure: Its role in grammar* (pp. 247–269). Berlin: Foris.

Lightfoot, D. (1999). *The development of language: Acquisition, change and evolution*. Malden/Oxford: Oxford University Press.

Ouhalla, J. (2011). Categorization and the interface levels. In R. Folli, & C. Ulbrich (eds.) *Interfaces in linguistics: New research and perspectives* (pp. 116–134). Oxford: Oxford University Press.

Ouhalla, J. (2012a). Lexical change and the architecture of the lexicon. In E. Torrego (Ed.), *Of grammar, words and verses* (pp. 41–66). New York and Amsterdam: John Benjamins. DOI: 10.1075/lfab.8.06ouh

Ouhalla, J. (2012b). The development of future markers in root-frames and satellite-frames: Arabic and French GO(ING) versus English and Berber GO(ING)+TO. To appear *in Journal of Semitic Studies, Supplement on Grammaticalization in Semitic*.

Rappaport Hovav, M., & Levin, B. (2010). Reflections on manner/result complementarity. In M. Rappaport Hovav, E. Doron, & I. Sichel (Eds.), *Lexical semantics, syntax, and event structure* (pp. 21–36). Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199544325.003.0002

Rice, F. A., & Sa'id, M. F. (1960). *Eastern Arabic*. Washington D.C.: Georgetown University Press. Reprinted in 2005.

Roberts, I., & Roussou, A. (2003). *Syntactic change: A minimalist approach to grammaticalization*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511486326

Shlonsky, U. (1997). *Clause structure and word order in Hebrew and Arabic: An essay in comparative syntax*. New York/Oxford: Oxford University Press.

Yod-dropping in *b-imperfect* verb forms in Amman*

Enam Al-Wer

This article presents analysis of data from the emerging dialect of Amman within the framework of 'dialect contact' and 'new dialect formation' (Trudgill, 1986, 2004). In particular, it focuses on the conjugation of 3rd person singular and plural verb forms with or without /j/. The dialect of Amman can be described as the output of contact between speakers of two major dialect groups: Jordanian and Palestinian. In both groups of dialects, the imperfective takes the form *b* + *Imperfect* form, but whereas in Jordanian dialects *Yod* is dropped from the stem in the *b-imperfect* form in all environments, in Palestinian dialects it is dropped in open syllables only. The analysis reveals two important innovative patterns of conjugation: (i) *Yod* is dropped everywhere except where it carries 'person' information, and (ii) in *hamza*-initial verbs ('glottal-initial verbs' such as /ʔakal/ '*to eat*') speakers with Jordanian heritage introduce a new form for 1st person singular and reallocate an existing form to the 3rd person singular masculine.

Keywords: sociolinguistics, contact, Jordan, Amman, dialect formation

1. Dialect contact and new dialect formation

The study of the linguistic outcome of face-to-face interaction between speakers of mutually intelligible dialects has been brought to the forefront of sociolinguistic research especially since the publication of Trudgill (1986). The model presented by Trudgill builds on the simple observation that speakers tend to modify their speech in inter-dialectal communications. The modification may involve *convergence*, i.e. making one's speech more similar to the speech of one's interlocutors, as well as *divergence* from the speech of one's interlocutors, as predicted by the

* Acknowledgment
I am grateful to Jonathan Owens who read and commented on a number of previous versions of the analysis presented in this article. Many of his suggestions, both in analysis and presentation, have been included. Any remaining errors are my own.

principles of the "accommodation theory" (Giles, Taylor & Bourhis, 1973). The model is expanded to account for cases of contact-induced long-term linguistic modifications, diffusion of innovations across space, and to cases of transplanted dialects and new dialect formation.

Since Trudgill's seminal work in this field, a wide range of situations has been investigated, including cases of dialect boundary regions (Britain, 1991, 2001, 2002), rural immigrant groups in the city (Kerswill, 1993), leveling in new towns (Kerswill & Williams, 2000), diffusion of innovations from large urban centers to smaller towns (Hernández-Campoy, 2003; Trudgill 1983, 1986), and multicultural settings (Cheshire, Kerswill, Fox & Torgersen, 2011).

New dialect formation is the topic of two important publications by Gordon, Campbell, Hay, Maclagan, Sudbury & Trudgill (2004) and Trudgill (2004), which focus in particular on the formation of New Zealand English. Both accounts are based on recordings collected between 1946 and 1948 by the New Zealand Broadcasting Corporation, which provided samples of the speech of the first locally born non-aboriginal New Zealanders (descendants of the early immigrants from the British Isles). Gordon et al. (2004) present a detailed quantitative sociolinguistic account of the emergence and focusing of a new dialect. Trudgill (2004), on the other hand, is less conventional in that he sets aside social factors, such as 'prestige' and 'identity', which are claimed to play no role in *tabula rasa* conditions, and accounts for the formation of New Zealand English within a deterministic model based essentially on the demographic representation of the linguistic features in the original mix of dialects, leveling out of localized/marked and minority features and drift.[1] (Also see Trudgill 2008a).

In Arabic-speaking communities, rapid urbanization and various socio-political developments have led to large-scale population movements and thus to dialect contact situations. A number of such cases dealing with different cities in the Arab world have been investigated in recent years such as the 'Amman Project' on which this article draws (see Al-Wer, 2002, 2003, 2007), Hachimi (2005), Al-Essa (2008), and Al-Qouz (2009). The range of situations and settings that await investigation in Arabic-speaking communities is wide and varied[2].

In analyzing linguistic features found in contact situations, it is sometimes difficult to ascertain whether the features are *contact-induced* rather than changes that would have occurred had contact not existed. In the case of Amman, this issue

1. Trudgill (2004) has been widely debated and critiqued. In particular, see the discussion in *Language in Society* by Mufwene (2008), Tuten (2008), Schneider (2008), Coupland (2008), Bauer (2008), and Holmes & Kerswill (2008); and the rejoinder by Trudgill (2008b) in the same volume.

2. Potentially especially interesting situations are the countries of the Gulf where in many cases the expatriate Arabic-speaking groups are larger than the native populations.

is immaterial since this is not a case of continuation of change in a dialect but the *formation* of a new dialect from scratch, as will be explained presently.

A historical sketch of Amman

Modern Amman was built roughly on the site of Rabbath Ammon, 'the Great City of the Ammonites', dated to 10th century BC. It was renamed 'Philadelphia' by the Ptolemies (3rd century BC). The Romans (1st century BC) kept the name 'Philadelphia', and the city became part of their Decapolis. It was probably the Umayyad rulers of Damascus (7th century) who restored the city's original Semitic name, but Amman lay in ruins during the Umayyad era and remained so until the mid 19th century.

The first settlers in the modern history of Amman arrived in 1876. They were speakers of Adyge, or what is popularly known in Arabic as *šarkasi* 'Circassian', a northwestern Caucasian language.[3] According to Ottoman records, in 1906 Amman had 5,000 Circassian settlers and virtually no Arabic-speaking inhabitants. The city therefore does not have a pre-20th century history of being an Arabic-speaking city.

Amman was resurrected in 1921 as the capital of the newly formed Emirate of Transjordan, which attracted immigrants from various locations. For the most part, the internal migrants came from the nearby city of Sult, at the time the largest urban settlement east of the River Jordan. From Palestine, the vast majority of the early settlers in Amman were from Nablus, many of whom had in fact been living in Sult for some time before moving into Amman. The early settlers also included a few merchant families and political activists from Damascus, which was then under the French mandate. The later migrants came from the south, particularly from Kerak and Madaba, and also from the Palestinian cities of Jerusalem, Haifa, and Jaffa. By the 1930s, the city's population, including the Circassians, had grown to 10,000. Internal migration continued steadily through the 1940s, but the most sudden population increase occurred as a result of the 1948 and 1967 wars and the displacement of Palestinians. In the period between 1950 and 1990, the population of Amman increased more than fifteen times. It is currently estimated to be 1.6 million.

2. The population

The third generation Circassians in Amman became native speakers of Jordanian Arabic. Currently, they make up roughly 2% of the city's population. The majority of Amman's inhabitants belong to two major groups: the 'Jordanians', i.e. the

3. For details about the Circassian community and language in Jordan, see Al-Wer 1999.

migrants from other East Bank locations; and the 'Palestinians', i.e. the migrants from Palestinian towns and villages. For convenience, I will simply use the labels 'Jordanian' and 'Palestinian' to refer to these groups, although neither group is homogeneous. It is noticeable, nevertheless, that while the Palestinian society is sharply stratified in terms of urban (townsfolk) and rural (peasants), which might explain the persistence of the sharp linguistic distinctions in Palestinian dialects between town dialect and village dialect even where the villages are in close proximity to the town, the Jordanian society is largely tribal, a semi-egalitarian system which cuts across social and linguistic distinctions between town and village to a large extent (*cf.* Salibi, 1993, chapter 6). The Palestinian system of social organisation was simply transplanted in Amman. One observes, for instance, that even though the Palestinian townsfolk and the villagers have become citizens of the same locality (Amman), they continue to form two sharply distinctive communities.

No official statistics are available on the current demographic representation of the two groups that make up the population of Amman.[4]The majority of the Palestinians in Amman come from two backgrounds: rural, speakers of *fallaħi* dialects, and urban, speakers of *madani* dialects.[5] The *fallaħi* dialects are marked by a number of localized linguistic features that are peculiar to rural Palestinian dialects (especially the central and northern rural dialects), notably [k] for /q/ and unconditioned affrication of /k/. On the other hand, the urban Palestinian dialects are very similar to other city dialects in the region, such as the dialects of Damascus and Beirut. The results from my research in Amman so far (details below) show very clearly that none of the features that have become characteristic of the new dialect are rural Palestinian in origin. The competition in Amman is between Jordanian linguistic features and urban Palestinian features, and therefore it is the demographic representation of the Jordanian and the urban Palestinian dialects that is likely to influence the making of the new dialect. By a rough calculation, the Jordanian and urban Palestinian dialects probably have equal demographic representation in the city.

3. The research

The data presented in this article come from the Amman Project, on-going research that aims to trace the formation of the city's dialect from the earliest stages. The data presented in this article were obtained in two stages using standard sociolinguistic

4. If such statistics do exist, they do not seem to be publically available.

5. Both of these sub-groups of Palestinian dialects are akin to the Sedentary type of Arabic dialects. The Jordanian dialects are classified as southern Levantine dialects but they are considered to descend from the Bedouin type (see Cleveland, 1963).

interviews. The first stage included three generation of speakers from four families (8 grandparents, 8 parents, and 14 of their children). Two of these families originally came from the city of Sult and two families originally from the Palestinian city of Nablus; this cohort of speakers was meant to represent speakers from the two major input varieties (Sult and Nablus). The second stage focuses on the third generation from both dialectal heritages (Jordanian and urban Palestinian). So far there are 35 speakers in this group, 20 female and 15 male. These data are supplemented by additional recordings from the following sources: 6 younger speakers (all 17–20 years of age, three female) from Sult, and two male speakers from Nablus who were visiting relatives in Amman when the research was being conducted but who normally live in Nablus. In addition, I drew on data from my earlier research in Sult, collected in 1987, and a follow-up smaller collection in 1997.[6]

For the purpose of the analysis of the feature presented in this article, 1150 imperfect verb forms were extracted. Of these, 490 are 3rd person singular and plural masculine forms. Bearing in mind that free interviews, as opposed to elicitations, limit the possibility of obtaining the full paradigm of conjugations or enough tokens per speaker of certain conjugations, the data used in the current analysis show some clear and consistent patterns, as explained below.[7]

The informants all live in West Amman, which is the wealthier and more cosmopolitan part of the city, and they mostly come from the upper middle class. Strictly speaking, the findings should be understood to be applicable within the geographical and social limitations of the sample of speakers interviewed so far. In the less affluent areas of Amman, familial networks are considerably closer. Extended families often reside in the same neighbourhoods and the younger generations tend to spend most of their leisure time within the family circle. In West Amman, on the other hand, where leisure facilities are much more widely available, the youngsters form intimate peer group relations and spend most of their leisure time away from their homes and their families. The differences between East and West Amman may be compared to the differences between localized, tight-knit working class communities and middle class suburban communities in cities in the west.

4. The three generations and their overall sociolinguistic profile

Including the early migrants, Amman only has three generations of native inhabitants. The first generation arrived in the city as adults, the second generation

6. For details about the data from Sult, see Al-Wer & Herin (2011).

7. Further research using elicitation methods is currently underway, which will supplement the data presented in this article and allow us to quantify the feature.

includes the children who arrived with their parents (or were born in the city), and the third generation represents the first native-born generation in the city. In three generations, the dialect has undergone a considerable degree of focusing to the extent that it is now possible to speak of a distinctive Ammani dialect.

Trudgill notes that the outcome of dialect contact can be a new "relatively focused and discrete variety" (Trudgill, 1986, p. 91), and demonstrates this type of development by the case of Burträsk Swedish, investigated by Thelander (1979, cited in Trudgill, 1986: 91–93). The contact situation in Amman has resulted in a similar situation. The stages of the formation of the new dialect in Amman are summarized below (for further details, see Al-Wer, 2007).

First generation speakers arrived in the city as adults; they spoke the dialects of the Jordanian or Palestinian towns from which they migrated, and to a large extent they continue to do so. However, the dialects spoken by this generation do show some divergence from the original dialects by leveling out the most localized and most marked features, a process which Trudgill calls *rudimentary leveling* (Trudgill, 2004, pp. 89–93). The most obvious feature which undergoes leveling at this stage from the Jordanian side (Sult) is /k/ affrication, which is totally absent in the data from first generation male and female Jordanians. The traditional dialect of Sult affricates /k/ conditionally (mostly in the vicinity of front vowels such as /ke:f/ > /ʃte:f/ 'how') (see Palva, 1994; Herin, 2010). Another localized feature that undergoes leveling at this stage is gender distinction in the third person plural pronouns, pronominal suffixes, and endings. The speakers tended to neutralize gender distinction in these forms. One of the examples of rudimentary leveling in the speech of Palestinians (Nablus) is lowering of /e:/ to /a:/ in items like /mbe:riħ/ > /mba:riħ/ 'yesterday' or /se:ʕə/ > /sa:ʕə/ 'hour'. Extremely raised variants of /a/ in these examples are generally stereotypical features of Palestinian dialects (most notable in the dialect of Jerusalem), and the tokens with the lowered vowels in the speech of this generation of Palestinian speakers can be considered a convergence or accommodation to Jordanian dialects, which have a low vowel in these items.

The second generation exemplifying the second stage of dialect formation show extreme inter-speaker and intra-speaker variability as a result of using a mixture of the two norms to the extent of appearing chaotic, although their speech can still be identified as either Jordanian or Palestinian, mostly through vocalic and morphosyntactic features. For example, in the data, the same speaker used three forms of the 2nd person suffix: Jordanian *-ku,* Palestinian *-kon* and the new Amman form *-kum* in /ki:f ħa:l-ku/, /ki:f ħa:l-kon/ and /ki:f ħa:l-kum/ 'how are you'. The Palestinian male speakers and the Jordanian female speakers use a mixture of Palestinian [ʔ] and Jordanian [g] for /q/.

The third generation, all of whom were born in the city, do not adhere to their parents' dialects faithfully, but in some cases they focus the patterns that appear in their parents' generation, while for many features they set totally new patterns, and use new features which were not present in the original mix. The mixture and variability witnessed in the second generation is considerably reduced, and some orderly linguistic behavior is established. We find a high degree of stability of usage of certain linguistic features, for example: the use of the stop variants of interdental sounds, which is almost consistent in the speech of the third generation; use of fudged linguistic forms, such as raising of pausal /a/ along the pattern found in urban Palestinian (though the raised vowel is phonetically identical to the raised vowel found in Jordanian dialects); and many new patterns and new features, like the use of innovative *-kum,* which does not occur in the input dialects.[8]

5. The data and analysis

Yod in imperfect verb forms

This feature concerns the conjugation of verbs in the imperfect of 3rd person masculine forms with or without /j/, as in /bjiʕmal/ vs. /biʕmal/ 'he does', /bjiʕrif/ vs. /biʕrif/ 'he knows', /bjilʕabu/ vs. /bilʕabu/ 'they play'. The Arabic dialects that use the *b-imperfect* form (principally Levantine and Egyptian dialects) vary with respect to this feature between dialects that maintain *Yod* in all cases, Cairene Arabic being an example, and those which drop it categorically, as in Jordanian dialects (see further below).

In the input dialects from the first and second generations in Amman, the patterns that were found most consistently in the data are summarized below. For the purpose of this section, I have omitted variations, but these will be examined later.

Urban Palestinian[9]

(1) *hamza*-initial (glottal-initial) verbs /ʔakal/ 'to eat', /ʔaxađ/ 'to take'

	First generation	*Second generation*
3rd pers	bja:kul/bja:xud	bja:kul/bja:xud
1st pers	ba:kul/ba:xud	ba:kul/ba:xud

8. Detailed analyses of these features can be found in Al-Wer (2003) & Al-Wer (2007).

9. I am aware that some Palestinian dialects, e.g. Jaffa and may be Haifa, have forms of the *hamza*-initial verbs that are identical to the forms found in Jordanian dialects (with /o:/ rather than /a:/ or /ja:/). The forms reported in this article are the ones recorded from 1st and 2nd generation speakers who originally come from Nablus.

The forms under (1) show that /j/ carries person information, and there is no vowel distinction between the forms in the 1st and 3rd person. The presence of /j/ in these forms is therefore functional.

Urban Palestinian

(2) other verbs: /ħaka/ 'to speak', /ʕirif/ 'to know'

	First generation	*Second generation*
3rd pers	bjiħki/bjiʕrif	bjiħki/bjiʕrif
1st pers	baħki/baʕrif	baħki/baʕrif

In the forms in (2), the presence of /j/ can be considered a redundant feature, since the person information is also signalled in vowel distinctions, /i/ versus /a/.

Jordanian

(3) *hamza*-initial verbs /ʔakal/ 'to eat', /ʔaxađ/ 'to take'

	First generation	*Second generation*
3rd pers	bo:kil/bo:xið	bo:kil/bo:xið or bo:xid
1st pers	bo:kil/bo:xið	bo:kil/bo:xið or bo:xid

The forms in(3) are identical for both persons. The supplementary data from speakers in the city of Sult confirm that in the traditional dialect no morphological marking of person in the 3rd and 1st person singular in these verbs is present.

Jordanian

(4) other verbs: /ħaka/ 'to speak', /ʕirif/ 'to know'

	First generation	*Second generation*
3rd pers	biħki/biʕrif	biħki/biʕrif
1st pers	baħki/baʕrif	baħki/baʕrif

In these forms, the person information is signalled in vowel distinctions, /i/ for 3rd person and /a/ for 1st person.

In the speech of the third generation, the forms listed below are found.

Ammanis of Palestinian parents (i.e. Palestinian input):

(5) *hamza* initial verbs /ʔakal/ /ʔaxad/

3rd pers	1st pers
bja:kul/bja:xud	ba:kul/ba:xud

(6) Other verbs: ħaka, ʕirif

3rd pers	1st pers
biħki/biʕrif	baħki/baʕrif

The forms in (5) and (6) show that Ammanis of Palestinian parents maintain /j/ where it is functional only, i.e., in *hamza*-initial verbs, but they drop it where it is redundant. In the speech of this group, the forms they use in (6) are identical to the forms found in the traditional Jordanian dialects in that they rely on vowel distinctions only.

Ammanis of Jordanian parents (i.e. Jordanian input):

(7) *hamza*-initial verbs /ʔkal/ /ʔaxad/

3rd pers	1st pers
bo:kil/bo:xid or bo:kol/bo:xod	ba:kul/ba:xod

(8) Other verbs: ħaka, ʕirif

biħki/biʕrif	baħki/baʕrif

The examples in (7) show that Ammanis of Jordanian parents now diverge from their parents' dialects by marking person morphologically through vowel change in the *hamza*-initial verbs. They use identical forms to those used by their parents in (8).

Variation in the data and discussion

To some extent, the patterns presented in (1)–(8) in the previous section are abstractions, which represent the patterns found to be used most consistently. But, the data contained variations, which I will examine presently.

As shown in the examples in (5), third generation Palestinians drop *Yod* from the *non-hamza* initial verbs. There were only three exceptions to this pattern. Two were instances of the verb *ħaka* 'to speak', which was produced by one individual as /bjiħki/ 'he speaks'; the other was an example of /bjilʕabu/ 'they play' in the speech of a different speaker.

Based on these data, it can be concluded that the third generation speakers of Palestinian heritage drop redundant /j/, and by doing so they converge to the speech of the third generation speakers of Jordanian heritage with respect to this feature. The data from the Jordanian speakers in Amman and from Sult provide evidence to the fact that Jordanian dialects are *Yod*-free.

Further examination of the data from the first and second generation Palestinians reveals some important patterns of variation. Already in the speech of the first generation there are examples of *Yod*-dropping, and more such examples occur in the data from the second generation Palestinians. Thus, there is an *apparent* convergence to the Jordanian pattern on the part of the first and second generation Palestinians. However, by examining the lexical items in which *Yod* was dropped in the speech of 1st and 2nd generations, it can be noticed that in all of them the dropping of historical /j/ can be accounted for by a morphophonemic rule, according to which /j/ is dropped in an open syllable.[10] Thus we find:

10. I had initially analysed this feature as concerning a particular class of verbs, namely 'hollow verbs' ('ajwaf') since all occurrences of *Yod*-less forms in my data at that stage occurred in such verbs. Jonathan Owens (personal communication) suggested that the relevant feature is rather concerned with syllable structure, not a certain class of verbs. I thank him for bringing this to my attention. Further data, which were obtained through elicitation techniques and are currently

(9) ji.ʔu:l → biʔu:l 'he says'
ji.ru:ħ → biru:ħ 'he leaves'
ji.ɣi:b → biɣi:b 'he vanishes'
ji.ku:n → biku:n 'he is/in the state of being'
ji.bi:ʕ → bibi:ʕ 'he sells'

On the other hand, /j/ is maintained consistently in the examples listed in 10 (where /j/ occurs in closed syllables):

(10) jiʔ.dar → bjiʔdar/ 'he is able to'
jis.ʔal → bjisʔal 'he asks'
jiʕ.mal → bjiʕmal 'he does'
jiħ.ki → /bjiħki/ 'he say'
jil.ħaʔ → /bjilħaʔ/ 'he follows'

It appears therefore that the Palestinian dialects which were transplanted in Amman had a morphophonemic rule which dropped /j/ in open syllables, yielding /biʔu:l/, /biru:ħ/, etc. (without /j/), but /bjilʕab/, /bjiħki/, etc. (with /j/). It should be noted that the conclusion that the dialect spoken by the earlier generations of Palestinians in Amman already dropped /j/ in a certain environment is further supported by the data collected from the two Nablusi speakers, which contained a total of 7 examples of *Yod*-less items, all of which were forms of *ra:ħ* '*t*o go' and *k*a:n '*to be*', whereas all other verbs occurred with *Yod*.

In her account of the historical phonology of the Jerusalem dialect, Garbell (1958: 325) mentions dropping of /j/ in /biku:n/ < /bijku:n/ as a length feature which was lost in unstressed syllables (during the 18th-20th centuries). [11] She also mentions alterations between /bisʔal/ and /bjisʔal/ as a feature which could be found in the speech of "Christians and young people" (p. 325). In my data, which come from Nablusi Palestinians, the *Yod*-less items are confined to the environment mentioned above, but the observations provided by Garbell may be taken as further evidence that in the Palestinian dialects dropping of *Yod* was not altogether absent.

To my knowledge, there is no information in the literature about the maintenance/dropping of *Yod* in other Levantine dialects. Informal inquiries on my part about the forms used by the native speakers of the dialects of Beirut and Damascus indicate that the pattern used in these dialects is identical to the Palestinian pattern

being analysed for a different feature, confirm that /j/ is indeed dropped in open syllables; more *Yod*-less examples include /biʒurr/ 'he pulls'; /biħutˤ/ 'he puts'; /biħibb/ 'he likes/loves'; /bixarrib/ 'he spoils'; /biʒallis/ 'he fixes'.

11. This is the only reference I could find that mentions variation in the use of *Yod* in these verb forms in a Palestinian dialect.

I found in Amman, i.e. *Yod* is maintained in closed syllables and dropped in open syllables. Outside the Levant, the *b-imperfect* form is a well-known feature of the dialect of Cairo. My informal inquiries with speakers of Cairene Arabic indicate that *Yod* is maintained in all cases.[12] Thus in Cairo one finds forms like /bijʔu:l/ 'he says' and /bijgurr/ 'he pulls'. It therefore appears that concerning *Yod* in the 3rd person verb forms the *b-imperfect* dialects of the Levant and Egypt can be arranged along a continuum: at one end are the most conservative dialects that do not drop *Yod* in any environment (Cairo); at the other end are the most innovative dialects that drop it in all environments (Jordanian); and in between there are dialects that drop *Yod* variably (conditionally, in open syllable: Damascus, Beirut, urban Palestinian). Assuming that the *b-imperfect* is a shared innovation, which Levantine and some Egyptian dialects inherited from a common source, the pattern found in Jordanian dialects, whereby *Yod* is dropped unconditionally, may be viewed as the continuation to completion of a change that had affected Levantine dialects at an earlier stage.

To conclude, the sequence of events in Amman can be reconstructed in the following way:

i. At the time of migration to Amman, the *Yod*-less form in the Palestinian dialects was the marked form, derived by a phonological rule.
ii. In Amman, speakers of such dialects came in contact with *Yod*-free dialects. In successive generations, and through exposure to and contact with the Jordanian dialects, the conditions that governed *Yod*-dropping in the original dialects were broken.
iii. As the frequency of occurrence of redundant *Yod* decreased, the third generation no longer had evidence for its usage among their peers.
iv. In the final stage, *Yod*-less forms became the unmarked choice. The factor that eventually led to the unmarking of these forms was the contact with *Yod*-free dialects, i.e. the Jordanian dialects.

As to the variation found in the data from the Jordanian speakers with respect to morphological person-marking in the *hamza*-initial verbs, recall that in the input dialects (example (3)), the forms for 3rd person and 1st person are identical: /bo:kil/ 'I eat/he eats'. There are no examples of this pattern at all in the third generation speakers from Jordanian backgrounds. So, the distinction /ba:kul/ 'I eat' versus /bo:kil/ or /bo:kol/ 'he eats' must have been established fairly quickly. The data show, however, that person-marking starts in the second generation of Jordanian speakers in Amman, and the variation found in this generation is

12. I thank Ghada Khattab, Hanadi Ismail and Manfred Woidich for responding to my queries concerning the forms used in Beirut, Damascus and Cairo.

inter-speaker only – i.e., the speakers either make the distinction consistently, or they do not make it at all. The supplementary data I have from the six younger speakers in Sult show that the traditional feature (i.e., no person marking) is largely the norm even among the younger speakers in Sult. It appears, therefore, that the emergence of person distinctions for these verbs in the speech of Jordanians in Amman was induced by contact with urban Palestinian dialects. One would expect for there to be some difficulty in acquiring a distinction (as opposed to losing one); but notice that the innovative derivation /ba:kul/ is already available in the environment.

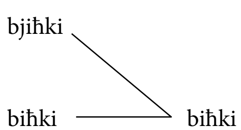
Finally, four tokens of the Palestinian form /bja:kul/ were found in the speech of two third-generation female Jordanian speakers, and two tokens of the Jordanian form /bo:kil/ were found in the speech of one third generation male Palestinian speaker. Although the data are strongly suggestive that there is a correlation between speakers' choices and dialectal background with respect to this feature, the exceptions listed here are not insignificant. They are at least an indication that features which were originally peculiar to a certain dialect in the original mix are used by the third generation regardless of their dialectal heritage.[13]

3. Conclusion

Prolonged contact between closely related dialects, as is the situation in Amman, provided there is a degree of social integration of the various linguistic groups (cf. Siegel, 1985) can lead to linguistic leveling and simplification as part of a process of koineization. Koineization normally results in: (i) reduction of the degree of variability which was originally present; and (ii) focusing and stabilization of a new linguistic system (see Kerswill & Williams, 2000; Trudgill, 1986).

The data on the development and stabilization of verb forms with and without *Yod* and the restructuring of constraints on its occurrence as a result of the contact situation in Amman can be understood as a case of simplification. The following diagrams re-state the findings in these terms:

(a) 1st & 2nd generations 3rd generation



13. These *hamza*-initial verbs are currently being analysed further using a larger pool of elicited data.

Diagram (a) shows that the simplification process has resulted in reduction in the number of forms used in the new dialect.

(b) Palestinians (1st, 2nd & 3rd) bja:kul ba:kul

Jordanians (1st & 2nd) bo:kil

Jordanians (3rd) bo:kil ba:kul

What can be seen in diagram (b) is that the Jordanian ambiguous form /bo:kil/ is rendered unambiguous by adding a separate form to mark person distinctions. In this restructuring of the system, existing forms are used: the function of the heritage form is reduced to mark 3rd person only, and the form /ba:kul/, a form that is already available in the linguistic environment, is added to mark 1st person. From the perspective of the traditional Jordanian dialects, this development appears to be a complication rather than a simplification since it resembles a 'split' and adds a form (rather than reducing the number of forms). On the other hand, it can be argued that the traditional Jordanian system of assigning two persons to one morphological form is cognitively complex, given that the same dialects mark 1st and 3rd persons morphologically in all other categories of verbs. The traditional Jordanian system is all the more complex for the third generation Jordanians in Amman given that in this context it is a minority form and therefore not as frequently encountered. In semantic terms, the development leads to regularization of person representation: each person is assigned a unique form; 3rd person form is maintained and assigned exclusively to 3rd person; and the form /ba:kul/, which is already available in the community, is assigned exclusively to 1st person. Hence there is a reduction in meaning-form assignments.

The end result is a reduction in verb forms, from five forms (/bjiħki/, /biħki/, /bja:kul/, /ba:kul/, /bo:kil/) to four forms (/biħki/, /bja:kul/, /bo:kil/, /ba:kul/) as a result of the leveling out of the redundant form /bjiħki/. By reducing the number of morphological forms the meaning is not impacted (no change in *referential meaning*). Therefore, the informational value of these developments must lie in the extra-linguistic sphere and be represented at the level of the 'community'. Koineization of the imperfect verb forms is simultaneously koineization of communal dialect differences.

References

Al-Essa, A. (2008). *Najdi speakers in Hijaz: A sociolinguistic investigation of dialect contact in Jeddah*. Unpublished Ph.D. thesis, University of Essex.

Al-Qouz, M. (2009). *Dialect contact, acquisition and change among Manama youth, Bahrain*. Unpublished Ph.D. thesis, University of Essex.

Al-Wer, E. (1999). Language and identity: The Chechens and the Circassians in Jordan. Proceedings of the first international conference on Arabic-English contrastive and comparative studies, Dirasat (special issue). Deanship of Academic Research, (University of Jordan), pp. 253–268.

Al-Wer E. (2002). Jordanian and Palestinian dialects in contact: Vowel raising in Amman. In M. Jones, & E. Esch (Eds.), *Language Change: The interplay of internal, external and extra-linguistic factors* (pp. 63–79). Berlin: Mouton de Gruyter.

Al-Wer E. (2003). New dialect formation: the focusing of -kum in Amman. In D. Britain, & J. Cheshire (Eds.), *Social dialectology: In honour of Peter Trudgill* (pp. 59–67). Amsterdam: John Benjamins. DOI: 10.1075/impact.16.06alw

Al-Wer E. (2007). The formation of the dialect of Amman: from chaos to order. In C. Miller, E. Al-Wer, D. Caubet, & J. C. E. Watson (Eds.), *Arabic in the city: Issues in dialect contact and language variation* (pp. 55–76). New York/London: Routledge.

Al-Wer, E., & Herin, B. (2011). The lifecycle of Qaf in Jordan. *Langage & Société*, 138, 59:76. DOI: 10.3917/ls.138.0059

Bauer, L. (2008). A question of identity: A response to Trudgill. *Language in Society* 37(2), 270–3. DOI: 10.1017/S0047404508080330

Britain, D. (1991). *Dialect and space: A geolinguistic study of speech variables in the Fens*. Unpublished Ph.D. dissertation, University of Essex.

Britain, D. (2001). Where did it all start?: Dialect contact, the 'Founder Principle' and the so-called (-own) split in New Zealand English. *Transactions of the Philological Society*, 99, 1–27. DOI: 10.1111/1467-968X.00072

Britain, D. (2002). Diffusion, levelling, simplification and reallocation in past tense BE in the English Fens. *Journal of Sociolinguistics*, 6(1), 16–43. DOI: 10.1111/1467-9481.00175

Cheshire, J., Kerswill, P., Fox, S., & Torgersen, E. (2011). Contact, the feature pool and the speech community: The emergence of Multicultural London English. *Journal of Sociolinguistics*, 15(2), 151–196. DOI: 10.1111/j.1467-9841.2011.00478.x

Cleveland, R. (1963). A classification of the Arabic dialects of Jordan. *Bulletin of the American Schools of Oriental Research*, 171, 56–63. DOI: 10.2307/1355607

Coupland, N. (2008). The delicate constitution of identity in face-to-face accommodation: A response to Trudgill. *Language in Society* 37(2), 267–270. DOI: 10.1017/S0047404508080329

Garbell, I. (1958). Remarks on the historical phonology of an eastern Mediterranean Arabic dialect. *Word*, 14, 303–337.

Giles, H., Taylor, D. M., & Bourhis, R. Y. (1973). Toward a theory of interpersonal accommodation through speech: Some Canadian data. *Language in Society*, 2, 177–192. DOI: 10.1017/S0047404500000701

Gordon, E., Campbell, L., Hay, J., Maclagan, M., Sudbury, A., & Trudgill, P. (2004). *New Zealand English: Its origins and evolution*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511486678

Hachimi, A. (2005). *Dialect leveling, maintenance and urban identity in Morocco*. Unpublished Ph.D. thesis, University of Hawaii.

Herin, B. (2010). *Le parler arabe de Salt: Phonologie, morphologie et éléments de syntaxe*. Unpublished Ph.D. dissertation, Université Libre de Bruxelles.

Hernandez-Campoy, J. M. (2003). Exposure to contact and the geographical adoption of standard features: Two complementary approaches. *Language in Society*, 32(2), 227–255. DOI: 10.1017/S0047404503322043

Holmes, J. & Kerswill, P. (2008). Contact is not enough: A response to Trudgill. *Language in Society*, 37(2), 273–277. DOI: 10.1017/S0047404508080342

Kerswill, P. (1993). Rural dialect speakers in an urban speech community: the role of dialect contact in defining a sociolinguistic concept. *International Journal of Applied Linguistics*, 3, 33–56. DOI: 10.1111/j.1473-4192.1993.tb00042.x

Kerswill, P., & Williams, A. (2000). Creating a new town koine: children and language change in Milton Keynes. *Language in Society*, 29, 65–115. DOI: 10.1017/S0047404500001020

Mufwene, S. S. (2008). Colonization, population contacts, and the emergence of new language varieties: A response to Peter Trudgill. *Language in Society* 37(2), 254–258. DOI: 10.1017/S0047404508080299

Palva, H. (1994). Bedouin and sedentary elements in the dialect of Es-Salt: diachronic notes on the sociolinguistic development. In D. Caubet, & M. Vanhove (Eds.), *Actes des premières journées internationales de dialectologie Arabe de Paris* (pp. 459–469). Paris: Institut National des Langues et Civilisations Orientales.

Salibi, K. (1993). *The modern history of Jordan*. London: I.B. Tauris & Co.

Schneider, E. W. (2008). Accommodation versus identity? A response to Trudgill. *Language in Society* 37(2), 262–267.

Siegel, J. (1985). Koines and koineization. *Language in Society*, 14(3), 357–378. DOI: 10.1017/S0047404500011313

Thelander, M. (1979) Språkliga variations modeller tillämpade på nutida Burträsktal. Stockholm: Almqvist & Wiksell..

Trudgill, P. (1983). *On dialect*. Oxford: Basil Blackwell.

Trudgill, P. (1986). *Dialects in Contact*. Oxford: Basil Blackwell.

Trudgill P. (2004). *New-dialect formation: The inevitability of colonial Englishes*. Edinburgh: Edinburgh University Press.

Trudgill, P. (2008a). Colonial dialect contact in the history of European languages: On the irrelevance of identity to new-dialect formation. *Language in Society* 37(2), 241–254.

Trudgill, P. (2008b). On the role of children, and the mechanical view: A rejoinder. *Language in Society*, 37(2), 277–280.

Tuten, D. N. (2008). Identity formation and accommodation: Sequential and simultaneous relations. *Language in Society*, 37(2), 259–262. DOI: 10.1017/S0047404508080305

Syntax

Prosodic constituency and locality in Levantine Arabic

Long-distance negative concord

Frederick M. Hoyt
University of New England

This paper examines negative concord sentences in Southern Levantine Arabic (Palestinian and Jordanian), providing evidence that locality restrictions on negative concord licensing are in fact restrictions on prosodic rather than syntactic locality. While negative concord is generally a clause-local dependency, a set of exceptions is examined in which the licensing relationship crosses subordinate clause boundaries. These examples involve a set of subordinating verbs with a high frequency in the Maamouri, Buckwalter, Graff, & Jin (2006a,b) corpus. Acoustic analysis of these data shows a strong correlation between the frequency of a subordinating verb in the corpus, its acceptability with non-local negative concord and reduced prosodic prominence in its pronunciation. This suggests that non-local negative concord licensing correlates with a subordinating verb structure being pronounced as a single prosodic constituent.

Keywords: negative concord, locality restrictions, prosodic locality, subordinating verbs, Southern Levantine Arabic

1. Introduction

This paper explores the relationship between syntactic constraints on negative concord in Southern Levantine and the prosodic properties of negative concord sentences.[1] Negative concord is well-known from the Romance languages, Slavic and others, but also occurs in Southern Levantine sentences such as in (1a):

1. I thank the organizers and participants of ALS 26. Particular thanks must go to Reem Khamis-Dakwar and Enam al-Wer for their comments and encouragement and to an anonymous reviewer for remarkably detailed and supportive notes on a draft of this paper.

(1) a. ***maː**=ʔakalt* ***wala*** *iši* *l=yoːm.*
not=ate.1s not.even thing the=day
"I didn't eat even one thing today," "I didn't eat a single thing today."
(Elicited datum)

b. ***maː**=ʔakalt* ***aiy*** *iši* *l=yoːm.*
not=ate.1s any thing the=day
"I didn't eat ANYthing today." (Elicited datum)

The sentence contains a sentential negation morpheme *maː-* "not" and a negative scalar focus particle *wala* "not even one, not a single." It appears to contain two negation morphemes but has an interpretation equivalent to (1b) containing only one negation morpheme.

The paper begins with examples of negative concord sentences in which an "n-word" inside a subordinate clause can be licensed by a negation clause in a higher clause (*long-distance negative concord*, or LDNC). LDNC appears to be an idiosyncrasy of certain subordinating verbs from different syntactic and semantic categories, as shown in the following examples:

(2) a. *ʔana* ***miš*** *ʕaːrif [afham*
I not knowing 1s.understand
wala *kilme min kalaam=ak.].*
not.even word.FS from speech=2MS
"I can't understand even one word of your speech." (Elicited datum)

b. ***maː**=ḥaːwalt [in=ni aḥki* ***wala*** *maʕ ḥada].*
not=tried.1s that=1s 1s.speak not.even with one
"I didn't try to talk even with even one person." (Elicited datum)

c. ***maː**=b=afakkir [inn=ha bitḥibb*
not=IND=1s.think that=3FS IND=3FS.like
wala *waːḥad min=hum.]*
not.even one from=3p
"I don't think that she likes even one of them." (Elicited datum)

The verbs in question are shown to be syntactically and semantically heterogeneous and therefore not a natural class in grammatical terms. However, what they do have in common is that they occur with a high frequency in the Linguistic Data Consortium (LDC) Levantine Call-Home corpus (Maamouri, Buckwalter, Graff, & Jin, 2006a & 2006b), a corpus (810,324 words) of Levantine Arabic speech (including Jordanian, Lebanese, Palestinian and Syrian data). This suggests that frequency of occurrence may have something to do with these verbs' transparency to long-distance negative concord.

I review a set of related generalizations about data in languages such as Italian, German and Japanese.[2] According to these, the scopes with which focus-sensitive items are interpreted correspond to the prosodic size of the constituents with which they combine. I follow Yamashita (2008) in referring to these generalizations collectively as the *Prosody-Scope Correspondence*. Sentences in which the Prosody-Scope Correspondence is observed are pronounced with focus intonation consisting of pitch peaks on the focal item and on its licensor or associate and with a region of reduced prominence (Poser, 1984; Selkirk & Tateishi, 1991, Ishihara, 2003, 2007; *inter alia*) between them.

With this in mind, I examine the sentences in the LDC corpus that show long-distance negative concord and show that the focus-intonation pattern can be observed in them, suggesting that the Prosody-Scope Correspondence is a property of Southern Levantine Arabic. The paper concludes with a discussion what is to be done to confirm the hypothesis and to further the study of intonational phonology in Levantine Arabic (El-Hassan, 1990; Chahal, 1999, 2001; Kulk, Odé, and Woidich, 2003) and other dialects (Abdalla, 1960; Hellmuth, 2006, 2011).

The paper is organized as follows: Section 2 provides a brief overview of negative concord sentences in Levantine Arabic. Section 3 introduces long-distance negative concord. Section 3.1 presents a range of verbs allowing long-distance negative concord and 3.2 discounts the possibility of treating it as an instance of *restructuring*. Section 4 presents the main hypothesis of the paper, that long-distance negative concord in Southern Levantine Arabic is subject to a locality restriction defined in terms of prosodic constituency and, in particular, that the constituent with which an n-word combines must be pronounced with an intonation melody consistent with focal backgrounding. The section begins in 4.1 with a review of the literature on prosodic locality in languages such as Italian and German in 4.1.1 and Japanese in 4.1.2. In 4.2 the generalizations reviewed in 4.1.1 and 4.1.2 are extended to Arabic. Section 5 concludes.

2. Negative concord in Southern Levantine Arabic

Southern Levantine Arabic is one of the many languages in which negative concord takes place (Hoyt, 2006, 2010; Lucas, 2009; Al-Sarayreh, 2012), where negative concord is understood according to the following definitions (Giannakidou, 2000; Watanabe, 2004):

2. See Hasegawa (1994), Deguchi & Kitagawa (2002), Ishihara (2003, 2007), Kitagawa & Fodor, (2003), Sugahara (2003), Blaszczak & Gärtner (2005), Kitagawa (2005), Kitagawa & Fodor, (2006), and Yamashita (2008) among others.

(3) Negative Expression: An expression the interpretation of which necessarily entails the meaning of predicate negation.
(4) N-word: A negative expression that can be used as a fragment answer.
(5) Negative concord: The failure of an n-word *X* to be interpreted as contributing negative meaning when in syntagm with another negative expression *N*. We say that *N* licenses *X* .

N-words in Southern Levantine Arabic

The inventory of n-words in Southern Levantine Arabic according to these definitions includes:

(6) a. The negative scalar focus particle *wala* "not even (one), not a single";
b. The homophonous additive particle *wala* "nor";
c. The "never-words" *ʔabadan*, *bilmarra* "never, not once, not at all";
d. The negative minimizer *hawa* "nothing" (lit. "air").

Of these, noun phrases and prepositional phrases prefixed with *wala* ("*wala*-phrases") have the widest syntactic distribution as they have both argumental and adverbial uses:

(7) a. Noun phrases: *wala iši* "not one thing," *wala ḥada* "not one person," *wala marra* "not even once," *wala nitfe* "not one bit," etc.
b. Prepositional phrases: *wala maʕ ḥada* "not even with one person," *wala la=waːḥad* "not even to one person," *wala b=iši* "not even with one thing," etc.

For this reason the following discussion focuses on scalar-*wala*.

The form *wala* has several homophonous uses, including "and not," "nor" and others. These are separate lexical items, given that they can co-occur with negative-scalar *wala* (for detailed discussion see Hoyt 2010):

(8) a. *l=yo:m* ***wala*** *ʔakalt* ***wala*** *iši.*
the=day not ate.1s not.even thing
"Today I didn't actually eat a single thing." (Elicited datum)

In sentences in which scalar-*wala* is subject to the licensing requirement, morphemes which are acceptable licensors for *wala*-phrases include the following:

(9) Sentential negation morphemes: *ma:-*, *ma:-…-š*, *-š*, *miš/mu*, *ma:ni/mani:š*, etc.

(10) *bidu:n* "without," *bidu:n-ma* "without (doing)":

a. *bniṭḥan=o ʔawwal marra l=ḥaal=o*
ind=1s.grind=him first time to=self=his
bidu:n *wala iši min l=ʔiḍafaat.*
without not.even thing from the=additives
"We grind it the first time by itself, without a single one of the additives." (Elicited datum)

b. *ke:f b=aχally šabb yiʕtarif inno b=yiḥibb=ni*
how IND=1s.let boy 3MS.avow that IND=3MS.love=me
bidu:n=ma *ʔaḥki maʕ=o wala kilme?*
with=that 1s.talk with=3MS not.even word
"How do I let a boy say that he loves me without my having spoken a single word with him?" (Elicited datum)

(11) *qabl* "before," *qablma* "before (doing):

a. *ʔana ḥammalit kull il=fayru:saat illi ʔinta ḥaaṭṭ=ha*
I load.1s all the=viruses rel you.MS put=3FS
gabil *wala waḥde štaɣalat.*
before not.even one.FS worked.2FS
"I downloaded all the viruses that you uploaded before a single one ran." (Elicited datum)

b. ***gabilma*** *ygu:l wala kilme gaalat=l=o*
before.that 3MS.say not.even word said.3FS=to-3MS
ʔanqað=ni w=b=aʕṭi:=k bo:se.
save=1s and=IND=1s.give=3MS kiss
"Before he said a single word, she said to him 'Save me and I'll give you a kiss.'" (Elicited datum)

(12) Subordinating verbs that entail the negation of their complements:

a. *manaʕ-yimnaʕ* "forbid, prevent (someone from doing)"
manaʕ *wala wa:ḥad yiftaḥ is=sanduuq.*
forbade not.even one 3MS.open the=box
"He forbade even one person to open the box." (Observed datum)

b. *baṭṭal-ybaṭṭil* "stop, cease, quit (doing)"
χalaaṣ, ***baṭṭalt*** *aḥky wala kilme.*
finished, stopped.1s 1s.say not.even word
"Fine, I have stopped saying a single word." (Elicited datum)

c. *rafaḍ-yurfuḍ* "refuse (to do)";
bess ʔana ***rafaḍt*** *aakil wala gaṭʕa.*
but I refused.1s 1s.eat not.even piece
"...but I refused to eat a single piece." (Elicited datum)

The bold-faced expressions in (9–12) are all interpreted as anti-morphic or, equivalently, anti-veridical operators, in that they are equivalent in meaning to classical negation.[3] *Wala*-phrases cannot be licensed by anti-additive or "merely" downward entailments, which are able to license negative polarity interpretations for words such as the following:[4]

(13) a. *ʔaiy* emphatic "any" (c.f. English emphatic "ANYthing");
b. *iši* (Jordanian/Palestinian), *šiː* (Syrian/Lebanese) "(one) thing, anything";
c. *ḥada*, *waːḥad* "(one) person, anyone";
d. *ʕumr* "ever"

Anti-additive or merely-downward-entailing contexts include the following (see Hoyt 2010, 130–132 for detailed examples):

(14) a The scope of pre-verbal *wala*-phrases
b. Comparative adjectives
c. Questions
d. Antecedent clauses of conditional sentences
e. Downward-Entailing Quantifiers (*kull* "each, every, all"; *qali:l* "few")

As indicated in the glosses given above in (2) and in what follows, *wala* is glossed variously as "not even one," "not one" or "not a single." In theoretical terms, it is a negative scalar focus particle,[5] interpreted as follows:

(19) a. It selects or associates with a singular indefinite NP: *wala iši* "not even one thing, not a single thing" vs. **wala ʔašyaː* "not even things";

3. An antimorphic operator is an operator that is both anti-additive and anti-multiplicative, meaning that both of DeMorgan's Laws apply to it:

(1) i. $OP\,(p \wedge q) \Leftrightarrow OP\,(p) \vee OP\,(q)$ (Anti-additivity)
ii. $OP\,(p \vee q) \Leftrightarrow OP\,(p) \wedge OP\,(q)$ (Anti-multiplicativity)

See Zwarts (1996) and Wouden (1994) for discussion of antiadditive and antimorphic operators. An anti-veridical operator is one for which the following inference holds:

(2) $OP\,(p) \Rightarrow \neg p$

See Zwarts (1995) and Giannakidou (1997, 1998) for discussion.

4. A downward entailing operator is one for which the following entailment holds:

(1) $P \subseteq Q$ and $\neg Q(x) \Rightarrow \neg P\,(x)$

Anti-additive and antimorphic operators are necessarily also downward entailing.

5. See Rooth (1992), Krifka (1995b), Israel (1996, 2001), Rullmann (1996), and Lahiri (1998) among many others.

b. It triggers a set of focal alternatives ranging over (non-null) cardinality values;
{I ate n things: $n \geq 1$};

c. It negates the minimum alternative in this set and implicates or entails negation of all higher alternatives:
{I didn't eat one thing and I didn't eat n things $n \geq 1$}

For example, in (1) above, *maː = ʔakalt wala iši waːḥad l = yoːm* "I didn't eat even one thing today" *wala* associates with the singular indefinite noun phrase *iši* "(a) thing," triggering a set of alternatives {I ate n things today: $n \geq 1$} and asserting that the speaker didn't eat one thing and also didn't eat any number of things greater than one. This follows standard analyses of focus semantics and the meaning of English *even* and its translation equivalents in various languages (see references cited above for discussion).

Typically, *wala*-phrases are pronounced with a strong stress accent on the first syllable of *wala* and with a strong accent on the most prominent syllable of the common noun with which it associates. In other words, *wala waːḥad* in (1) above would be pronounced as *WA.la WAː.ḥad* (with capitals indicating strong accentuation). This suggests that *wala*-phrases are typically pronounced with strong focal accentuation, although it is not clear that they necessarily do so.

An overview of negative concord

As shown in Hoyt (2010), *wala*-phrases occur in many syntactic configurations and are subject to the licensing requirement in only some of those. In brief, *wala*-phrases at the left edge of a clause need not be licensed and do not undergo negative concord.

(20) a. ***wala waːḥad*** *min=ku b=ifham=ni.*
not.even one from=you.MP IND=3MS.understand=me
"Not a single one of you understands me." (Elicited datum)

b. ***wala ktaːb*** *ʕirifit miːn kæːn illi katab=u.*
not.even book knew.1s who was rel wrote=3MS
"Not one book did I know who it was who wrote [it]." (Elicited datum)

Native speakers generally express a strong preference for an n-word following the predicate to co-occur with negation marking on that predicate (21a), indicating that, in the absence of negation marking on the predicate, the sentence is unacceptable (21b). The constrast in (21a-b) shows the typical pattern of negative concord sentences in Spanish, Italian, Romanian, etc.:

(21) a. ***ma:**=kalt* ***wala*** *iši* *l=yo:m.*
not=ate.1s not.even thing the=day
"I didn't eat a single thing today." (Elicited datum)

b. **ʔakalt* ***wala*** *iši* *l=yo:m.*
ate.1s not.even thing the=day
"I ate not a single thing today." (Elicited datum)

In the acceptable example (21a) the *wala*-phrase *wala iši* "not even one thing, not a (single) thing" undergoes negative concord with and hence licensed by the negation morpheme *ma:*- "not" on the clausal predicate *akalt* "I ate." In contrast, the unacceptable example in (21b) shows *wala iši* occurring without negation-marking on the verb and is hence unlicensed.[6]

Negative concord licensing and locality

Negative concord in Southern Levantine Arabic (Blau, 1960; Cowell, 1964; Hoyt, 2006, 2010; Lucas, 2009) is generally a clause-local relation: It is only acceptable between a negation morpheme preceding the clausal predicate and an n-word that is a dependent of the same clause. Native speakers generally reject sentences in which an n-word inside a subordinate clause or noun phrase is licensed by a negation morpheme scoping over it, as in the following schema:

(22) NEG ...V1 ...[IP/NP ...*wala*-NP ...]

For example, licensing fails when a *wala*-phrase is inside a relative clause (23a), inside a construct-state noun phrase (24a) or inside a subordinate clause (25a):

(23) *Inside Relative Clause*:
a. ****ma:**=fi:* *ḥada* *[*RC *ʕind=u* ***wala*** *maʕlu:ma].*
not=exist one at=3MS not.even information
"There isn't anyone who has even one bit of information."
(Elicited datum)

b. ***ma:**=fi:* *ḥada* *[*RC *ʕind=u* ***ʔaiy*** *maʕlu:ma* *].*
not=exist one at=3MS any information
"There isn't anyone who has ANY bit of information."
(Elicited datum)

(24) *Inside construct state NP:*
a. ****ma:**=šuft* *[*NP *walad* ***wala*** *wa:ḥad* *min=hum].*
not=saw child not.even one from-them
"I didn't see the child of even one of them." (Elicited datum)

---

6. Exceptions to this generalization do arise (as detailed in Hoyt 2010) but are not relevant to the present discussion.

b. ***ma:***=*šuft* [NP *walad* ***ʔaiy*** *wa:ḥad min=hum].*
not=saw child any one from=them
"I didn't see the child of any one of them." (Elicited datum)

(25) *Inside Subordinate Clause:*
a. ****maː***=*waʕatt* [IP *aḥki maʕ* ***wala*** *waːḥad min=hum].*
not=promised.1s 1s.speak with not.even one from=3p
"I didn't promise to speak with a single one of them."
(Elicited datum)

b. ***maː***=*waʕatt* [IP *aḥki maʕ* ***ʔaiy*** *waːḥad min=hum].*
not=promised.1s 1s.speak with any one from=3p
"I didn't promise to speak with ANY of them." (Elicited datum)

As shown, all of these examples have acceptable paraphrases with the negative-polarity-sensitive particle *ʔaiy*, translatable as the emphatic use of English "any" (glossed as "ANY" in all-caps by Kadmon and Landman, 1993; Krifka 1995a) in place of *wala*. However, in some sentences, n-words in the complements of certain subordinating verbs can in fact be licensed by main-clause negation. These include *bidd-* "want" (26a), *ḥaːwal-yḥaːwil* "try" (26b), *fakkar-yfakkir* "think" (26c) and *qaːl-yquːl* "say" (26d):

(26) a. *biddi:=š* [IP *aḥki* ***wala*** *maʕ wa:ḥad fi:=hum].*
want.1S=NEG 1s.speak not.even with one in=3p
"I don't want to speak even with one of them." (Elicited datum)

b. *ʕumr* ***maː***=*ḥa:walti* [IP *tiḥki* ***wala*** *maʕ ḥada fi:=hum].*
Ever not=tried.2FS 2MS.speak not.even with one in=3p
"You didn't ever try to speak even with one of them." (Elicited datum)

c. ***maː***=*b=afakkir* [CP *inn=ha bi=thibb* ***wala hada*** *fi:=hum.]*
not=IND=1s.think that=she IND=2FS.like not.even one in=3p
"I don't think that she likes even one of them." (Elicited datum)

d. ***ma:***=*ʔult* [CP *ʔin=ny ḍadd=kum fi=****wala*** *šiy ʔult=u].*
not=said.1s that=1s against=2p in=not.even thing said.2p=3MS
"I didn't say that I was against you in even one thing you said [it]."
(Elicited datum)

I refer to these apparent exceptions to the locality of negative concord as *long-distance* negative concord (LDNC), where "long-distance" is intended in contrast to "clause-local".[7]

7. *Long-distance negative concord* is a term used by Piñar Lurrubia (1996); Przepiórkowski & Kupść (1997a); Matos (1999); among others.

Which verbs allow LDNC?

Not all subordinating verbs allow LDNC. Examination of elicited and corpus data shows that long-distance negative concord most typically occurs with a limited set of subordinating verbs:

(27) *bidd-* "want," *qidir-yiqdar* "can, be able," *ʕirif-yiʕraf* "be able, know how to," *ḥaːwal- yḥaːwil* "try," *χalla-yχalli* "let do, make do, have do," *læːzim* "must, have to, necessary," *mumkin* "can, might, possible," *qaːl-yquːl* "say," *fakkar-yfakkir* "think, believe," *kæ:n-yiku:n* "be," *ṣa:r-yṣi:r* "become," *rijiʕ-yirjaʕ* or *ʕaːwad-yʕaːwid* "return, do again," etc.

Some of these are auxiliaries (*læːzim* "must, have to, necessary," *mumkin* "can, might, possible, *kæːn-kæːn* "be," *ṣa:r-yṣi:r* "become," *rijiʕ-yirjaʕ* or *ʕaːwad-yʕaːwid* "return, do again"), and are expected to be transparent to local syntactic dependencies. The others are Arabic analogues of verbs that allow long-distance negative concord in other languages (see references above).

To investigate which verbs are transparent to LDNC, an experiment was done in Irbid, Jordan in December 2007 with four native speakers from a village in the rural northern region of the Irbid Governate. They were between 25 and 30 years of age, had bachelor degrees from Jordanian universities, were from the same clan and spoke essentially the same local dialect. They were shown lists of sentences containing *wala*-phrases, *ʔaiy*-phrases and bare indefinites within the scope of a matrix negation morpheme and all within the complement of a subordinating verb:

(28) NEG V1 ...[V2 ...*wala* NP ...]

The speakers were presented with discourse contexts in which the sentences might be uttered and were asked to grade the acceptability of the sentences in these contexts using magnitude estimation (Bard, Robertson, & Sorace, 1996; Cowart, 1997; Keller, 2000; Featherston, 2005). The verbs used in constructing the sentences were taken from the Linguistic Data Consortium Levantine Call-Home corpus (Maamouri et al., 2006 a, b), a corpus of 810,324 words.

Table 1 shows the frequency of subordinating verbs in the LDC corpus (in terms of overall numbers) with their average acceptability with long-distance negative concord for the four speakers (as a z-score). The tale shows that the verbs with average acceptability z-score (–1.27 or higher) are a proper subset of the more frequent verbs in the corpus (shown in italics).

The reader should note that there is a great deal of variation in native speaker judgments regarding the acceptability of LDNC, both across speakers and longitudinally for individual speakers. That being said, a strong correlation appears to exist between the relative frequency of subordinating verbs in the corpus and their acceptability with LDNC. The question is therefore: What (if anything) does frequency have to do with transparency to negative concord licensing?

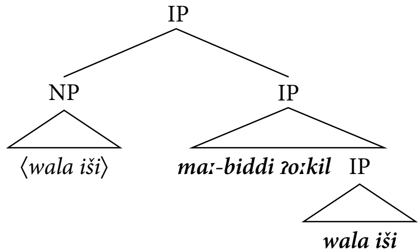
Table 1. Subordinating Verb Frequency and LDNC z-score for Maamouri et al. (2006b).

Verb	Gloss	Frequency (out of 810,324 words)	Acceptability w/LDNC (avg. z-score)
bidd-	want	7417	1
qaːm	stand	2364	0.28
ʕirif	be able	12125	0.25
χalla	let	2726	0.15
rajaʕ	return	1141	0.14
kaːn	be	9483	−0.08
qidir	be able	793	−0.08
šaːf	see	915	−0.13
laːzim	must	829	−0.55
twaqqaʕ	believe	55	−1.01
ansa	forget to	406	−1.01
kirih	hate to	135	−1.01
mumkin	can	586	−1.11
ḥabb	like to	5700	−1.27
šakk	doubt	101	−1.5
ʕirif	know that	12125	-1.5
nakar	deny	14	−1.5
manaʕ	prevent	277	−1.5
ḥaːwal	try	140	−1.5
jabbar	make do	84	−1.5
naṣaḥ	advise	108	−1.5
tðakkar	remember to	120	−1.5
samaḥ	allow	111	−1.5
qarrar	decide to	3	−1.5
tjannab	avoid	3	−1.5
ʕazam	invite	69	−1.5
χaːf	fear to	360	−1.5
simiʕ	hear	3019	−1.5
qaːl	say	6072	−1.5
tnaddam	neglect to	28	−1.5
iʕtaraf	admit	5	−1.5
waʕad	promise to	54	−1.5

LNDC as syntactic movement?

A popular analysis of long-distance negative concord is that an n-word in a subordinate clause undergoes syntactic movement out of the subordinate clause, adjoining to the clause containing its licensor. This allows the licensing mechanism (however that may be analyzed in particular proposals) to be established locally.[8] Long distance negative concord can therefore be treated as a special case of local negative concord and so allowing for a unified analysis.

(29) a. *maː=biddi ʔoːkil wala iši.*
not=want.1s 1s.eat not.even thing
"I don't want to eat even one thing."

b.


A movement analysis of negative concord makes incorrect predictions regarding "split-scope" interpretations that *wala*-phrases can have (see Hoyt, 2010, for detailed discussion).

Another possibility is that LDNC in Southern Levantine is a kind of *restructuring*. This is suggested by the observation that the verb meanings associated with high LDNC-ACCeptability are familiar from the literature on "restructuring" or "complex-predication formation," familiar symptoms of which are "clitic-climbing" or auxiliary selection in the western Romance languages, long-distance scrambling in the western Germanic languages, or long-distance agreement in Hindi-Urdu. While analyses differ in their details, the intuition they try to capture is that restructuring is a subordination structure analyzed grammatically as a single clause.

Long-distance negative concord has been analyzed as restructuring in a number of languages, including Slavic languages such as Polish and Serbo-Croatian (Progovac, 1993; Dziwirek, 1998, *inter alia*). For example, the following Polish examples show that the n-word *nikogo* "no one" can be licensed inside an infinitival complement but not inside a subjunctive clause:

8. See Rizzi (1978), Aissen & Perlmutter (1983), Miyagawa (1987), Bayer & Kornfilt (1989), Butt (1995), Dziwirek (1998), Andrews & Manning (1999), Cinque (2001), Wurmbrand (2001, 2005), Chung (2004), Stejapanović (2004), and Hoyt (2006) among many others.

(30) a. **Janek* ***nie*** *powedział* [subj *ze kocha* ***nikogo***].
Janek not said that love no-one
"Janek didn't say that he loved anyone."

b. *Janek* ***nie*** *kazał Ewie* [inf *zwrócić się do*
Janek not ordered Eve-DAT turn-infin ref to
nikogo *o pomoc*].
no-one for help
"Janek didn't tell Eve to turn to anyone for help."

These analyses suggest the possibility that long-distance negative concord in Southern Levantine might be an instance of restructuring. This was explicitly argued by Hoyt (2006). However, corpus data and fieldwork conducted in Jordan in 2007–2008 suggests that things are not so clear. In particular, there is much variation both in native speaker judgments regarding the acceptability of LDNC as well as in the classes of verbs with which it is acceptable.

This suggests that the verbs that allow LDNC cannot be defined as a semantic natural class. They also cannot be defined as a syntactic natural class as they vary in terms of how much structure they allow in the subordinate clause, as shown in (2) above. Some (such as *bidd-* "want" or *ʕirif* "know, be able" allow only bare verbal complements while others (such as *ḥaawal* "try" or *ʔiʕtaqad* "believe") allow subordinating conjunctions in LDNC sentences. This suggests that, contra Hoyt (2006), restrictions on long-distance negative concord are not a grammatical matter. The following theoretical questions therefore arise: (i) why is there so much inter-speaker variation in terms of the verbs that allow LDNC, and (ii) why does the acceptability of LDNC seem to correlate so strongly with frequency?

Prosodic locality

I explore the possibility that prosodic constituency may play an important role in LDNC, and in particular that the verbs which allow it are also verbs which can undergo some degree of prosodic reduction (Monachesi, 2005) or are at least more susceptible to pitch weakening than are other verbs. I build on a claim by Blaszczak & Gärtner (2005) that the scopal domains of n-words in Italian and German correlate with prosodic contiguity of the constituents with which they combine. I hypothesize that Southern Levantine Arabic n-words are frequently pronounced with contrastive focus and, as such, must combine with a constituent containing a licensor and which is pronounced with some degree of reduced prominence, as is characteristic of constituents pronounced as "background" to a focal constituent.

The connection between LDNC and high-frequency verbs then might be explained in terms of prosodic reduction of high-frequency lexical items (Heine, 1993; Bybee & Schiebman, 1999). I present examples of LDNC in audio data from Maamouri et al. (2006a, 2006b) and annotated for intonational constituency (Hellmuth, 2006) supporting the generalization.

Prosodic conditions on scope interpretation

Before going further with Southern Levantine Negative Concord, I briefly review work on relationships between prosodic constituency, focus and prosodic locality in other languages.

Condition on extended scope taking

Blaszczak & Gärtner (2005) argue that Condition on Extended Scope Taking (Kayne, 1998) is a prosodic effect. One instance of this involves negative concord sentences in Italian as shown in (31). The generalization is that an Italian n-word such as *nessuno* "no one" must be separated from its licensing negation morpheme by a contiguous string of words (shown in brackets subscripted with σ) – including the verb of which it is an argument – in order to have the negative concord interpretation (31a). If the word order is changed so that the verb of which the n-word is an argument is not part of the contiguous string, then the sentence can only have a "double negation" interpretation (31b):

(31)	a.	*(σ* ***non***	*voglio*	*che*	*venga)*	***nessuno****.*		
		Not	want.1s	that	come	no-one		
		"I want someone to come."						Negative Concord
	b.	*(σ* ***non***	*voglio*	*che*	***nessuno***	*(σ*	*venga)*	
		not	want.1s	that	no-one		come	
		"I don't want no one to come."						
		"I don't want anyone to come."						Double Negation

Likewise, German n-words such as *niemanden* "no one" can be interpreted as taking scope only over contiguous constituents. For example, in (32a) *niemanden* has two scope interpretations: one in which it scopes over the constituent *zu grüssen versprach* "promised to greet" with the meaning "she did not promise to greet anyone," and another in which it takes scope only over "to greet," meaning "she promised not to greet anyone." In (32b), however, *niemanden* is separated from *zu grüßen* "to greet" and only has the narrow scope reading:

(32) a. *daß sie* ***niemanden*** *(σ zu grüßen versprach)*
That she no-one to greet promised
"...that she did not promise [to greet anyone]."
"...that she promised [not to greet anyone]."

b. *daß sie* ***niemanden*** *(σ versprach) (σ zu grüßen)*
that she no-one promised to greet
"...that she promised [not to greet anyone]."

Blaszczak & Gärtner's (2005) generalization is therefore that n-words in Italian and German are interpreted as taking scope over constituents that are pronounced as a single prosodic unit.

Similar generalizations involving question words and some negative polarity items are found in Japanese.[9] In Japanese constituent questions, a question word such as *dare* "who" or *nani* "what" must be licensed by a question particle. The following sentence from Yamashita (2008) contains the question particles *ka* (in a subordinate clause) and *no* (in the main clause) and the question word *nani* "what" (question words and particles are indicated in boldface):

(33) a. *Naoya*=GA *Mari*=GA ***nani=o*** *(σ nomiya=de non=da* ***ka)***
N.=NOM M=NOM what=ACC bar=LOC drink=PAST Q
Yumi=NI *tsutae*=TA ***no?***
Y.=DAT tell=PAST Q
"Did Naoya tell Yumi [what Mari drank at the bar]?"

b. *Naoya=ga Mari=ga* ***nani=o*** *(σ nomiya=de*
N.=NOM M=NOM what=ACC bar=LOC
non=da ka Yumi=ni tsutae=ta) ***no?***
drink=PAST Q Y.=DAT tell=PAST Q
"What was it that Naoya told Yumi [whether Mari drank it at the bar]?"

The question word that falls within the scope of two question particles (one of which is subordinate to the other) can be interpreted with scope associated with either. For example, if *nani* in (33a) and (33b) is associated with *ka* in the embedded clause, the sentence is interpreted as a yes-or-no question (33a), while if *nani* is associated with the *no* in the main clause, the sentence is interpreted as a constituent question (33b).

In either case, the sentence is pronounced with what Ishihara (2007) calls a focus intonation pattern, which consists of the following:

9. See Deguchi & Kitagawa (2002), Ishihara (2002, 2003, 2005, 2007), Kitagawa & Fodor, (2003, 2006), Hirotani (2005), Kitagawa (2005) and Yamashita (2008).

(34) a. A pitch excursion (or peak) in the F0 with which the focused constituent (in this case, a question word) is pronounced;
b. Pitch compression or downtrend in the F0 with which the words following the focused phrase are pronounced (Poser, 1984; Pierrehumbert & Beckman, 1988; Selkirk & Tateishi, 1991; Sugahara, 2003, *inter alia*);
c. Pitch reset on the particle or morpheme with which the focus associated (in this case, a question particle).

Likewise, the exclusive particle *shika* "only" has to be licensed by a negation morpheme:[10]

(35) a. *John=ga Mary=to=**sika** awa=**nakat**=ta.*
John=NOM Mary=with=NPI meet=NEG=TNS
"John met only Mary."
"John didn't meet [anyone] but Mary."

b. **John=ga Mary=to=**sika** at=ta.*
John=NOM Mary=with=NPI meet=TNS

The *shika*-phrase and its licensing negation usually must be in the same clause:

(36) **Bill=ga Pam=ni [John=ga Mary=to=**shika** atta=to]*
Bill=NOM Pam=DAT John=NOM Mary=with=only met=C
*tutae=**nakat**=ta.*
tell=NEG=TNS
"Bill only told Pam [that] John met Mary."

However, a *shika*-phrase can be licensed non-locally if and only if it occurs inside a non-finite control complement and if the non-finite complement is pronounced as a contiguous string with compressed pitch adjacent to the negation morpheme that licenses the *shika*-phrase (37b).

(37) a. *Naoya=wa Mari=ni* ***sono ramu=shika*** *(σ nomiya=de)*
Naoya=top Mari=DAT that rum=only bar=LOC
*noma=nai=yoo=ni iwa=**nakat**=ta.*
drink=not=tns=C tell=NEG=TNS
"Naoya didn't tell Mari to drink [only the rum] at the bar."
"It was only the rum that Naoya told Mary not to drink at the bar."

10. See Muraki (1978), Kato (1985), Hasegawa (1994), Aoyagi & Ishii (1994) and Hirotani (2005). Japanese *shika* resembles Arabic *illa* "only, except for, other than, but" in usage.

b. *Naoya=wa* *Mari=ni* ***sono ramu=shika*** *(σ nomiya=de*
Naoya=top Mari=DAT that rum=only bar=LOC
noma=nai=yoo=ni) *iwa=**nakat**=ta.*
drink=not=TNS=C tell=NEG=TNS
"Naoya didn't tell Mari to drink [only the rum] at the bar."
"It was only the rum that Naoya told Mary not to drink at the bar."

Based on these observations, Yamashita (2008) proposes the Prosody-Scope Correspondence:

(38) *The Prosody-Scope Correspondence*: The scope of a focal phrase is determined and indicated by the extent of the post-focal reduction in prominence between the phrase and the particle that licenses it.

The parallel with Blaszczak & Gaertner's (2005) generalizations above should be clear: constituents pronounced with focal intonation and with scopal interpretations take scope over a sister constituent that is pronounced as a continuous prosodic unit.

Prosodic locality in southern levantine LDNC?

I hypothesize that a similar generalization can be made about long-distance negative concord sentences in Southern Levantine: namely, that n-words must be local to their licensors in terms of prosodic constituency. As was discussed above (9), *wala* is interpreted with focal semantics and the noun phrase with which it associates is frequently if not always pronounced with at least some degree of focal prominence. Accordingly, the generalizations above predict that focal intonation on a *wala*-phrase will correlate with reduced prominence on the string of words separating the *wala*-phrase from the negation morpheme that licenses it:

(39) (σ NEG ...reduced prominence ...) *wala*-NP

Examination of a selection of negative concord sentences found in the Maamouri et al. (2006a,b) corpus appears to confirm the prediction. The audio segments for the sentences were extracted from the corpus and analyzed using the Praat software package and the ProsodyPro script, which extracted mean F0 and mean duration values for each prosodic word (i.e. each lexical word along with whatever clitics it hosts). The relative values for mean F0 and duration for the words in the sentence were then compared in order to determine:

(40) i. What the relative mean F0 and duration values were for the focused constituent (the *wala*-phrase) and its licensor (the negation morpheme);
ii. Whether words intervening between the licensor (and its lexical host) and the *wala*-phrase were pronounced with lower mean F0 or duration than were the licensor verb complex and the *wala*-phrase.

The following are the examples of LDNC found in (Maamouri et al., 2006 a,b). Each shows pairs of mean F0 and duration for each prosodic word (in the format F0/DUR). Words are grouped according to trends in the F0 and duration values with high values in bold[11]:

(41) a.

*(**b=yismaḥ=il=naː=š***	*inšuːf)*	*(wala*	*ʔišy.)*
IND=3MS.LET=to=1p=NEG	1p.SEE	not.even	thing)
334.2/.68	353.8/.32	354.9/.19	295.9/.42

“He doesn’t let us see even one thing.” (fla_0100: 467.8-471.18: Lev, F)

b.

*(maː=**biddy***	*nḍayʕ)*	*(wala*	*waʔt)*
not=want.1s	1p.lose	not.even	time
322.5/.47	295.1/.45	318.5/.29	324.1/.27

“I don’t want us to lose any time.” (fla_0107: 482.28-493.87: Lev,F)

c.

*(ma=**ḫaːwalt**=iš*	*tiʕtariḍ)*	*(wala*	*marra)?*
not=tried.2ms=NEG	2MS.resist	not.even	once
158.4/.61	157.9/.38	148.5/.28	154.4/.35

“You didn’t try to resist even once?” (fla_0247: 155.43-159.36: Lev,M)

d.

*(maː=**ḫaːwalt***	*itsakkir*	*il=χaṭṭ)*	*(wala*	*marra)?*
not=tried.2MS	2MS.close	the=line	not.even	once
128.2/.43	125.5/.36	125.2/.42	156/.17	131.1/.36

“You haven’t tried to hang up even once?”

(fla_0626: 459.46-464.06: Leb,F)

e.

*(maː=ʕam=**b=aʔdar***	*aʕmil)*	*(wala*	*šiː)*
not=prog=ind=1s.BE.able	1s.do	not.even	thing
211.9/.65	207.5/.44	245./.25	349.5/.23

“I’m not able to do a single thing.” (fla_1041: 97.22-107.87: Leb,F)

f.

(muš-mumkin	*tunʔuð)*	*(wala*	*waːḥad)*
not-possible	2ms.save	not.even	one
280.6/.66	259.5/.25	299.1/.25	253.7/.36

“You can’t save even one.” (fla_1139: 179.93-186.81: Leb,F)

g.

*(maː=**laːzim***	*nfariʔ)*	*(wala*	*waːḥad)*	*(min miyye)*
not=should	1p.leave	not.even	one	from hundred
126.5/.40	125.2/.34	151/.16	161.2/.25	293.4/.64

“We mustn’t leave even one out of a hundred.”

(fla_1524: 194.11-202.81: Leb,M)

---

11. Citations include: the name of the recording in which the datum was found; its time stamp within the file; the nationality of the speaker (where identified) and the speaker’s gender. The LDC data are transcribed impressionistically based on the audio for each example.

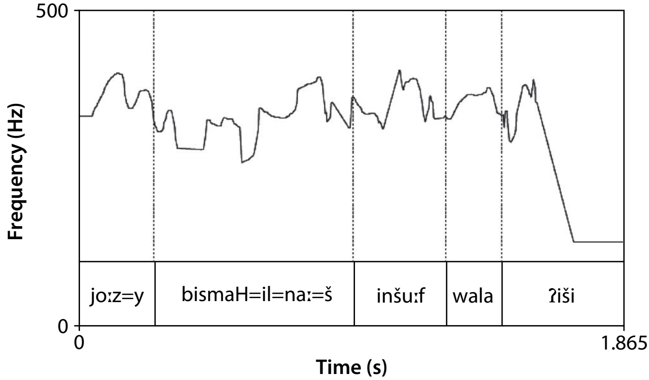


Figure 1. Pitch Track for (41)a.

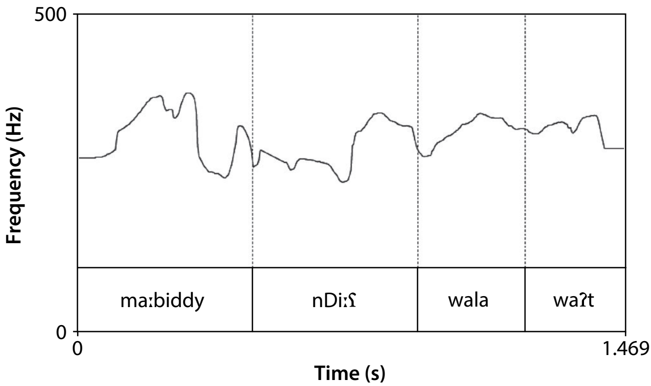


Figure 2. Pitch Track for (41)b.

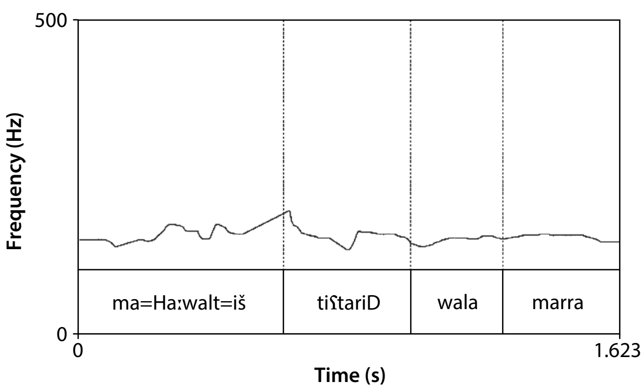


Figure 3. Pitch Track for (41)c.

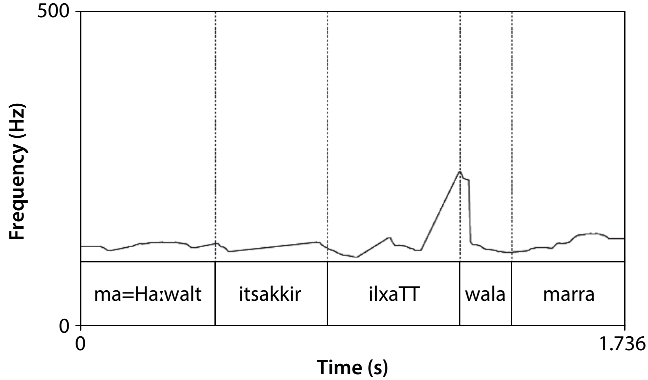


Figure 4. Pitch Track for (41)d.

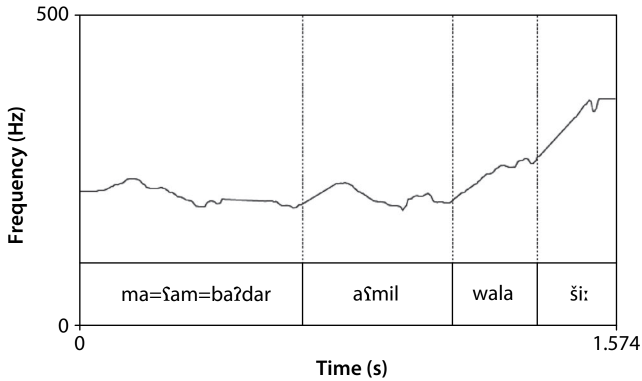


Figure 5. Pitch Track for (41)e.

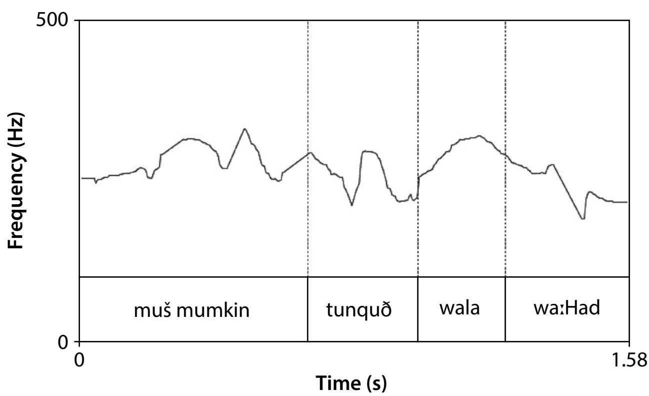


Figure 6. Pitch Track for (41)f.

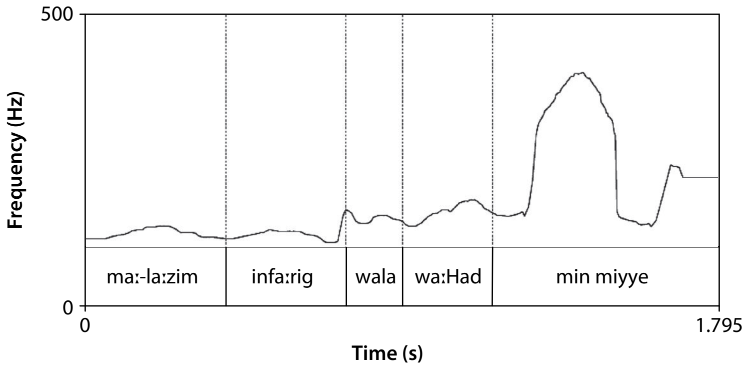


Figure 7. Pitch Track for (41)g.

Table 2. Mean F0 for examples (41)a–(41)g (in Hz: high values for each phrase in bold).

Prosodic Constituency and Locality in Levantine Arabic

Example	Source File	neg (+host)	embedded verb	wala	NP
(41a)	fla_0100: 467.8-471.18	334.2	**353.8**	**354.9**	295.9
(41b)	fla_0107: 482.28-493.87	**322.5**	295.1	318.5	**32.4.1**
(41c)	fla_0247: 155.43-159.36	**158.4**	157.9	148.5	**154.4**
(41d)	fla_0626: 459.46-464.06	**128.2**	125.5+125.2	**156**	131.1
(41e)	fla_1041: 97.22-107.87	**212**	207.6	245.5	**349.5**
(41f)	fla_1139: 179.93-186.81	**280.6**	259.5	**299.1**	253.7
(41g)	fla_1524: 194.11-202.81	**126.5**	125.2	151.0	**161.2**

Table 3. Mean duration for examples (41)a–(41)g (in seconds: high values in bold).

Example	Source File	neg (+host)	embedded verb	wala	NP
(41a)	fla_0100: 467.8-471.18	**.68**	.32	.19	**.42**
(41b)	fla_0107: 482.28-493.87	**.47**	.45	**.29**	.27
(41c)	fla_0247: 155.43-159.36	**.61**	.38	.28	**.35**
(41d)	fla_0626: 459.46-464.06	**.43**	.36+.42	.17	**.36**
(41e)	fla_1041: 97.22-107.87	**.65**	.44	.25	.23
(41f)	fla_1139: 179.93-186.81	**.66**	.32	.25	**.35**
(41g)	fla_1524: 194.11-202.81	**.40**	.34	.16	**.25**

The results (shown in Tables 2 and 3) show that mean pitch and mean duration reliably fall upon the prosodic word consisting of the licensing negation and the matrix subordinating verb. The only example in which a higher F0 mean occurs on the subordinate verb is (41a). However, in this example the negation-V1 complex is pronounced with more than twice the duration of the subordinate verb. This suggests that pitch and duration may work together in signaling degrees of relative prominence. The results are consistent with the prediction that the degrees of prominence on the words falling between the *wala*-phrase and its licensor are lower than the degrees with which the V1 complex or the *wala*-phrase are pronounced.

Is the reduction in mean F0 and duration on the subordinate clauses in the examples is an instance of reduced prominence? In the literature on languages with prominence (such as Japanese), the reduced prominence between a licensor and the question word or NPI it licenses is widely argued to be a prosodic constituent referred to as the Major Phrase or Intermediate Phrase (Poser, 1984; Pierrehumbert and Beckman, 1988; Selkirk and Tateishi, 1991, *inter alia*). If this is the true in Southern Levantine, then the generalization can be refined to the following:

(42) (MaP NEG ...reduced prominence ...) *wala*-NP

In other words, the locality condition would be that a *wala*-phrase must combine with a major phrase prosodic constituent containing its licensor.

One might ask whether the decreased prominence on the subordinate clauses involves de-accenting. Indeed, in (4) and (41) e., the subordinating verb *baʔdar* "I can" appears to lack a pitch accent altogether: The two syllables in the word are pronounced with peaks at almost the same pitch rather than with a higher pitch on the syllable that would typically be accented (the first in this case). The word therefore appears to be de-accented. De-accenting has also been identified in Lebanese Arabic by Mitchell (1993) and Chahal (1999, 2001).

However, while de-accenting (as in (41) e.) may be a sufficient condition for allowing long-distance negative concord, de-accenting appears to be entirely absent in Egyptian Arabic (Hellmuth, 2005, 2006, 2011), yet Egyptian also has long-distance negative concord:

(43) a. *ma=šuft=iš* [CP *inn=u* *kal* *wala* *rɣi:f].*
not=saw.1S=NEG that=3MS ate not.even loaf
"I haven't seen that he ate a single piece of bread." (Woidich, 1968, 153)

b. *ʔana* *miš* *ʕaawiz* [IP *titɣayyar* *wala* *ḥaaga].*
I not want 3FS.change not.even thing.FS
"I don't want a single thing to change."
(Internet datum (accessed 7/2012))

I conclude that de-accenting is not a necessary condition for LDNC.

Assuming that LDNC is subject to a prosodic locality condition; and that the domain of locality is the domain of reduced prominence (I follow the literature in calling this the Major Phrase), the following schema express the generalization about when LDNC is possible:

(44) i. (MaP NEG ...) *wala*-NP
ii. *(MaP NEG ...) (MaP ...) *wala*-NP

This hypothesis leads to the question: What is the connection (if any) between verb frequency (as reflected in Table 1) and transparency to long-distance negative concord?

Words that are used with a high relative frequency in speech are often pronounced with reduced prominence, meaning with reduced pitch or without pitch, they are pronounced with shorter duration, etc. (Heine, 1993; Bybee & Schiebman, 1999; Joan & Thompson, 2000; Ladd, 2008). As such, the question might be whether pitch is lowered or weakened more on high-frequency subordinating verbs in Levantine Arabic than on others with a lower frequency. If transparency to LDNC is correlated to prosodic weakening (in the form of greater reduction in prominence), then the prosodic locality condition hypothesized above would predict that verbs which block LDNC are pronounced with greater prominence and therefore resist being included in a prosodically subordinate position. Investigating this goes far beyond the scope of this paper. However, the following predictions need to be tested:

a. i. Are the verbs which block LDNC subject to reduced prominence generally? The hypothesis would predict otherwise.
ii. Do other focus-sensitive operators (e.g. other n-words such as *ʔabadan* or *bilmarra* "never" and *ḥitta* "even") likewise trigger reduction in prominence on backgrounded phrases in the way that *wala* does? If so, are they subject to similar prosodic locality conditions? The hypothesis would predict that they would be.
iii. Are negative sentences with *ʔaiy*-phrases also subject to reduced prominence? If so, why aren't they subject to the same prosodic restrictions that *wala* is subject to?
iv. What is the domain of reduced prominence in Southern Levantine Arabic?

These questions need to be investigated in terms of the interaction of syntactic structure, pragmatics and prosody in Southern Levantine Arabic and in Levantine Arabic more generally.

3. Summary

I have shown evidence that locality restrictions on negative concord in Southern Levantine Arabic are to be characterized in prosodic terms. Verbs that are transparent to LDNC are syntactically and semantically heterogeneous but have a high rate of occurrence in naturally-occurring speech, suggesting that transparency to LDNC may be a frequency effect of the sort discussed by Bybee and others (Heine, 1993; Bybee & Schiebman, 1999; Ladd, 2008).

Furthermore, examination of LDNC sentences from Maamouri et al. (2006 a,b) are consistent with generalizations noted for Italian, German and Japanese, according to which the constituents over which certain operators are interpreted as taking scope correspond to prosodic rather than syntactic constituents. In particular, the prosodic constituent in question appears to be the domain of reduced prominence, which is observed to take place in German, Japanese and Italian (as discussed above) and which appears to be taking place in the Levantine Arabic data examined above. The question is therefore whether a correlation can be drawn between the frequency of a verb in speech and its succeptibility to appearing in prominence configurations.

References

Abdalla, A. G. (1960). *An instrumental study of the intonation of Egyptian Arabic*. Unpublished Ph.D. dissertation, University of Michigan.

Aissen, J., & Perlmutter, D. (1983). Clause reduction in Spanish. In D. M. Perlmutter, & C. M. Rosen (Eds.), *Studies in relational grammar* (pp. 360–403). Chicago, IL: University of Chicago Press.

Al-Sarayreh, A. (2012). *The licensing of negative sensitive items in Jordanian Arabic*. Unpublished Ph.D. dissertation, University of Kansas.

Andrews, A. D. & Manning, C. D. (1999). *Complex predicates and information spreading in LFG*. Stanford, CT: Center for the Study of Language and Information.

Aoyagi, H., & Ishii, T. (1994). On NPI licensing in Japanese. In N. Akutsuka (Ed.), *Japanese/Korean linguistics 4* (pp. 295–311). Stanford, CT: Center for the Study of Language and Information.

Bard, E., Robertson, D., & Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language*, 72, 32–68. DOI: 10.2307/416793

Bayer, J., & Kornfilt, J. (1989). Restructuring effects in German. In E. Engdahl, M. Reape, M. Mellor, & R. Cooper (Eds.), *Parametric variation in Germanic and romance*. Edinburgh: University of Edinburgh Center for Cognitive Science.

Blaszczak, J., & Gärtner, H.-M. (2005). Intonational phrasing, discontinuity, and the scope of negation. *Syntax*, 8, 1–22. DOI: 10.1111/j.1467-9612.2005.00072.x

Blau, J. (1960). *Syntaxes des palästinensischen Baurndialektes Bir Zeits*. Waldorf, Hessen: Verlag für Orientkunde.

Butt, M. (1995). *The structure of complex predicates in Urdu*. Stanford, CT: Dissertations in Linguistics, Center for the Study of Language and Information.

Bybee, J. & Schiebman, J. (1999). The effect of usage on degree of constituency: The reduction of don't in American English. *Linguistics*, 37, 575–596. DOI: 10.1515/ling.37.4.575

Chahal, D. (1999). A preliminary analysis of Lebanese Arabic intonation. In Proceedings of the 1999 Conference of the Australian Linguistic Society (pp. 1–17).

Chahal, D. (2001). *Modeling the intonation of Lebanese Arabic using the autosegmental-metrical framework: A comparison with English*. Unpublished Ph.D. dissertation, University of Melbourne.

Chung, S. (2004). Restructuring in Chomorro. *Syntax*, 7, 199–233. DOI: 10.1111/j.1368-0005.2004.00070.x

Cinque, G. (2001). "Restructuring" and functional structure. In L. Bruge (Ed.), *Venice working papers in linguistics* (Vol. 11, pp. 45–127). Venice: University of Venice.

Cowart, W. (1997). *Experimental syntax: Applying objective methods to sentence judgments*. Los Angeles, CA: Sage Publications.

Cowell, M. (1964). *A reference grammar of Syrian Arabic*. Washington, DC: Georgetown University Press (Washington).

Deguchi, M., & Kitagawa, Y. (2002). Prosody and wh-questions. In M. Hirotani (Ed.), *Proceedings of the North Eastern linguistics society* (pp. 73–92). Amherst, MA: Graduate Linguistics Student Association.

Dziwirek, K. (1998). Reduced constructions in universal grammar: Evidence from the Polish object control construction. *Natural Language and Linguistic Theory*, 16, 53–99. DOI: 10.1023/A:1005952030641

El-Hassan, S. A. (1990). Intonation in the educated spoken Arabic of Jordan: The patterning of accents. *Abhath al-Yarmouk (Yarmouk Studies)*, 8, 7–31.

Featherston, S. (2005). Magnitude estimation and what it can do for your syntax. *Lingua*, 115, 1525–1550. DOI: 10.1016/j.lingua.2004.07.003

Giannakidou, A. (1997). The landscape of polarity items. Unpublished Ph.D. dissertation, Groningen University.

Giannakidou, A. (1998). *Polarity sensitivity as (Non)veridical dependency*. Amsterdam: John Benjamins. DOI: 10.1075/la.23

Giannakidou, A. (2000). Negative…concord? *Natural Language and Linguistic Theory*, 18, 457–523. DOI: 10.1023/A:1006477315705

Hasegawa, N. (1994). Economy of derivation and A'-movement in Japanese. In M. Nakamura (Ed.), *Current topics in English and Japanese* (pp. 1–25). Tokyo: Hituzi Syobo Publishing Company.

Heine, B. (1993). *Auxiliaries: Cognitive forces and grammaticalization*. Oxford: Oxford University Press.

Hellmuth, S. (2005). No de-accenting in (or of) phrases: Evidence from Arabic for cross-linguistic and cross-dialectal prosodic variation. In S. Frota, M. Vigario, & M. Freitas (Eds.), *Prosodies* (pp. 99–112). Berlin: Mouton de Gruyter.

Hellmuth, S. (2006). *Intonational pitch accent distribution in Egyptian Arabic*. Unpublished Ph.D. dissertation, School of Oriental and African Studies, University of London.

Hellmuth, S. (2011). Acoustic cues to focus and givenness in Egyptian Arabic. In Z. Majeed Hassan, & B. Heselwood (Eds.), *Instrumental studies in Arabic Phonetics* (pp. 301–323). Amsterdam: John Benjamins.

Hirotani, M. (2005). *Prosody and LF interpretation*. Unpublished Ph.D. dissertation, University of Massachusetts at Amherst.

Hoyt, F. (2006). Long-distance negative concord and restructuring in Palestinian Arabic. In P. Dekker, & H. Zeiljstra (eds.), *Concord phenomena and the syntax-semantics interface* (pp. 27–32). Malaga: European Summer School of Logic, Language and Information.

Hoyt, F. M. (2010). *Negative concord in Levantine Arabic*. Unpublished Ph.D. dissertation, University of Texas at Austin.

Ishihara, S. (2002). Invisible but audible wh-scope marking: Wh-constructions and deaccenting in Japanese. In L. Mikkelsen, & C.Potts, (Eds.), *Proceedings of West Coast conference on formal linguistics 21* (pp. 180–193). Somerville, MA: Cascadilla Press.

Ishihara, S. (2003). *Intonation and interface conditions*. Unpublished Ph.D. dissertation, Massachusetts Institute of Technology.

Ishihara, S. (2005). Prosody-scope match and mismatch in Tokyo Japanese wh-questions. *English Linguistics*, 22, 2. DOI: 10.9793/elsj1984.22.347

Ishihara, S. (2007). Intonation of sentences with an NPI. In S. Ishihara (Ed.), *Interdisciplinary studies on information structure 9: Proceedings of the 2nd Workshop on Prosody, Syntax, and Information Structure* (pp. 77–119). Potsdam: University of Potsdam.

Israel, M. (1996). Polarity sensitivity as lexical semantics. *Linguistics and Philosophy*, 19, 619–666. DOI: 10.1007/BF00632710

Israel, M. (2001). Minimizers, maximizers and the rhetoric of scalar reasoning. *Journal of Semantics*, 18, 297–331. DOI: 10.1093/jos/18.4.297

Joan, B. & Thompson, S. A. (2000). Three frequency effects in syntax. *Berkeley Linguistics Society*, 23, 378–388.

Kadmon, N. & Landman, F. (1993). Any. *Linguistics and Philosophy*, 16, 353–422. DOI: 10.1007/BF00985272

Kato, Y. (1985). *Negative sentences in Japanese*. Tokyo University: Sophia Linguistica, 19.

Kayne, S. (1998). Overt vs. covert movement. *Syntax*, 1, 128–191. DOI: 10.1111/1467-9612.00 006

Keller, F. (2000). *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality*. Unpublished Ph.D. dissertation, University of Edinburgh.

Kitagawa, Y. (2005). Prosody, syntax and pragmatics of wh-questions in Japanese. *English Linguistics*, 22, 302–346. DOI: 10.9793/elsj1984.22.302

Kitagawa, Y., & Fodor, J. D. (2003). Default prosody explains neglected syntactic analyses of Japanese. In W. McClure (Ed.), *Japanese/Korean linguistics 12* (pp. 267–279). Stanford, CT: Center for the Study of Language and Information.

Kitagawa, Y., & Fodor, J. D. (2006). Prosodic influence on syntactic judgements. In G. Fanselow, C. Féry, R. Vogel, & M. Schlesewsky (Eds.), *Gradience in grammar: Generative perspectives* (pp. 336–358). Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199274796.003.0017

Krifka, M. (1995a). The semantics and pragmatics of negative polarity items. *Linguistic Analysis*, 25, 209–357.

Krifka, M. (1995b). The semantics and pragmatics of polarity items. *Linguistic Analysis*, 25, 1–49.

Kulk, F., Odé, C., & Woidich, M. (2003). The intonation of colloquial Damascene Arabic: A pilot study. In E. H. Berkman (Ed.), *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam 25* (pp. 15–20). Amsterdam: University of Amsterdam.

Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press. DOI: 10.1017/CBO9780511808814

Lahiri, U. (1998). Focus and negative polarity in Hindi. *Natural Language Semantics*, 6, 57–123. DOI: 10.1023/A:1008211808250

Lucas, C. (2009). *The development of negation in Arabic and Afro-Asiatic*. Unpublished Ph.D. dissertation, University of Cambridge.

Maamouri, M., Buckwalter, T., Graff, D., & Jin, H. (2006a). *Levantine Arabic QT training data set 5 (speech)*. DVD-ROM. LDC Catalog No. LDC2006S29.

Maamouri, M., Buckwalter, T., Graff, D., & Jin, H. (2006b). *Levantine Arabic QT training data set 5 (transcripts)*. DVD-ROM. LDC Catalog No. LDC2006T07.

Matos, G. (1999). Negative concord and the scope of negation. *Catalan Working Papers in Linguistics*, 7, 175–190.

Mitchell, T. F. (1993). *Pronouncing Arabic* (Vol. 2). Oxford: Oxford University Press.

Miyagawa, S. (1987). Restructuring in Japanese. In T. Imai, & M. Saito (eds.), *Issues in Japanese linguistics* (pp. 273–300). Dordrecht: Foris.

Monachesi, P. (2005). *The verbal complex in Romance: A case study in grammatical interfaces*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199274758.001.0001

Muraki, M. (1978). The Sika-nai construction and predicate restructuring. In J. Hinds, & I. Howard (Eds.), *Problems in Japanese syntax and semantics* (pp. 155–177). Tokyo: Kaitakusha.

Pierrehumbert, J., & Beckman, M. (1988). *Japanese tone structure*. MIT Press.

Piñar Lurrubia, P. (1996). *Negative polarity licensing and negative concord in the Romance languages*. Unpublished Ph.D. dissertation, University of Arizona.

Poser, W. (1984). *The phonetics and phonology of tone and intonation in Japanese*. Unpublished Ph.D. dissertation, Massachusetts Institute of Technology.

Progovac, L. (1993). Locality and subjunctive-like complements in Serbo-Croatian. *Journal of Slavic Linguistics*, 1, 116–144.

Przepiórkowski, A., & Kupść, A. (1997a). *Negative concord in polish*. Technical Report 828, Institute of Computer Science, Polish Academy of Sciences.

Rizzi, L. (1978). A restructuring rule in Italian syntax. In S. J. Keyser (ed.), *Recent transformational studies in European languages* (pp. 113–158). Cambridge, MA: MIT Press.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language and Linguistic Theory*, 1, 75–116.

Rullmann, H. (1996). Two types of negative polarity items. Proceedings of the North East Linguistic Society, 26, 335–350.

Selkirk, E., & Tateishi, K. (1991). Syntax and downstep in Japanese. In C. Georgopoulos, & R. Ishihara (Eds.) *Interdisciplinary approaches to language: essays in honor of S.-Y. Kuroda* (pp. 519–544). Dordrecht: Kluwer Academic Publishers. DOI: 10.1007/978-94-011-3818-5_28

Stejapanović, S. (2004). Clitic-climbing and restructuring with "finite clause" and infinitive complements. *Journal of Slavic Linguistics*, 12, 173–212.

Sugahara, M. (2003). *Downtrends and post-FOCUS intonation in Japanese*. Unpublished Ph.D. dissertation, University of Massachusetts at Amherst.

Suñer, M. (1995). Negative elements, island effects, and resumptive no. *The Linguistic Review*, 12, 233–273. DOI: 10.1515/tlir.1995.12.3.233

Watanabe, A. (2004). The genesis of negative concord: Syntax and morphology of negative doubling. *Linguistic Inquiry*, 4, 559–612. DOI: 10.1162/0024389042350497

Woidich, M. (1968). *Negation und negative Sätze im Ägyptisch-Arabischen*. Unpublished Ph.D. dissertation, Ludwig-Maximilians-Universität (München).

Wouden, T. v. d. (1994). Negative contexts. Unpublished Ph.D. dissertation, Rijksuniversiteit Groningen.

Wurmbrand, S. (2001). *Infinitives, restructuring and clause structure*. Berlin: Mouton de Gruyter.

Wurmbrand, S. (2005). Verb clusters, verb raising, and restructuring. In M. Everaert, & H. von Riemsdijk (Eds.), *The Blackwell companion to syntax* (pp. 227–341). Blackwell.

Yamashita, H. (2008). Prosody and the syntax of shika-NPIs in Tokyo Japanese and its implications for the theory of grammar. In S. Clarke, M. Hirayama, K. Kim, & E. Suh (Eds.), *Toronto working papers in linguistics 28: Proceedings of the International Conference on East Asian Linguistics* (pp. 375–394). Toronto: Department of Linguistics, University of Toronto.

Zeijlstra, H. (2004). *Sentential negation and negative concord*. Unpublished Ph.D. dissertation, Amsterdam University.

Zwarts, F. (1995). Nonveridical contexts. *Linguistic Analysis*, 25, 286–312.

Zwarts, F. (1996). Three types of polarity items. In F. Hamm, & E. Hinrichs (eds.), *Plural quantification* (pp. 177–238). Dordrecht: Kluwer.

Negation and the subject position in San'ani Arabic*

Elabbas Benmamoun and Khalid Al-Asbahi

In this paper, we discuss a topic that has not received extensive attention within the lively debate about the syntax of sentential negation in Arabic varieties. The topic concerns the interaction between the subject of the sentence, particularly the pronominal subject, and the sentential negative. In most Arabic varieties, the pronominal subject precedes the sentential negative on a par with its lexical subject counterpart. San'ani Arabic, however, allows for the pronominal subject to follow the sentential negative. Assuming that the structure of sentential negation, and of the clause in general, is to a large extent uniform across dialects, we will provide an account that derives the somewhat peculiar pattern in San'ani Arabic and contrast it with a different pattern in Moroccan Arabic that illustrates a different manifestation of the interaction between negation and the pronominal subject. San'ani relies on cliticization (without movement) while Moroccan Arabic relies on head movement. In the course of the paper, we introduce data that have not figured prominently in the debate about sentential negation in Arabic varieties and also draw diachronic parallels between San'ani Arabic and Classical Arabic, where cliticization may have been the process that gave rise to the negative *laysa*.

Keywords: Sentential negation, Moroccan Arabic,San'ani Arabic,cliticization

1. Morpho-syntax of sentential negation in Arabic varieties

Most of the research on sentential negation in Arabic varieties, including both the so-called colloquial varieties and Standard Arabic, has focused on the position of the sentential negative and its interaction with the verb and its extended functional domain, particularly tense (Ouhalla, 1991, 1992, 1993, 1994, 2002; Eid, 1993, Benmamoun, 1992, 2000; Fassi Fehri, 1993; Shlonsky, 1997;

* Benmamoun's research for this paper was supported by NSF grant BCS 0826672 (Benmamoun, PI).

Brustad, 2000; Onizan, 2005; Bahloul 2007, Lucas, 2007; Soltan, 2007, 2011; Hoyt, 2010; Aoun, Benmamoun & Choueiri, 2010; among many others). The paradigm cases of the intimate connection between sentential negation and the verb come from Standard Arabic, where the form of the sentential negative co-varies with tense.[1]

(1) a. yanbaħu l-kalb-u
bark.3MS the-dog-NOM
'The dog is barking/barks'

b. nabaħa l-kalb-u
barked.3MS the-dog-NOM
'The dog barked'

c. sa-yanbaħu l-kalb-u
FUT-bark.3MS the-dog-NOM
'The dog will bark'

(2) a. laa yanbaħu l-kalb-u
NEG bark.3MS the-dog-NOM
'The dog is barking/barks'

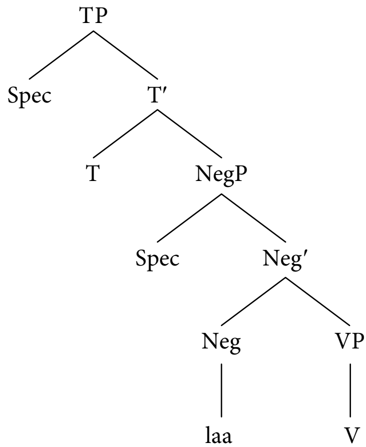
b. lam yanbaħ l-kalb-u
NEG.PAST bark.3MS the-dog-NOM
'The dog did not bark'

c. lan yanbaħa l-kalb-u
NEG.FUT bark.3MS the-dog-NOM
'The dog will not bark'

The fact that the form of the negative varies according to tense even though the form of the verb remains relatively the same has led to the analysis whereby negation and tense merge with each other, assuming an analysis where tense and negation head their own phrases in the syntax as illustrated in (3).

1. As indicated below, some of our data were taken from the LDC Arabic corpora (www.ldc.upenn.edu).

(3)



Benmamoun (2000) argues that the reason the verb does not carry tense in standard Arabic is due to its inability to move across the negative head which is a closer host. On the other hand, in the spoken Arabic dialects, the verb does carry tense in the context of sentential negation. This is the case for example, in Moroccan Arabic and Jordanian Arabic in the present, past, and future tenses:[2]

(4) a. ma-kəydir-š lli ʕlih (MA: native speaker)
NEG-do.3MS-NEG what on.him
‘He does not do what he is supposed to do’

b. ma-ħdərti-š f-lwəqt (MA: native speaker)
NEG-arrive.3MS -NEG in-time
‘You did not arrive on time’

c. ma-ɣadi-š yaʕqəl ʕlik (MA: native speaker)
NEG-FUT-NEG remember.3MS on-you
‘He will not remember you’

(5) a. maa-bnokil ʔilla beyḍ baladi (LDC Levantine CallHome)
NEG-EAT.1p but eggs organic
‘We do not eat but organic eggs’

b. wallah maa-rtaħit yom (LDC Levantine CallHome)
by god NEG-rested.1s day
‘By god, I did not rest one single day’

2. The negative *maa* in SA does occur with finite verbs and thus patterns with the sentential negative in the dialects. *maa*, however, is not as frequent, particularly in the context of verbs as Benmamoun et al. (2014a) show based on a corpus study. The difference between *maa* and *laa* is that the latter may attract tense while the former does not.

Table 1. The distribution of the sentential negation in verbal contexts in Arabic varieties.[3]

Tense	Standard	Moroccan	Jordanian/Levantine
Future	lan	ma-š	maa or maa-š
Present	laa	ma-š	maa or maa-š
Past	lam	ma-š	maa or maa-š
Imperative	laa	ma-š	maa or maa-š

The Moroccan Arabic negative is interesting in that it is made up of two parts, a proclitic and an enclitic. The two occur together and independent of other hosts, which suggests that their realization as enclitic and proclitic on a head could be due to merger of the head with sentential negation. In Moroccan Arabic, this merger certainly applies to verbs and could also be due to the movement of the verb through the negative projection. Unlike Standard Arabic, the negative actually merges with the verb, and the whole complex ends up in tense. The main patterns are summarized in Table 1.

In addition to the negation that occurs mostly in the context of verbs, there is another negation that occurs in the context of non-verbal predicates. This is illustrated by Moroccan Arabic in (6).

(6) l-wəld maši hna
the-boy NEG here
'the boy is not here'

This pattern arises because there is no head that moves to sentential negation and merges with it.[4] Table 2 summarizes the main distribution of negation in Standard Arabic, Moroccan Arabic, and Levantine Arabic.

Thus, the structure in (3) and the attendant syntactic analysis in terms of head movement and the related constraints help account for the distribution of sentential negation in Standard Arabic and a number of spoken modern Arabic dialects. However, this is by no means the consensus within approaches to the syntax of sentential negation in Arabic varieties. For example, Soltan (2007) provides very

3. The patterns in Table I represent only a subset of the varieties that exist within each region. Within each country and region, there is significant variability.

4. Non-verbal predicates merge with negation in Moroccan Arabic and other dialects (see Benmamoun, 2000 and Benmamoun et. al., 2014b). We assume that there is no null head, such as a null copula, in verbless sentences. Benmamoun (2000) provides a number of tests based on Case, modals, locality, and head movement to argue against such a head. For example, in Standard Arabic, overt copulas in past tense, future tense, and imperative sentences assign accusative Case to the predicate (see (9) below), but in verbless sentences, the predicate is nominative, which is the default Case that emerges when no syntactic Case assigner is present.

Table 2. The distribution of sentential negation in Arabic varieties in verbal and non-verbal contexts.

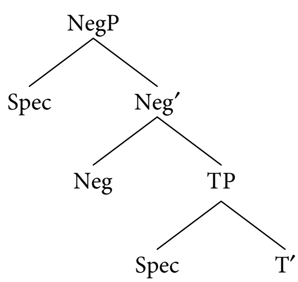
	Verbal Negation	Example	Non-Verbal Negation	Example
Levantine	maa...(š)	ma-ʔižat(-š) Nadia 'Nadia did not come'	miš	huwe miš hon he neg here 'he is not here'
Moroccan	maa...š(i)	ma-žat-š Nadia	maaši	huwa maši hna he neg here 'he is not here'
Standard	laa/ lam/ lan/ maa	lam ta'ti Nadia neg.came Nadia "Nadia did not come"	laysa	huwa laysa huna He not here "he is not here"

plausible arguments that the negative projection is located higher than tense, an analysis adopted by Benmamoun et al. (2014a) as an alternative to (3) for all Arabic varieties. One such argument that Soltan provides comes from the location of the future marker in Egyptian Arabic, which clearly precedes the future tense marker as illustrated in (7):

(7) miš ha-yiskut-u ʕala kida ʔabadan (Egyptian blog)
NEG fut-silent-3p on this ever
'They will never remain silent about it'

According to this analysis, the negative head precedes the temporal projection of the clause as illustrated in (8).

(8)



In (7), Egyptian Arabic deploys the negative that we find in the context of non-verbal predicates, but the temporal marker precedes sentential negation, contrary to the prediction of the structure in (3). Benmamoun et al. (2014) provide numerous arguments to support the analysis that the negative projection is located higher than the tense projection across Arabic varieties (including Standard Arabic, Moroccan Arabic and San'ani Arabic). Suffice it to say that the interaction between tense, the predicate, and negation requires an alternative account if it turns out that the structure in (8) is correct, which we assume it is.

2. Sentential negation and the subject

In all previous analyses, the status of the subject did not figure prominently in the debate about the syntax of sentential negation. However, there is interesting evidence in Arabic that the subject does interact with negation and that this interaction can shed further light on the issue of the position of the subject and the location of sentential negation. Two pieces of evidence are critical in this regard. First, Standard Arabic has a negative *laysa* that agrees with the subject and assigns accusative Case to the predicate:

(9) a. laysa l-bayt-u kabiir-an
NEG.3MS the-house-NOM big-ACC
'The house is not big'

b. lays-at l-madiinat-u kabiirat-an
NEG.3FS the-city-NOM big-ACC
'The city is not big'

In fact, *laysa* patterns with verbs in that it agrees with the subject in number, person, and gender and can license null subjects:[5]

(10) a. lays-uu fii l-bayt-i
NEG-3MP in the-house-GEN
'They are not in the house'

b. las-na fii l-bayt-i
NEG-3FP in the-house'
'They are not in the house'

5. Benmamoun (2000) challenges the view that *laysa* is a verb and provides a diachronic account for its peculiar inflectional behavior, namely the fact that it carries agreement morphology that usually occurs in the context of past tense sentences though it occurs mainly in present tense contexts. See Benmamoun et al. (2014a) for a more detailed analysis that compares *laysa* with negative pronouns in the dialects. If *laysa* were a genuine verb, one would expect it to display imperfective morphology that is the expected pattern in present tense sentences. Moreover, verbs, including copular verbs, are not allowed in non-generic present tense sentences with non-verbal predicates, such as adjectives, prepositions, and nouns (Benmamoun, 2000). Were *laysa* to contain a verbal element or to be a verb itself, it would be an exception to this generalization that as far as we know exists in all Arabic varieties. In short, there are no compelling empirical or conceptual reasons for positing a verbal element in the context of the negative *laysa*.

Second, in the so-called verbless sentences in Arabic dialects, such as Moroccan Arabic, the subject must precede sentential negation regardless of whether it is a lexical noun phrase or a pronoun.

(11) a. l-wəld maši hna
the-boy NEG here
'The boy is not here'

b. huwa maši hna
he NEG here
'He is not here'

(12) a. *maši l-wəld hna
NEG the-boy here
'The boy is not here'

b. *maši huwa hna
he NEG here
'He is not here'

In (9–12) we see a striking contrast between Moroccan Arabic and Standard Arabic. In Standard Arabic, the subject can follow sentential negation but it cannot do so in Moroccan Arabic. If the structure in (8) is correct, the movement of the subject in Moroccan Arabic may have nothing to do with tense, as argued in Benmamoun (2000). Rather, it could be due to some property of negation that requires its specifier to be filled by the subject; or to some property of the clause that prefers the highest specifier position to be filled by the subject, or for the head to carry subject agreement (so called EPP).[6] This account would potentially provide an analysis for both Standard Arabic and Moroccan Arabic. In Moroccan Arabic, the subject must move to the highest subject position to fulfill the EPP requirement of the clause or sentential negation. On the other hand, in Standard Arabic, this requirement is fulfilled by the agreement on the negative, as argued in Benmamoun (2000) following Alexiadou & Anagnostopoulou (1999).[7] This could explain why *laysa* can precede or follow the subject, an optionality that has to do with the ability to satisfy the EPP property of the clause through an XP specifier or a head endowed with agreement features. It seems, then, that the only viable analysis is the one that allows for Standard Arabic to leave its subject lower (thanks to the agreement features on negation) but forces it to move higher in Moroccan

6. Note that the status of the EPP (originally Extended Projection Principle) is still unclear within the Minimalist Program. Originally, it was limited to the distribution of subjects but recently its scope has been expanded to include wh-constructions and other movement constructions.

7. This could also account for why the Standard Arabic negatives, *laa*, *lam*, and *lan*, must be adjacent to the verb which carries agreement features (Benmamoun 2000).

Arabic – and possibly other modern dialects such as Egyptian Arabic, since the latter do not allow negation to carry agreement to satisfy the EPP requirement without XP movement to the higher Spec position.[8] However, there is an Arabic variety that on the surface seems to contradict this neat distribution. The Sana'ani Arabic dialect of Yemen allows the pronominal subject to follow or precede sentential negation. We will show that rather than contradicting the analysis, Sana'ani Arabic actually supports it as long as we allow for cliticization to also satisfy the EPP requirement. We turn to this variety next.

Sentential negation in San'ani Arabic patterns to a large extent with its counterpart in Moroccan Arabic and Egyptian Arabic. The sentential negative consists of a proclitic *ma* and an enclitic *š* (Watson, 1993). The verb and its aspectual markers are flanked by the two negative markers as illustrated in (13):

(13) a. ma-bein-a-lʕab-š
NEG-ASP.1s-play-NEG
'I am not playing'

b. ma-kunt-š daari
NEG-was.1s-NEG knowing
'I did not know'

c. ma-ça-yidra-š ba-l- makaan
NEG-FUT-3MS.know-NEG of-the-place
'He will not know the place'

Unlike Moroccan Arabic, and on a par with eastern dialects, particularly in the Gulf region, San`ani uses *laa* or its reduced form *la* in negative imperative as an alternant with *ma-š*:

(14) a. la-ti-gambir-š hana
NEG-2-sit-NEG here
'Do not sit here'

b. ma-ti-gambir-š hana
NEG-2-sit-NEG here
'Do not sit here'

If we confine our attention to the above data, San'ani Arabic appears to pattern with the other Arabic dialects, particularly the Moroccan Arabic dialects where negation merges with the verbs in all tenses, and also with the imperative verb. San'ani also provides one of the strongest pieces of evidence for the structure in

8. The lower position of the subject does not have to be the lexical projection but could be specifier of the tense projection itself.

(8), where negation is located above the projection of tense. The critical data are given in (15a, b) and the configuration of the clause with the ellipsis fragment of the sentence is given in (15c):

(15) a. hu musaafir bass ʔana **maʕ**
he travel but I NEG
'He will travel but I will not

b. hu musaafir bass ʔana **maši**
he travel but I NEG
'He will travel but I will not

c. ....[bass [$_{\text{NegP}}$ ʔana **maši** [$_{\text{TP}}$.....]]]

In (15a, b) we see that the negative is stranded on its own with no host. In (15a) only the proclitic *ma* appears with a voiced pharyngeal as coda but in (15b) both components of negation, the enclitic and the proclitic appear as fragments of ellipsis after the subject. The interpretation of the elided constituent includes the predicate and its tense, which under any reasonable analysis would mean that at least TP is elided, thus giving further support to the idea that negation is higher than TP as in the representation in (8).

3. Interaction between the subject and negation in San`ani and Moroccan Arabic

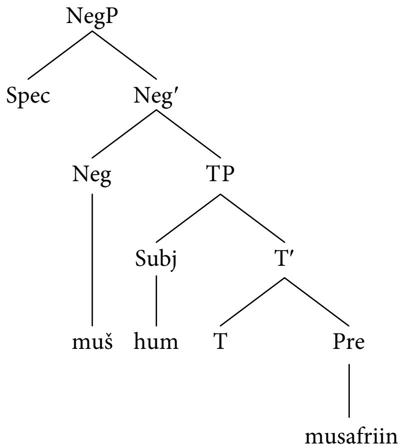
With this brief description of San'ani sentential negation in mind, we can now turn to the distribution of the subject in the context of sentential negation. Consider the following sentences:

(16) a. muš hum musafriin
NEG they traveling
'They are not traveling'

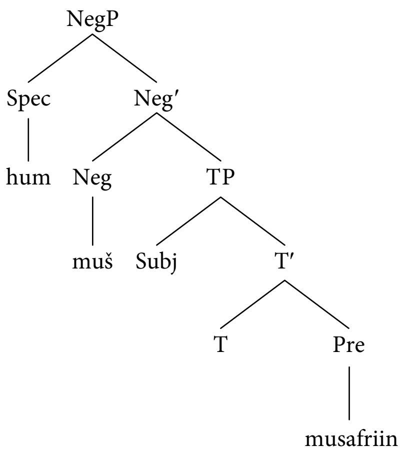
b. muš ʔant ba-l-beit ʔabadan
NEG you in-the-house ever
'You are never in the house'

The sentences in (16), where the pronominal subject follows the verb, are grammatical with no special intonation on the negative. In this respect, San'ani Arabic contrasts sharply with Moroccan Arabic, where the equivalent sentences are ungrammatical. In the former, the pronominal subject is located in the specifier of NegP (17a), and in the latter, it is located in the specifier of TP (17b).

(17) a.



b.



A number of questions arise but we will focus on the following two:

i. Is the subject in the specifier of NegP, or is it in a higher projection?
ii. What drives the movement of the subject above TP?

Starting with the first question, there is strong evidence that the subject is in the specifier of the negative projection. The evidence comes from negative quantifiers and negative polarity items in Arabic varieties. In San'ani Arabic, very much like in many Arabic dialects (Kenstowicz, 1989; Hoyt, 2010), a negative subject such as *wala + NP* precludes the presence of the negative head *ma-š*.

(18) wala-ħad haane
no-one here
'There is no one here'

The complementary distribution between the head of sentential negation and the negative subject when the latter precedes sentential negation follows from economy conditions under the assumption that only one negative element can occupy the negative projection and license it.[9] Support for this analysis comes from Moroccan Arabic, where the NPI *ħətta + NP* does not preclude the presence of the negative head *ma*.

(19) ħətta wəld ma-hna
even boy NEG-here
'No boy is here'

The difference between San'ani and Moroccan can be attributed to the fact that the San`ani subject in (18) contains a negative element *laa*. This is not the case for the Moroccan Arabic subject in (19), which in turn requires the presence of the negative head.[10] That is, the *wala* phrase in San'ani is specified for negative features thanks to *laa*, but the *ħətta* phrase in Moroccan is not. This is plausible because *ħətta* is not negative but rather a presuppositional and focus particle which can occur in non-negative contexts (Benmamoun, 1995). In San'ani, the *wala* subject is in the Spec of negation, which obviates the presence of the negative head.[11]

Turning to the second question about the motivation for the movement of the subject to a position higher than TP, this movement is obligatory for Moroccan Arabic subjects and optional for San`ani Arabic subject pronouns. However, San'ani Arabic lexical subjects pattern with Moroccan Arabic subjects in that they must occur in the Spec of the negative projection (20).

(20) a. ʔal-ħadiiga miš haliya
the-park NEG beautiful
'the park is not beautiful'

b. *miš ʔal-ħadiiga haliya
NEG the-park beautiful

Therefore, the question must be refined to focus on why San`ani pronominal subjects can remain in a lower position on a par with lexical subjects in Standard

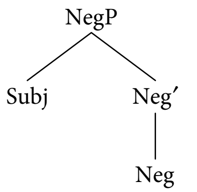
9. The idea is that preferably only one position (head or specifier) needs to be overt.

10. Moroccan NPIs are in complementary distribution with the negative *š*, but that complementarity is not sensitive to grammatical function or syntactic position. It obtains regardless of whether the NPI is an NP in the pre-negative or post-negative position or whether it is a maximal projection or head (see Benmamoun, 1997, 2006).

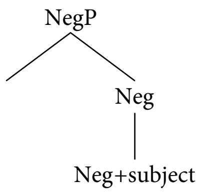
11. When the *wala* phrase occurs in the pre-negative position with an overt negative head, we get a double negation interpretation. This could be due to the presence of two negative projections, with the *wala* phrase occupying one and the head occupying another, or alternatively we may be dealing with an instance of constituent negation. We leave this open.

Arabic. It seems that the default option is for the subject to move to the specifier of negation, possibly to fulfill the (EPP) requirement that the negative projection host a nominal element, such as the subject, in its head position as in Standard Arabic, where the head carries subject agreement or must be adjacent to a verb inflected for subject agreement. The two options are illustrated below:

(21) a.



b.



Since the lexical subject is a maximal projection, merger with a head is presumably not an option due to whatever constraint blocks the movement of maximal projections to head positions (traditionally blocked under Structure Preservation). This leaves only one option open for lexical subjects, namely the (21a), where the subject moves to the Spec of NegP. If the subject is a pronominal, two options are possible. If it moves as a maximal projection (which would be the case if the whole DP moves), only option (21a) is available. On other hand, if the subject is a pronoun occupying the head of the nominal projection containing the pronominal subject, it can merge with the negative head via head movement and incorporation. This, we believe, is what happens to Moroccan Arabic pronouns in sentences such as (22):

(22) ma-huwa-š hna
NEG-he-NEG here
'He is not here'

What other options can there be? The pattern in (22) is not possible in San'ani (23); but it seems that San'ani displays a third option whereby the pronominal subject stays in the Spec of TP and cliticizes onto the negative head.

(23) *ma-huwa-š haane
NEG-he-NEG here

The idea is that the representation in (17a) allows for the pronominal subject in the specifier of TP to cliticize onto the negative *miš* in NegP. That is, instead of head movement, San'ani Arabic opts for cliticization, which is essentially rebracketing. The cliticization option is most likely the reason why the negative seems to inflect for gender agreement in San'ani. Consider the sentence in (24):

(24) ʔal-ħadiiga **miš**-i haliya
the-park NEG-2FS beautiful
'The park is not beautiful'

The gender agreement on the negative is a vestige of the pronominal subject that cliticized onto the negative, a process that is most likely behind the historical development of the inflected negative *laysa* in Standard Arabic and the variants of the negative *muu* in some Gulf varieties (Holes, 1990). In this scenario, there are two paths for pronouns in Arabic varieties in the context of negation. Under one path, the negative merges with the pronoun, possibly as an instance of head movement. This is the Moroccan Arabic option that is also found in other varieties such as Egyptian (Eid, 1993). Under another path, the pronoun stays in a projection lower than negation but cliticizes onto it. This is the San'ani Arabic option and possibly the option that led in part to the rise of *laysa* in Classical Arabic. These two options yield the two types of negative pronoun paradigms in Table 3.

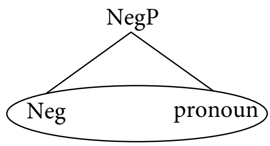
The contrast between the two paradigms is striking, but under the present analysis, it appears that these are the only two options for pronouns because of the nature of the configuration from which they originate. Negation occupies a projection above the tense projection that hosts the subject. The merger between negation and the pronominal subject can only obtain in two ways: either through head

Table 3. Negative pronouns in Arabic varieties.

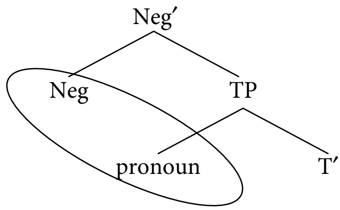
Pronoun	Sana'ani	MA
1sm	mišana	manaš
1sp	mišeħne	maħnaš
2ms	mišant	mantaš
2fs	mišanti	mantiš
3ms	miššu	mahuwaš
3fs	mišši	mahiyaš
2mp	mišantum	mantumaš
2fp	mišantum/ mišantan	mantumaš
3mp	miššum	mahumaš
3fp	miššum/ miššan	mahumaš

movement of the pronoun to the head of negation, to create a complex head, or through encliticization, where no movement takes place but the two elements phonologically lean on each other and form a morphological unit in the morph-phonological component.[12] The two options, movement and cliticization, are illustrated in (25) and (26) respectively.

(25) Head Movement



(26) Cliticization



Under the movement option in (25), the pronoun is adjoined to the head of the negative projection. By contrast, under the cliticization option in (26), the two elements are located in their respective projections throughout the syntactic component and are merged (through cliticization) in the post-syntactic component.[13] The cliticization option maybe another way to satisfy the EPP requirement.[14] If this is correct, the EPP requirement can be satisfied through XP movement to a specifier position, agreement on the head, or cliticization. The typology of the interaction between sentential negation and subjects in Arabic seems to exemplify all three options.

12. The third option, namely cliticization of the pronoun in the specifier of negation onto the head of NegP is ruled out in Arabic, where cliticization and prosodic rebracketing (as in the Construct State) seem to require a sequence where the head is to the left, possibly due to a c-command requirement.

13. Cliticization would then be another way to satisfy the requirement that negation be paired with a subject. The options are (i) a subject in the Spec of negation, (ii) a pronoun that moves into the head position of negation, or (iii) a pronoun that cliticizes onto the negative head. Clearly, the relation between the subject and negation and the difference between lexical and pronominal subjects require further research.

14. The cliticization option may also be available for *maa* in Classical Arabic, which can precede the subject and seems to be adjacent to it.

4. Conclusion

Arabic varieties provide fertile grounds for testing current linguistic approaches. This is clearly the case in the context of sentential negation. In one variety, such as Standard Arabic, negation carries subject agreement on a par with verbs; in another variety, particularly San'ani Arabic, the pronominal subject may follow the sentential negative and cliticize onto it; and in yet a third variety, such as Moroccan Arabic, the pronominal subject can merge with negation or precede it on a par with lexical subjects. In this paper, we have put forward an analysis that maintains a uniform structure for the Arabic varieties, particularly San'ani Arabic and Moroccan Arabic, but each deploys different options that the grammar makes available within the limits of the configuration, the operations that can be carried out, and the constraints on those operations. San'ani opts for the cliticization route and derives negative pronominal paradigms where the pronoun is clearly encliticized onto negation and ultimately becomes an agreement affix as is the case in San'ani and Standard Arabic. Moroccan Arabic opts for the movement option and ends up with the pronoun inside the negative head.

Obviously, a number of questions remain, chief among them the question about the reason for the movement of the subject, lexical and pronominal, to the specifier of the negative head. We suggested an analysis, along current minimalist lines, that negation requires a specifier or a subject, but this account must remain tentative because there is no independent reason why this should be the case. Another question is why the cliticization option is not available in varieties such as Moroccan Arabic and why the head movement option is not available in San`ani. In other words, it is unclear why the two varieties make different choices and what is behind those choices. These questions remain open for now.

References

Alexiadou, A., & Anagnostopoulou, E. (1999). EPP without spec, IP. In D. Adger, S. Pintzuk, B. Plunkett, & G. Tsoulas (eds.), *Specifiers: Minimalist approaches* (pp. 93–109). Oxford: Oxford University Press.

Aoun, J., Benmamoun, E., & Choueiri, L. (2010). *Arabic syntax*. Cambridge: Cambridge University Press.

Bahloul, M. (2007). *Structure and function of the Arabic verb*. New York: Routledge.

Benmamoun, E. (1992). *Inflectional and functional morphology: Problems of projection, representation and derivation*. Unpublished Ph.D. dissertation, USC, Los Angeles.

Benmamoun, E. (1995). Negative polarity and presupposition. In M. Eid (Ed.), *Perspectives on Arabic linguistics VII* (pp. 47–66). Amsterdam: John Benjamins.

Benmamoun, E. (1997). Licensing of negative polarity items in Moroccan Arabic. *Natural Language and Linguistic Theory*, 15, 263–287. DOI: 10.1023/A:1005727101758

Benmamoun, E. (2000). *The feature structure of functional categories: A comparative study of Arabic dialects*. Oxford: Oxford University Press.

Benmamoun, E. (2006). Licensing configurations: The puzzle of head negative polarity items. *Linguistic Inquiry*, 37(1), 141–149. DOI: 10.1162/ling.2006.37.1.141

Benmamoun, E., Abunasser, M., Al-Sabbagh, R., Bidaoui, A., & Shalash, D. (2014). *The Location of sentential negation in Arabic varieties*. Brill's Annual of Afroasiatic Languages and Linguistics 5: 83-116.

Benmamoun, E., Abunasser, M., Al-Sabbagh, R., Bidaoui, A., & Shalash, D. (2014a). *Variations on the same theme: The negative copula in Arabic*. In S. Farwaneh and H. Ouali (Eds.), Perspectives on Arabic Linguistics XXIV–XXV (pp. 121–138). Amsterdam: John Benjamins.

Brustad, K. (2000). *The syntax of spoken Arabic*. Washington, DC: Georgetown University Press.

Eid, M. (1993). Negation and predicate heads. In M. Eid, & G. K. Iverson (Eds.), *Principles and prediction: The analysis of natural language* (pp. 135–152). Amsterdam: John Benjamins. DOI: 10.1075/cilt.98.13eid

Fassi Fehri, A. (1993). *Issues in the structure of Arabic clauses and words*. Kluwer: Dordrecht. DOI: 10.1007/978-94-017-1986-5

Hoyt, F. (2010). *Negative concord in Levantine Arabic*. Unpublished Ph.D. dissertation, University of Texas, Austin.

Holes, C. (1990). *Gulf Arabic*. London: Routledge.

Kenstowicz, M. (1989). The null subject parameter in modern Arabic dialects. In O. Jaeggli, & K. Safir (Eds.), *The null subject parameter* (pp. 263–275). Dordrecht: Kluwer. DOI: 10.1007/978-94-009-2540-3_9

Lucas, C. 2007. *Jespersen's Cycle in Arabic and Berber. Transactions of the Philological Society* 105: 398–431..

Onizan, N. (2005). Function of negation in Arabic literary discourse. Unpublished Ph.D. dissertation, University of Kansas, Lawrence.

Ouhalla, J. (1991). *Functional categories and parametric variation*. London: Routledge. DOI: 10.4324/9780203412015

Ouhalla, J. (1992). Focus in standard Arabic. *Linguistics in Potsdam*, 1, 65–92.

Ouhalla, J. (1993). Negation, focus and tense: The Arabic maa and laa. *Rivista di Linguistica*, 5, 275–300.

Ouhalla, J. (1994). Verb movement and word order in Arabic. In D. Lightfoot, & N. Hornstein (Eds.), *Verb movement* (pp. 41–72). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511627705.004

Ouhalla, J. (2002). The structure and logical form of negative sentences in Arabic. In J. Ouhalla, & U. Shlonsky (Eds.), *Themes in Arabic and Hebrew syntax* (pp. 299–320). Dordrecht: Kluwer. DOI: 10.1007/978-94-010-0351-3_11

Shlonsky, U. (1997). *Clause structure and word order in Hebrew and Arabic: An essay in comparative semitic syntax*. Oxford: Oxford University Press.

Soltan, U. (2007). *On formal feature licensing in minimalism: Aspects of standard Arabic morphosyntax*. Unpublished Ph.D. dissertation. University of Maryland, College Park.

Soltan, U. (2011). Licensing of the NPI ʕumr and its implications for sentential negation in Cairene Egyptian Arabic. *The 2011 linguistic society of America extended abstracts*. http://www.elanguage.net/journals/index.php/lsameeting/article/view/1457

Watson, J. (1993). *A syntax of Sanʕani Arabic*. Wiesbaden: Otto Harrassowitz

Splitting Neg: The morphosyntax of sentential negation in Cairene Egyptian Arabic revisited*

Usama Soltan
Middlebury College

The goal of this paper is to revisit sentential negation patterns in Cairene Egyptian Arabic and propose a novel analysis for their morphosyntactic properties in this Arabic dialect. In particular, it is argued that the distribution of negation patterns takes place in the mapping from the syntax to the morphology, that the head hosting negation is higher than T on the clausal hierarchy, and that the Neg-domain in this dialect is split into two separate heads, one encoding semantic negation, and another expressing formal negativity. Implications of the proposed analysis are discussed with regard to the interaction between negation and polarity items, variation in negation across different Arabic dialects, and negation patterns in negative disjunction and negative concord structures.

Keywords: Sentential negation, morphosyntax, negative polarity, Egyptian Arabic.

* For their insightful comments and questions, I'd like to extend my thanks to Elabbas Benmamoun, Stuart Davis, Fred Hoyt, Sarah Ouwayda, Martin Walkow, and members of the audience at both the *26th Arabic Linguistics Symposium* at Columbia University, and the *3rd Annual Tampa Workshop in Linguistics* at University of South Florida. I'm especially thankful to an anonymous reviewer, who provided substantial comments on the manuscript, which led to several improvements in the paper. I am also indebted to Middlebury College for granting me an academic leave, during which I was able to conduct my study of negation in Arabic dialects, some of which is reported on in this paper. Last but not least, special thanks are due to Karen Froud and Reem Khamis-Dakwar for their valuable editorial assistance and their generous patience during the writing, revising, and proofreading of this paper. Needless to say, all errors or shortcomings in this work remain entirely my responsibility.

1. Introduction: Patterns of sentential negation in Cairene Egyptian Arabic

Many of the Arabic dialects spoken today exhibit a two-pattern negation system: A *discontinuous* pattern, where a predicate appears between two negative elements forming a single prosodic word, and an *independent* pattern, where the predicate follows the negation marker without them forming a unit. In Cairene Egyptian Arabic (CEA, henceforward), for example, perfective verb forms are negated in the discontinuous pattern, as shown in (1) below, where the verb appears between the proclitic *maa* and the enclitic *-š*, both of which are glossed as NEG.[1]

(1) maa-nim-t-i-š
NEG-sleep.PERF-1SG-EV-NEG
'I did not sleep.'

By contrast, the predicate in a copular verbless sentence like (2) is negated by the independent marker *miš*.[2]

(2) ʔanaa miš muhandis
I NEG engineer
'I am not an engineer.'

In this paper, I will occasionally refer to the discontinuous pattern in (1) as the *maa…š* pattern, and to the independent pattern in (2) as the *miš* pattern.

The study of sentential negation patterns in CEA and other modern Arabic dialects has received considerable attention in the generative literature over the past twenty-five years or so, because of its theoretical and empirical implications for linguistic analysis (see, for example, Eid 1993; Shlonsky 1997; Benmamoun 2000; Ouhalla 2002; Hoyt 2007; Aoun et al. 2010; among others). In particular, one can identify three major questions in this literature, as formulated in (3).

1. The following abbreviations are used in the glosses of data: 1, 2, 3 for first, second, and third person, respectively; SG = singular; PL = plural; M = masculine; F = feminine; NEG = negation; FUT = future; ASP = aspect; PERF = perfective; IPFV = imperfective; PTCP = participial; Q = question-particle; ACC = accusative; IND = indicative; EV = epenthetic vowel. Arabic emphatic consonants are marked by symbols with dots underneath them in the transcription of the data. I will also consistently transcribe the negation marker in EA as *maa*, under the assumption that it has an underlying long vowel that undergoes shortening in actual pronunciation. While nothing hinges on this phonological assumption at all, it does gain some support from the fact that the negation marker is pronounced with a long vowel when occurring in *ʕumr* (= 'ever') constructions, which are illustrated in Section 4 of this paper.

2. For some speakers, *miš* may also be pronounced as *muš*. As far as I can tell, there are no morphosyntactic consequences of this phonological difference. For consistency, I will use *miš* in the examples throughout the paper.

(3) a. What conditions regulate the distribution of the two negation patterns?
b. Where is Neg located in Arabic clause structure; i.e., is it higher or lower than T(ense)?
c. What is the grammatical status of the *-š* enclitic of the negation morpheme?

In this paper, I revisit previous answers to the questions in (3) with reference to sentential negation patterns in CEA, pointing out some empirical and theoretical problems with these earlier approaches. Instead, I propose an analysis of the distribution of the two negation patterns in CEA, whereby the negative domain is split into two heads, both of which are placed higher than T in the clausal hierarchy, and where the distribution of the two negation patterns follows from head movement (or lack thereof) in the mapping from the syntax to the morphology. Implications of the proposed analysis are then discussed with regard to facts related to (i) the interaction between negation and polarity-sensitive items in CEA, (ii) variation in the negation strategies utilized across Arabic dialects, and (iii) negative disjunction and negative concord structures, where a third pattern of sentential negation is occasionally used.

The paper is organized as follows: In Section 2, I discuss the distribution of the *maa…š* and *miš* patterns in CEA, showing that an analysis in terms of head movement applying in the mapping from the syntax to the morphology is to be theoretically and empirically preferred over one where such movement is assumed to take place exclusively in the syntax. Section 3 discusses two possible structures for the position of Neg on the clausal hierarchy: The low-Neg analysis, where Neg is projected lower than T, and the high-Neg analysis, where Neg is located above T. Based on attested negation patterns from Egyptian Arabic dialectal variation and child language, I conclude in this section that the high-Neg analysis is more empirically adequate. In Section 4, I discuss two previous analyses of the grammatical status of the *-š* enclitic of the negation morpheme, showing that they both fail to capture certain empirical facts from contexts with polarity-sensitive items in CEA. Section 5 presents an analysis of sentential negation in CEA whereby the Neg-domain is argued to be split between two heads, both of which occupy a position higher than T. The distribution of the two negation patterns is then argued to follow from a morphological algorithm of head movement applying in the mapping from the syntax to the morphology. In Section 6, I explore three types of empirical consequences for the proposed analysis. In particular, I revisit the facts from polarity-sensitive item contexts that have been argued in Section 4 to be problematic under previous analyses, showing that they can receive a principled account under the split-Neg analysis proposed here. I then explore the empirical implications of a split-Neg analysis for the study of variation in negation strategies across Arabic dialects in light of the Jespersen's cycle. Finally, a third pattern of negation, where sentential negation is marked by means of the particle *laa* and the term *walaa*, is illustrated by examples from negative disjunction and negative concord structures in CEA, and an account for this pattern is sketchily explored within the proposed analysis. Section 7 sums up the conclusions of the paper.

2. Distribution of discontinuous and independent negation in CEA

In the relevant literature on sentential negation in Arabic dialects, the first question typically discussed has to do with how each one of the two negation patterns illustrated earlier in (1) and (2) is derived and why each appears in certain contexts but not in others. One influential approach in this regard, first proposed in Benmamoun (2000) and later adopted in Aoun et al. (2010), is based on positing a fundamental distinction between [+Past] and [–Past] contexts within a feature-checking framework along the lines of Chomsky's (1995) Minimalist Program. More specifically, under this analysis, past tense T is specified for both a [+D] and a [+V] feature, whereas present tense T is specified only for a [+D] feature. Under minimalist assumptions, the [+V] feature will attract V in past tense contexts, as in the partial structure in (4a) below, but no such attraction takes place in present tense contexts for the simple reason that there is no [+V] feature on T, as shown in (4b).

(4) a. $[_{\text{TP}}\ \text{T}_{[\text{+Past, +D, +V}]}\ [\text{V}_{\text{P}}\ \ldots\text{V}\ldots]]$

b. $[_{\text{TP}}\ \text{T}_{[\text{–Past, +D}]}\ [\text{V}_{\text{P}}\ \ldots\text{V}\ldots]]$

If Neg were located lower than T, as assumed under this analysis, it follows that past tense verb forms will always appear in the discontinuous negation pattern, since V has to pick up Neg on its way to T, per the *Head Movement Constraint* (HMC; Travis 1984), as shown in (5a). By contrast, when T is [–Past], no verb raising takes place, and Neg is spelled-out as the independent negation marker in this case, as shown in (5b).[3]

(5) a. $[_{\text{TP}}\ \text{T}_{[\text{+Past, +D, +V}]}\ [_{\text{NegP}}\ \text{Neg}\ [\text{V}_{\text{P}}\ \ldots\text{V}\ldots]]]$

b. $[_{\text{TP}}\ \text{T}_{[\text{–Past, +D}]}\ [_{\text{NegP}}\ \text{Neg}\ [\text{V}_{\text{P}}\ \ldots\text{V}\ldots]]]$

3. Notice that this analysis crucially rests on the assumption that NegP is lower than TP in Arabic, a claim that will be empirically contested in the next section. Notice further that this analysis treats *maa* and *-š* as one composite element generated under the Neg head, an issue that will be also discussed in Section 3.

Under this approach, surface negation patterns reflect operations of head movement (or lack thereof) taking place in the syntax.

Despite its ability to account for a good range of sentential negation facts in Arabic dialects, a syntactic analysis along the above lines will always face an empirical problem accounting for the wide range of variation in sentential negation patterns that are attested in Arabic dialects. The main challenge has always been the fact that the distribution of the two negation patterns does not seem to follow a verbal-nonverbal or a past-nonpast contrast, and this is true both within the same dialect as well as across different dialects.

On the one hand, each negation pattern occurs in a variety of different grammatical contexts within the same dialect. For example, in CEA, in addition to occurring with perfective verb forms, as in (1), discontinuous negation also occurs with the present tense aspectual imperfective (6a), pronominals (6b), the existential expletive *fii(h)* (6c), and PPs whose complement is a pronominal clitic (6d).

(6) a. maa-bi-y-ruħ-š ʔil-gamʕa
NEG-ASP-IPFV-go.3SGM-NEG the-university
'He doesn't go to the university.'

b. maa-huu-š/maa-huwwa-a-š muhandis
NEG-he-NEG/NEG-he-EV-NEG engineer
'He is not an engineer.'

c. maa-fii-š ħadd hinaa
NEG-in.it-NEG someone here
'There is nobody here.'

d. maa-ʕand-uu-š ʕarabiyya
NEG-at-him-NEG car
'He doesn't have a car.'

Similarly, in addition to copular structures like (2), the independent negation marker *miš* may occur optionally with the present tense aspectual imperfective (7a), obligatorily with future verb forms (7b), and quite marginally with copular structures with predicate PPs (7c).

(7) a. miš bi-y-ruuħ ʔil-gamʕa
NEG ASP-IPFV-go.3SGM the-university
'He doesn't go to the university.'

b. miš ħa-saafir ʔiṭaalya taanii
NEG FUT-IPFV.travel.1SG Italy again
'I will not travel to Italy again.'

c. ?? miš ʕand-u ʕarabiyya
NEG at-him car
'He doesn't have a car.'

On the other hand, cross-dialectal variation shows that certain categories are able to host negation in some dialects, but not in others. For example, nouns and adjectives in CEA cannot host negation, but they can do so in Moroccan Arabic (MA), as discussed in Benmamoun (2000), and in Siʕiidi Egyptian Arabic, a dialect mainly spoken in southern Egypt (cf. Khalafallah 1969). Compare the ungrammatical negation patterns from CEA in (8) to the grammatical negation patterns from MA in (9), the latter taken from Aoun et al. (2010: 101).

(8) a. *Ahmad maa-doktoor-š
Ahmad NEG-doctor-NEG

b. *Ahmad maa-taʕbaan-š
Ahmad NEG-tired-NEG

(9) a. huwa maa-fəllaħ-š
he NEG-farmer-NEG
'He is not a farmer.'

b. huwa maa-ṭwil-š
he NEG-tall-NEG
'He is not tall.'

Given the lack of complementary distribution between the two negation patterns as well as the variation with regard to which categories may host negation across Arabic dialects, it is reasonable to conclude that the distribution of the two negation patterns is morphologically, rather than syntactically, conditioned. Under what I will refer to in this paper as the *morphological approach*, the negation morpheme is an affix in need of a host, which, in turn, needs to be adjacent to the affix for word-formation to take place.[4] Certain categories can serve as hosts for the negation affix, whereas others cannot, depending on the dialect.

At the same time, from a theoretical perspective, this approach to the distribution of negation patterns in Arabic dialects is compatible with a theory

4. While I will use the term 'morphological approach' to describe the analysis presented here for lack of a better term, it should be clearly understood that the main assumption of the proposal is that the head movements responsible for the Spell-out of negation patterns take place in the mapping from the syntax to the morphological component. As will be seen later in the paper, the syntactic structure of negative sentences handed over to the morphology is still an essential part of the proposed analysis.

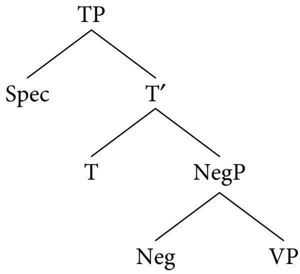
that takes head movement to be an operation of the morphological component, as has been suggested in Chomsky (2001) and Boeckx and Stjepanović (2001). Such an approach has the advantage of avoiding the recalcitrant theoretical problems that have been typically raised with regard to the status of head movement in the minimalist literature (e.g., its failure to satisfy the extension condition, its violation of the condition on chain uniformity, and the fact that it does not have semantic effects at LF).[5] Furthermore, relegating sentential negation in Arabic dialects to the morphological component saves us the trouble of having to invoke a set of ad hoc features in the syntax to generate the attested patterns, and filter out the unacceptable ones. Instead, such movements can take place in the mapping from the syntax to the morphology driven by the affixal properties of the heads involved, giving rise to the observed surface negation patterns. In Section 4 of this paper, I provide an implementation of such a post-syntactic analysis of CEA negation. Before we do that, however, we need to discuss two further issues in the morphosyntax of negation in CEA: The position of Neg in clause structure and the status of the enclitic *-š*. I do that in the next two sections.

3. The position of NegP on the clausal hierarchy in Arabic dialects

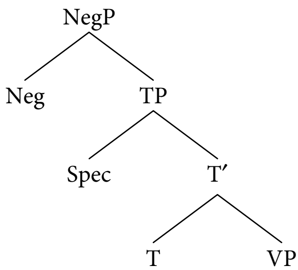
The second major issue in the syntax of negation in Arabic dialects has to do with the position of Neg in clausal structure. In this regard, one can identify two main approaches: (i) The *low-Neg* analysis, adopted in Benmamoun (2000), Ouhalla (2002), and Aoun et al. (2010), whereby Neg occupies a position lower than T; and (ii) the *high-Neg* analysis, assumed in Diesing and Jelinek (1995) and Soltan (2007), whereby Neg is located higher than T. Both approaches are diagrammatically represented in (10a,b) below, respectively.

5. Another problem with head movement is the so-called 'traffic rule' problem, as pointed out in Chomsky (2001). To illustrate, if T has both a [+D] and [+V] feature, the standard way of licensing these features is for a DP to move to SpecTP to license the [+D] feature, and for a verb to move to T to license the [+V] feature. There is nothing in principle, however, that prevents the checking of the categorial features of T from proceeding as follows: Move a VP to SpecTP to check the [+V] feature, coupled with a movement of a D head to T to check the [+D] feature. Further stipulations are needed to make sure that this 'traffic rule' violation does not occur. If head movement is relegated instead to the morphological component, no such problem arises. For a more recent discussion on the status of head movement in minimalist syntax, see Roberts (2010).

(10) a. Low-Neg approach



b. High-Neg approach



While a good range of negation facts can be accounted for under either structure in (10), there are two main empirical facts from attested negation patterns in Egyptian Arabic that pose a problem to the low-Neg analysis. I discuss each in turn.

The first empirical argument against the low-Neg analysis is that it fails to account for dialects where the independent negation marker *miš* actually appears with perfective verb forms, a pattern that is predicted to be unattested under a low-Neg analysis, due to the HMC. One such dialect is spoken in some areas of the Sharqiyyah province in northern Egypt, where sentences like (11) readily occur.

(11) ʔanaa miš liʕib-t
I NEG play.PERF.1SG
'I did not play.'

If Neg were lower than T, the negation pattern in (11) cannot be derived without V first skipping over Neg on its way to T, followed by Neg moving over the T complex, to generate the right word order. Both movements violate the HMC.[6] In addition, it is not clear how to motivate Neg-movement in that context. In brief, the negation pattern in (11) is simply underivable under standard assumptions, if Neg

6. I follow the standard assumption in the literature on Arabic syntax that perfective verbs have to move to T. For a full discussion of the issue as well as a range of empirical arguments in support of this assumption, see Benmamoun (2000) and Aoun et al. (2010).

were indeed projected below T. By contrast, under the high-Neg analysis, this negation pattern follows if in this particular dialect past tense T is not required to raise to Neg, hence giving rise to the *miš* pattern instead.[7]

A second related empirical fact in favor of the high-Neg analysis and against the low-Neg analysis comes from Egyptian children's speech, where the pattern of negation exemplified in (11) is rather common among children early on in their acquisition of negation in Egyptian Arabic (cf. Omar 1967). This means that there is a stage in negation acquisition where children overgeneralize the use of the *miš* pattern to all verb forms. If Neg were lower than T by default, as it is assumed under the low-Neg analysis, these utterances by children would be very hard to explain, given the standard assumption that the HMC is a universal principle of grammar. Under the high-Neg analysis, a possible explanation is available if children first assume that T, no matter what its tense specification is, does not need to raise to Neg; hence the use of the independent negation marker *miš* in such early utterances. Later on, based on positive evidence in the primary linguistic data, children realize that Neg has to conflate with past tense T (among other heads), and the discontinuous negation pattern emerges to replace these early non-adult-like *miš* pattern utterances.[8]

To sum up the discussion in this section, there is strong evidence from negation patterns in Sharqiyyah Egyptian Arabic as well as negative utterances produced by Egyptian children in the early stages of language acquisition that Neg has to be higher than T in CEA clause structure, and presumably in other Arabic dialects that exhibit similar negation patterns. A low-Neg analysis cannot account for such empirical facts.[9]

7. Note that, for most EA speakers, (11) is not a grammatical sentence, since Neg is always required to conflate with an adjacent T that is specified for past tense.

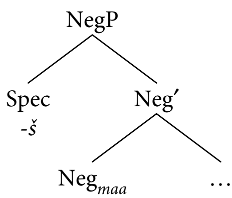
8. A reviewer points out that the acquisition facts regarding negation in EA may still be compatible with a low-Neg analysis, if we assume that a child's syntactic structure does not have TP at this stage yet. In the absence of a TP layer, the reviewer argues, the emerging pattern would not be surprising. This would have been a possible explanation if EA child language shows the occurrence of root infinitives, for example. But this is not the case, given the way Semitic verbal morphology works in terms of tri-consonantal roots and patterns (McCarthy 1979 and subsequent works), and the fact that both the perfective and imperfective patterns occur early in the acquisition process, thereby suggesting the presence of both the T and Asp layers in early child language. There is no infinitive category in the language that would occur if TP were missing. So, the acquisition evidence cited in (11) does show that there is indeed a TP layer in EA child speech at (if not prior to) the stage where children produce this particular form of negation.

9. An implicit assumption here, familiar from Ouhalla (1991) and Zanuttini (1997), is that the position of Neg is parametric: Some languages place Neg higher than T (e.g., CEA); others project T lower than Neg (e.g., Germanic).

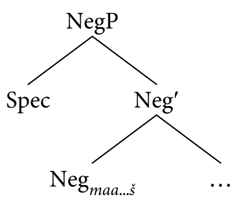
4. The grammatical status of the -*š* enclitic of the negation morpheme

A typical surface property of the negation morpheme in CEA (and in other dialects with similar bipartite negation) is that it includes a -*š* segment, which appears as an enclitic in the *maa…š* pattern, and as a subpart of the negation marker in the *miš* pattern. A question arises as to the grammatical status of -*š*. One potential analysis is to treat -*š* as a Spec of the Neg head *maa*, similar to what has been suggested for bipartite negation in languages like French (cf. Pollock 1989; Ouhalla 1990; Moritz and Valois 1994). An alternative analysis, argued for in Benmamoun (2000) and Aoun et al. (2010), treats the negation morpheme as a discontinuous Neg head, under which the two negative segments are generated.[10] The two analyses are represented in (12a,b), respectively.

(12) a.



b.



Aoun et al. argue that the discontinuous head analysis allows us to account for the variation attested in negation patterns in Arabic dialects (e.g., the fact that some dialects mark negation with *maa* only, while others use -*š* only). It is possible, however, for the analysis in (12a) to do the same. For example, dialects that mark negation with *maa* only can be argued to not project an inherent SpecNegP, while those that mark negation with -*š* only can be argued to have lost *maa* as a negative head and have instead come to treat -*š* as the head of NegP.[11]

10. Ouhalla (2002) proposes to treat -*š* as a variable bound by the negation operator. Syntactically, this analysis is more in the spirit of the SpecNegP analysis discussed in the text. In this paper, I adopt the standard assumption that negation does not create an operator-variable dependency.

11. Cross-dialectal variation in negation strategies in Arabic dialects reflects the workings (or lack thereof) of the *Jespersen's cycle* (Jespersen 1917). I will get back to this issue with illustrating data from different Arabic dialects in Section 6.

On the other hand, there is one grammatical context that seems to favor treating *-š* as SpecNegP rather than as part of a composite Neg head, namely, structures with so-called Negative Polarity Items (NPIs).[12] More specifically, it has been noted that in some Arabic dialects the *-š* enclitic is in complementary distribution with NPIs (Benmamoun 1997, 2006; Bahloul 1996). Consider, for example, the MA examples in (13), cited from Benmamoun (2006: 143).

(13) a. ma-qrit(*-š) ħətta kitab
NEG-came.1SG even book
'I didn't read any book.'

b. ma-ža(*-š) ħətta waħəd
NEG-came.3SGM even one
'No one came.'

c. ħətta waħəd ma-ža(*-š)
even one NEG-came.3SGM
'No one came.'

d. Nadya ʕummɔr-ha ma-žat(*-š)
Nadya ever-her NEG-came.3SGF
'Nadya never came.'

e. Omar baqi ma-ža(*-š)
Omar yet NEG-came.3SGM
'Omar hasn't come yet.'

This fact can be accounted for under the Spec-analysis of *-š*, if we assume that an NPI and *-š* both compete for SpecNegP. The discontinuous head analysis, however, does not have a natural way of explaining this fact. Rather, it has to assume a rule at the sub-morphemic level, and stipulate that such a rule can only target the *-š* enclitic but not the *maa* proclitic of the composite head.

Even though these NPI facts seem to suggest that treating *-š* as Spec of NegP has direct empirical consequences for MA, the same NPI facts in CEA suggest that this analysis cannot be maintained, since NPI licensing in CEA is not always in complementary distribution with *-š*. Rather, the only NPI that induces *-š*

12. Benmamoun (1997, 2006) uses the term NPIs in a general sense to refer to lexical elements that are sensitive to the polarity of the sentence, for which sometimes the term polarity-sensitive items is used instead (cf. Soltan 2014). To avoid any confusion that may result from using both terms, I will follow Benmamoun in the rest of the paper in using the term NPIs in its general sense, which refers to both elements of the *any*-type as well as *N-words* (the latter also called *negative concord items* in the literature). Nothing in the analysis of the relevant empirical facts hinges on this terminological distinction. For a discussion of polarity-sensitive items in EA, see Soltan (2014). See also footnote 21.

deletion is *ʕumr* (= 'ever'), and it does so only when it occurs in pre-negative (but not in post-negative) position, as shown in (14). The occurrence of other NPIs does not induce any such effect, and the surfacing of *-š* is obligatory in such contexts, as the data in (15–16) show for the NPIs *lisssah* (= 'yet'), *ʔayy* (= 'any'), and *xaaliṣ* (= 'at all').[13]

(14) a. ʕumr-ii maa-safir-t(*-š) Masr
ever-my NEG-travel.PERF-1SG-(*NEG) Egypt
'I have never travelled to Egypt.'

b. maa-safir-t*(-š) Masr ʕumr-ii
NEG-travel.PERF-1SG-*(NEG) Egypt ever-my
'I have never travelled to Egypt.'

(15) a. Mona lissah maa-safir-it-*(š)
Mona yet NEG-travel.PERF-3SGF-*(NEG)
'Mona has not travelled yet.'

b. Mona maa-safir-it-*(š) lissah
Mona NEG-travel.PERF-3SGF-*(NEG) yet
'Mona has not travelled yet.'

(16) a. ʔanaa maa-šuf-t-i-*(š) ʔayy ħaaga
I NEG-see.PERF-1SG-EV-NEG any thing
'I didn't see anything.'

b. ʔanaa maa-šuf-t-i-*(š) ħaaga xaaliṣ
I NEG-see.PERF-1SG-EV-NEG thing at all
'I didn't see anything at all.'

I take the asymmetry in behavior between different NPIs as evidence against the SpecNegP analysis of the *-š* segment.[14] I conclude, therefore, that neither the Spec-head analysis nor the discontinuous head analysis of negation is adequate enough to account for the CEA facts, hence the need for an alternative analysis. I propose this next.

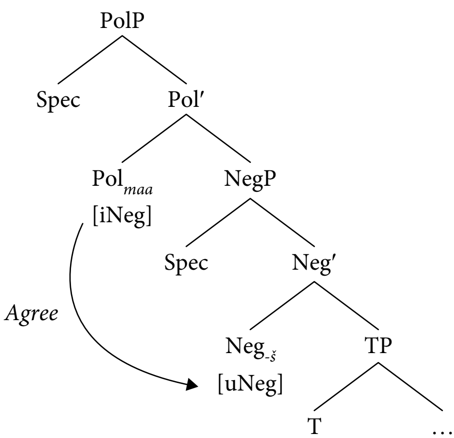
13. I follow Benmamoun (1997) in assuming that NPI-licensing in Arabic dialects may take place either under a Spec-head relation with Neg or under c-command by Neg. This is needed since these dialects allow NPIs to occur in pre-negative as well as post-negative position, as illustrated by the data in (14–15), for example.

14. Another argument, though a theory-internal one, has to do with whether or not multiple specifiers of a single head are allowed. In a framework that permits multiple specifiers (e.g., Chomsky 1995), accounting for the complementary distribution between two elements in terms of their 'competing' for a single Spec is not that straightforward.

5. Splitting Neg

To account for the morphosyntax of negation in CEA as well as a range of other empirical facts, I sketch out a *split-Neg analysis* in Soltan (2011), whereby *maa* and *-š* are treated as separate heads, located higher than T, but only *maa* is specified for semantic negation, while *-š* is marked solely for formal negativity, a property it may have acquired diachronically as it developed from the adverbial usage of the noun *šayʔ* (= 'a thing') as a minimizer to its usage as a negation marker (cf. the discussion in Section 6 below on the diachrony of *-š*). Under this analysis, *maa* is a (Pol)arity head, the locus of interpretable negation (marked as an [iNeg] feature), and *-š* is a Neg(ative) head specified for an uninterpretable negative feature (marked as [uNeg]).[15] Licensing of [uNeg] takes place under *Agree* between Pol and Neg, in a (modified) sense of Chomsky (2000, 2001), as illustrated in the tree in (17), ignoring irrelevant details up and down the tree.[16]

(17)



15. The split-Neg analysis adopted here is an adaptation of the approach to negation and negative concord proposed in Zeijlstra (2004, 2008). It should be pointed out that nothing hinges on the labels assigned to the two heads here. I follow Zanuttini (1997) in assuming that negation is expressed via a Polarity Phrase, and choose to use the standard term Neg for the formally negative enclitic *-š*. I find that to be a more convenient labeling than to call one head Neg_1 and the other Neg_2, which would not do a good job of highlighting the inherent lexical difference between the two heads as proposed here.

16. This is a modified version of Agree because the so-called *Probe* is actually an interpretable feature looking for a *Goal* with a matching uninterpretable feature in its c-command domain. This has been suggested in a number of places in the minimalist literature on syntactic agreement; see Zeijlstra (2004, 2008) and references cited there.

The presence of the formally negative head *-š* does not induce a double negation reading since its [uNeg] feature gets checked off and deleted by the [iNeg] feature of Pol, thereby generating only a single negation reading at LF.[17]

Given the structure in (17), we are in a position to formulate a morphological algorithm to derive the distribution of negation patterns in CEA, one along the lines of (18) below, where 'hosting head' is the key locus of dialectal variation.

(18) a. In contexts where Neg is adjacent to a hosting head *H*, *H* moves to Neg and then to Pol, and the discontinuous *maa-H-š* pattern arises.
b. Otherwise, Neg incorporates into Pol, giving rise to the *miš* pattern.

To illustrate from the contrast between perfective verb forms and prospective imperfective forms (cf. the two sentences in 1 and 7b), the two negation patterns are derived as in (19a,b), respectively, irrelevant details aside.

(19) a. $[_{\text{PolP}}\ \text{Pol}\ [_{\text{NegP}}\ \text{Neg}\ [_{\text{TP}}\ \text{T}_{[+\text{PAST}]}\ [_{v\text{P}}\ v\ [_{\text{VP}}\ \text{V} \ldots]]]]] \rightarrow$ [*maa-nim-t-i-š*]

b. $[_{\text{PolP}}\ \text{Pol}\ [_{\text{NegP}}\ \text{Neg}\ [_{\text{TP}}\ \text{T}_{[-\text{PAST}]}\ [_{\text{AspP}}\ \text{Asp}\ [_{v\text{P}}\ v\ [_{\text{VP}}\ \text{V} \ldots]]]]]]$ → [*miš ħa-saafir*]

Similar derivations can be given for each structure associated with one of the two negation patterns, where the presence of a hosting head gives rise to discontinuous negation (cf. the examples in 6), or, otherwise, the independent pattern surfaces (cf. the data in 7). Thus, structures where Neg is adjacent to $\text{T}_{[+\text{PAST}]}$, a pronominal, the existential expletive *fii(h)*, and PPs whose complement is a pronominal clitic are derived under step (18a).[18] By contrast, structures where Neg is adjacent to $\text{T}_{[-\text{PAST}]}$ in a verbless sentence or to a prospective aspectual head are derived under step (18b).[19] Structures where Neg is adjacent to other non-prospective aspectual

17. This is comparable to what happens in negative concord sentences, where the presence of multiple negative elements does not lead to a double negation interpretation, under the assumption that negative concord items are formally, but not semantically, negative (see, for example, Zeijlstra 2008).

18. To be more explicit, overt pronominal and expletive subjects occupy SpecTP, hence are adjacent to Neg and Pol. Similarly, PPs host negation only when they occur as fronted predicates, hence, they are also assumed to occupy SpecTP in such cases via predicate inversion, and are, therefore, a closer host to Neg and Pol than T is. These three categories, thus, behave like XPs for theta-assignment purposes and as X^0's for the purpose of head movement, comparable to clitics in other languages (e.g., Romance).

19. $\text{T}_{[-\text{Past}]}$ is always null in Arabic dialects, as evidenced by the absence of a present tense form of the copula, giving rise to so-called verbless sentences (see Benmamoun 2000 and references cited there for a discussion). It is no surprise, therefore, that it cannot function as a host for negation in such contexts, hence the occurrence of the *miš* pattern.

heads (e.g., habitual or progressive Asp) may be derived under either step in (18), thereby giving rise to the compatibility of either negation pattern with these particular verb forms (cf. 6a and 7a).

Similarly, under this analysis, dialectal variation with regard to which syntactic categories may host negation is a purely morphological fact about each dialect. For example, at Spell-out, nouns and adjectives adjacent to Neg and Pol behave under step (18a) in MA, but under step (18b) in CEA, thereby explaining the contrast in negation between the examples given earlier in (8) and (9). By the same token, the proposed analysis allows us to account for the range of dialectal variation noted earlier with regard to the occurrence of discontinuous negation with perfective verb forms, as in Sharqiyyah Egyptian Arabic and child language (cf. 11). Under this analysis, V in such dialects raises all the way to $T_{[+PAST]}$, but does not act as a hosting head, forcing step (18b) to apply. In Section 6 below, I discuss further cross-dialectal implications of the analysis proposed in this section.

In sum, a split-Neg analysis of sentential negation allows us to derive the distribution of the two negation patterns in CEA in the mapping from the syntax to the morphology, where notions such as affixality, adjacency, and hosting heads play a central role. It remains to see if the proposed analysis can also account for empirical facts outside the core facts of negation patterns. I explore three types of such empirical consequences in the next section.

6. Empirical consequences of the split-Neg analysis

In this section I explore the implications of a split-Neg analysis of negation in CEA in three empirical domains: The behavior of the *-š* enclitic in NPI contexts, the cross-dialectal variation in negation strategies in Arabic dialects in terms of the Jespersen's Cycle, and constructions in CEA where sentential negation is marked by *(wa)laa*. I discuss each in turn.

6.1 Empirical consequences I: Revisiting the behavior of *-š* in NPI contexts

In Section 4, I pointed out the morphosyntactic discrepancy in the behavior of different NPIs in CEA with regard to their interaction with the *-š* enclitic: When the NPI *ʕumr* (= 'ever') occurs in pre-negative position, *-š* is obligatorily deleted from the verbal complex. By contrast, the occurrence of the NPI *lissah* (= 'yet') in pre-negative position does not induce a similar effect, and the *-š* enclitic has to obligatorily surface on the verb. One advantage of the split-Neg analysis, as I argue in Soltan (2012), is that it allows us to provide a principled account for this morphosyntactic fact by targeting *-š* for deletion under certain conditions, a solution that is not readily available under either the SpecNegP or the discontinuous head analyses, as pointed out earlier in Section 4. In this subsection, I present a summary of the account in Soltan (2012) as one example of the kind of empirical consequences that the analysis presented here can have.

As it turns out, the key to the solution of the puzzle of *-š* deletion in CEA has to do with the 'formal negativity' (or lack thereof) associated with different NPIs. In particular, I argue, on the basis of diagnostic tests, that some NPIs are formally marked as negative, while others are not so marked, and that the overt realization of *-š* is only compatible with NPIs that have matching formal features.[20] Formal negativity can be determined by an item's synchronic behavior in the language and possibly by considering its diachrony as well. Below, I discuss this with regard to the two NPIs *ʕumr* and *lissah*, as well as the *-š* enclitic of the negation morpheme.

One test to determine the formal negativity of a lexical item has to do with whether or not it is compatible with nonnegative environments such as interrogatives or the protasis of a conditional. As it turns out, the NPI *ʕumr* may indeed occur in such contexts, as (20) shows, which suggests that it is nonnegative. By contrast, the NPI *lissah* cannot appear in questions, except in the presence of the overtly negative morpheme *wallaa* (21a), nor in conditionals (21b), which suggests that it is lexically negative.

(20) a.

ʔinta	ʕumr-ak	safir-t	Masr?
you	ever-you	travel.PERF.2SGM	Egypt

'Have you ever traveled to Egypt?'

b.

law	ʕumr-ak	safir-t	Masr	laazim	tə-zuur
if	ever-you	travel.PERF.2SGM	Egypt	must.PTCP	IPFV-visit.2SGM

ʔaswaan
Aswan
'If you ever travel to Egypt, you must visit Aswan.'

(21) a.

Ahmad	gih	*(wallaa)	lissah?
Ahmad	come.PERF.3SGM	or.not	yet

'Has Ahmad come or not yet?'

b.

*law	Ahmad	gih	lissah ...
if	Ahmad	come.PERF.3SGM	yet

'*If Ahmad arrived yet, ...'

20. The proposal is familiar from work on negative concord items, which have been argued to be 'negative,' as opposed to NPIs of the *any*-type, which are typically assumed to be nonnegative. For a recent discussion of negative concord, see Penka (2011) and references cited there.

A second diagnostic for the formal negativity of an NPI is whether it may occur as a fragment answer. As it turns out, while *ʕumr* does not occur in that function (22b), *lissah* does (23b).[21]

(22) a. ʔinta safir-t Masr ʔabl kida?
you travel.PERF-1SGM Egypt before this
'Have you traveled to Egypt before?'

b. *ʕumr-ii
ever-my
'Never.'

(23) a. huwwa Mona waṣal-it?
Q Mona arrive.PERF.3SGF
'Has Mona arrived?'

b. lissah
yet
'Not yet.'

Consider now the distribution of the -*š* enclitic. Like *lissah*, but unlike *ʕumr*, the -*š* enclitic may not occur in questions or conditionals, as the ungrammaticality of the examples in (24) shows, thereby suggesting that it is also an element lexically marked for formal negativity.

(24) a. *šuft-i-š Ahmad ʔin-nahaar-da?
see.PERF.2SGM-EV-NEG Ahmad the-day-this
Intended reading: 'Did you see Ahmad today?'

b. *law šuft-i-š Ahmad ʔin-nahaar-da ...
if see.PERF.2SGM-EV-NEG Ahmad the-day-this
Intended reading: 'If you saw Ahmad today, ...'

On the other hand, when we consider the diachrony of the three items, we also arrive at the same conclusions regarding the negativity status of each. *ʕumr* is historically derived from the noun *ʕumr* (= 'life/age'), and is, thus, nonnegative in origin. The origin of *lissah*, by contrast, is not as clear, but it may have been derived from *laysa*, a negation marker from Classical Arabic, or from the phrase

21. A reviewer points out that since *lissah* can function as a fragment answer, it should be treated as an N-word (in the sense of Laka 1990), not an NPI. Recall from footnote 12 that I am using the term NPI in its general sense of polarity-sensitive items. That terminological issue aside, the reviewer's observation is in full conformity with the analysis proposed in the text: Some elements that are licensed in the scope of negation induce -*š* deletion, but others do not. The proposed analysis accounts for this by appeal to the featural negativity (or lack thereof) of such terms, a property that can be empirically tested via the diagnostics used and illustrated in this section of the paper.

li-s-saaʕah (literally, 'to the hour') when used in negative contexts. As for *-š*, it is generally assumed to be a phonological reduction of the Classical Arabic noun *šayʔ* (literally = 'a thing') in its accusative adverbial NPI function, as in the Quranic verse in (25) below, hence its origin is also probably negative (Lucas 2010).

(25) ʔinna ʔallah-a laa ya-ð̣lim-u ʔan-naas-a šayʔ-an
COMP Allah-ACC NEG IPFV-be.unjust-IND the-people-ACC thing-ACC
wa-laakinna ʔan-naas-a ʔanfus-a-hum
and-but the-people-ACC selves-ACC-their
ya-ð̣lim-uun
IPFV-be.unjust-IND (*Qur'an* 10: 44)
'Allah is not unjust to people one bit; it is they who are unjust to themselves.'

To sum up the discussion so far, given the synchronic and diachronic evidence presented above, we may conclude that while the NPI *ʕumr* is nonnegative, both the NPI *lissah* and the *-š* segment of the negation morpheme appear to be formally negative. Adding the negation marker *maa*, which is the locus of semantic negation, to the list, we can summarize the grammatical properties of the four items as in the table in (26).

(26)	*-š*	*lissah*	*ʕumr*	*maa*
Diachronic origin	Noun used as an NPI	Perhaps from a negative usage	Noun meaning 'age/life'	Negative morpheme
Compatibility with nonnegative contexts	No	No	Yes	
Occurrence as a fragment answer	N/A	Yes	No	
Negativity status	Formal	Formal	Nonnegative	Semantic

Given (26), we are in a position to restate the empirical fact regarding the surfacing of *-š* (or lack thereof) in CEA NPI contexts, as follows: The *-š* enclitic, a formally negative element, disappears in the presence of a nonnegative NPI such as *ʕumr*, but is retained in the presence of a negative NPI such as *lissah*. The phenomenon, however, is sensitive to locality: *-š* disappears only when *ʕumr* is 'close by' (i.e., in pre-negative position), but not when it is relatively distant (i.e., in postverbal position), as shown earlier by the contrast between (14a) and (14b).[22] We may

22. Notice that *ʕumr* still has to be semantically licensed under c-command by an appropriate operator, either *downward entailing* in the sense of Ladusaw (1979), or *nonveridical* in the sense of Giannakidou (1998). Polarity-sensitive items like *lissah*, by contrast, have to be both formally and semantically licensed given their featural make-up.

thus restate this morphosyntactic fact in the form of the descriptive generalization in (27), where Neg is used to refer to *-š*, as it is narrowly defined in this paper.

(27) Within a local domain, Neg is not Spelled-out in the presence of an NPI that is formally nonnegative; otherwise it is phonologically realized.

In Soltan (2012), I propose to derive the generalization in (27) from a general interface condition on the Spell-out of formal features licensed in the syntax, as in (28), where 'local domain' is taken to correspond to 'a phase,' in the sense of Chomsky (2001), i.e., CP and *v*P.[23]

(28) *Minimize Formal Feature Mismatch* (MFFM): At Spell-out, minimize formal feature mismatch on licensees of the same licenser within a local domain.

If this analysis is on the right track, it provides an empirical argument in support of a split-Neg analysis, since only by splitting the Neg-domain into two heads, we are able to formulate a rule to target *-š* for deletion. Such a possibility is not straightforwardly available under non-split-Neg analyses of CEA negation.[24] In the rest of this section, I discuss two other grammatical contexts where *-š* deletes, showing that they also follow from the MFFM principle in (28).[25]

The phenomenon of *-š* deletion discussed in this section also takes place in some other restricted syntactic environments in CEA. Two idiomatic expressions that induce *-š* deletion are *ʕannak* and *ʔinšallah*. As an idiom, *ʕannak*, which is literally a PP meaning 'about you,' roughly translates as 'I don't give a damn [if you don't do X],' when used in negative contexts. Similarly, *ʔinšallah* (which is a phonologically modified form of *ʔin shaaʔa llaah*, the common expression meaning 'God's willing' in Arabic dialects), when used in negative

23. One advantage of the formulation of MFFM in (28) is that it predicts *-š* deletion in contexts where the mismatch has to do with other features than just formal negativity, as will be discussed later in this section.

24. A question arises regarding the behavior of *-š* in MA NPI contexts discussed earlier in Section 4 (cf. the examples in 13). There are two possibilities to consider here. One is that, unlike in CEA, principle (28) in MA is not sensitive to locality, hence forcing *-š* to delete nonlocally. Another potential explanation is that perhaps MA developed a French-style Spec-head configuration for its bipartite negation rather than a split-Neg structure, thereby explaining the complementary distribution between *-š* and NPIs in the dialect. At this point, I have no evidence to favor one explanation over the other.

25. There seems to be cross-linguistic evidence for the feasibility of MFFM, since multiple licensing configurations discussed in the minimalist literature are typically characterized by feature match, e.g., multiple nominative structures in Japanese derived via Multiple Agree (Hiraiwa 2001) and multiple negative concord structures in West Flemish derived via Binary Agree (Haegeman and Lohndal 2010). Such facts suggest that MFFM is on the right track as a general condition on multiple licensing configurations.

sentences, also has the admonishing interpretation 'I don't give a damn [if you don't do X].' The following two examples illustrate the usage of *ʕannak* and *ʔinšallah* in CEA.

(29) a. ʕann-ak maa-geet(*-š)
about-you NEG-come.PERF.2SGM-(*NEG)
'I don't give a damn if you don't come.'

b. ʔinšallah maa-geet(*-š)
God's willing NEG-come.PERF.2SGM-(*NEG)
'I don't give a damn if you don't come.'

It is reasonable to assume that both expressions have some feature that needs to be licensed by negation in such contexts. No matter what the exact nature of that feature may be, the presence of such expressions will always create a grammatical context where *-š* is in the vicinity of an XP with a mismatching feature, thereby forcing *-š* to delete, as observed in (29).

The same can be said about the oath expression *wallaahi* (= 'by God'), which, when used in negative contexts, is typically incompatible with *-š*, particularly when it is phonologically stressed, and/or syntactically modified by the adjective *ʔil-ʕaẓiim* (= 'the great'), to signal a more emphatic denial.[26]

(30) a. wallaahi l-ʕaẓiim maa-ʔaxad-t(*?-š) minn-uh ħaaga
by God the-great NEG-take.PERF.1SGM-(*?NEG) from-him thing
'I swear I did not take anything from him.'

b. wallaahi l-ʕaẓiim maa-ʕraf(*?-š) ʕann-uh ħaaga
by God the-great NEG-IPFV.know.1SGM-(*?NEG) about-him thing
'I swear I do not know anything about him.'

Under the analysis adopted in this section, this follows if, in those contexts where *wallaahi* carries an emphasized denial feature, it has to be licensed by semantic negation (Pol in our terms), which, in turn, creates feature mismatch with *-š* at the point of Spell-out, leading to *-š* deletion. What is interesting, however, is that, unlike the idiomatic expressions in (30), *wallaahi* may actually appear in post-negative position. When that happens, *-š* has to obligatorily surface, as shown by (31), whether or not *wallaahi* is phonologically and/or syntactically marked for emphasized denial, which is what we expect given that principle (28) is sensitive to locality.

26. As the '*?' indicates, the occurrence of *-š* in such contexts is not totally ungrammatical; it is definitely less preferable, though. I have no explanation at the moment for this case of grammaticality gradation.

(31) a. maa-ʔaxad-t-i-*(š) minn-uh ħaaga wallaahi l-ʕaẓiim
NEG-take.PERF-1SGM-EV-*(NEG) from-him thing by God the-great
'I swear I did not take anything from him.'

b. maa-ʕraf-*(š) ʔann-uh ħaaga wallaahi l-ʕaẓiim
NEG-IPFV.know.1SGM-*(NEG) about-him thing by God the-great
'I swear I do not know anything about him.'

In sum, I have argued in this section that an explanation for the morphosyntactic puzzle of *-š* deletion in polarity-sensitive contexts in CEA follows from a (potentially general) interface principle that regulates Spell-out in multiple licensing configurations. I have also shown that the principle proposed can naturally extend to account for other grammatical environments where *-š* gets deleted. I take that as evidence for the empirical adequacy of the split-Neg analysis.

6.2 Empirical consequences II: Dialectal variation and the Jespersen's cycle

In this section, I consider the empirical consequences of a split-Neg analysis for the study of negation variation across modern Arabic dialects. Since a full discussion of such variation is beyond the scope of this paper, the section mainly aims to provide the general implications of the proposed analysis for the major negation strategies used in today's Arabic dialects.

One tacit assumption of the analysis presented here is that splitting of Neg in CEA must have occurred at one point in the historical development of the dialect, reflecting the workings of the *Jespersen's cycle* (Jespersen 1917), which is typically posited in accounts of diachronic changes in negation marking in human languages. This predicts that dialects with no discontinuous negation will not have *-š* (or some similar morpheme) in the make-up of their independent negation marker. This is indeed the case in Syrian and Kuwaiti Arabic, for example, as discussed in Brustad (2000), who provides a comparative account of negation strategies in four Arabic dialects: Egyptian, Moroccan, Syrian and Kuwaiti. Since I have already illustrated sentential negation patterns in CEA in this paper, I only provide examples of sentential negation in Moroccan, Syrian, and Kuwaiti below.

(32) a. Nadia ma-ža-t-š (Moroccan; Benmamoun 2000: 69)
Nadia NEG-come.PERF.3SGF-NEG
'Nadia didn't come.'

b. Nadia ma-ši fə-l-madrsa (Moroccan; Benmamoun 2000: 73)
Nadia NEG-NEG in-the-school
'Nadia is not at school.'

(33) a. maa b-tu-fruq maʕ-u (Syrian; Brustad 2000: 280)
NEG ASP-IPFV-differ.3SGF with-him
'I doesn't make any difference to him.'

b. muu muʕallme (Syrian; Brustad 2000: 280)
NEG educated
'(She's) not educated.'

(34) a. maa yi-ħibb-ha (Kuwaiti; Brustad 2000: 280)
NEG IPFV-love.3SGM-her
'He doesn't love her.'

b. muu ħilwa (Kuwaiti; Brustad 2000: 280)
NEG pretty
'(She's) not pretty.'

Brustad gives *Table 1* to illustrate the contrasts in negation markers between the four dialects.[27]

Under the analysis proposed here, the negation marker distribution shown in Table 1 follows straightforwardly: Dialects like Moroccan and Egyptian, which underwent the Jespersen's cycle, created a split-Neg structure, where Pol is *maa* and Neg is *-š*. As we should expect, this *-š* head will typically show in both negation patterns, as the Moroccan and Egyptian cells in the table show. By contrast, dialects like Syrian and Kuwaiti, which have not undergone the Jespersen's cycle,

Table 1. Sentential negation markers in four Arabic dialects (Brustad 2000: 282).[28]

	Verbal negation	**Predicate negation**
Moroccan	*maa...š(i)*	*maaši*
Egyptian	*maa...š(i)*	*miš*
Syrian	*maa*	*muu*
Kuwaiti	*maa*	*muu*

27. Brustad's 'verbal negation' and 'predicate negation' for the most part correspond, respectively, to what we have called 'discontinuous negation' and 'independent negation,' in this paper.

28. I have adjusted the phonetic transcription in Brustad's table to conform to the transcription system used in this paper.

have no split-Neg structure, and as a result, show no *-š* segment in any of their negation patterns, which is what we expect if the analysis given here is correct.[29]

Furthermore, given the Jespersen's cycle, we should predict that some Arabic dialects would eventually convert a split-Neg configuration back into a non-split-Neg structure. In other words, we should expect some dialects to drop the *maa* proclitic entirely and promote *-š* to be the semantic negation operator. This has been reported to have happened already in the Lebanese dialect of Baskinta, cited in Abu Haidar (1979), and in one Jordanian dialect reported in Palva (1972). Illustrating examples are given in (35), as cited in Benmamoun (2000: 70).

(35) a. bi-t-ħibb-š šuɣl il-bayt
ASP-IPFV-like.3SGF-NEG work the-house
'She doesn't like housework.'

b. bədd-ii-š
want-my-NEG
'I don't want.'

Similarly, Hoyt (2007) and Lucas (2010) discuss evidence that, for some speakers of Palestinian Arabic, a transition is occurring from discontinuous negation to a pattern of negation where only *-š* is used. While these speakers still produce discontinuous negation, it has become customary for them to drop the *maa* head in certain contexts, subject to certain grammatical restrictions. Lucas (2010: 191) speculates that 'successive generations of children acquiring this system will gradually fail to learn and apply this restriction, and the postverbal construction will thus be extended throughout the whole of the verbal paradigm.' In our terms, speakers of this dialect are currently alternating between a split-Neg and a non-split-Neg structure in their grammar. If this grammatical trend continues to be

29. It should be noted here that I am assuming that a split-Neg is only one manifestation of the Jespersen's cycle. Another manifestation, which is more commonly assumed in the relevant literature, is that bipartite negation arises due to an element becoming a Spec of the original Neg head in a language, as is generally assumed for the development of French *ne…pas*, for example. This leaves it as an open possibility that negation in some Arabic dialects may have developed along the French-style Spec-Neg configuration. The NPI facts cited in Section 4 from Moroccan Arabic seem to suggest that this may have been the case in this dialect. Another possibility is that the two strategies, splitting Neg or creating a Spec for Neg, are not mutually exclusive, but can develop at different stages in the history of a language. Another interesting question still is whether dialects that complete the cycle and end up with postverbal negation only also change the position of NegP hierarchically from higher than TP to lower than TP. Hopefully, future investigations of negation strategies in Arabic dialects as well as their historical development can shed more light on these questions and others. For a detailed study of linguistic cycles, including the negative cycle, see Gelderen (2008, 2011); for a recent discussion of the history of negation in Arabic and Afro-Asiatic, see Lucas (2013).

robust in the primary linguistic data, new generations of children acquiring the dialect are more than likely to end up taking the dialect in the direction of a non-split-Neg, where *maa* disappears and *-š* becomes the marker for semantic negation, which would complete the Jespersen's cycle in this dialect.

Reasons of space as well as lack of knowledge of other relevant facts in a variety of Arabic dialects do not allow me to provide a full discussion of negation variation between more Arabic dialects here. However, it is clear that the proposed split-Neg analysis has the potential to explain how contrasts in the negation systems of these dialects can be accounted for, including how the Jespersen's cycle has manifested itself differently across dialects. It is hoped that this topic will continue to garner more interest in future research on Arabic negation.

6.3 Empirical consequences III: Sentential negation with *(wa)laa* in CEA

In this section, I discuss a third pattern of negation in CEA that has not received much attention in the literature, offering a tentative analysis for its occurrence within the current proposal. This pattern, which is rather limited in its usage, occurs in negative disjunction (ND) and negative concord (NC) structures only. It is characterized by sentential negation being marked by either *laa* or *walaa* in ND structures, and by the term *walaa* in NC contexts. The use of this negation pattern is also semantically marked in the sense that it is felt to express an emphatic type of negation. I provide illustrating examples for its usage in both ND and NC constructions below.

Laa is a negative marker from Classical Arabic, but is not as commonly used to mark sentential negation in the modern dialects as *maa* is. *Laa* is still used, however, for other negative functions. For example, in CEA, *laa* is still productive in the grammatical context of negative disjunction, where it marks disjuncts in the scope of negation. In such contexts, *laa* also appears in free alternation with *walaa*, as illustrated in the sentences in (36) below. (For convenience, in the data below, I mark the free alternation between *laa* and *walaa* as *(wa)laa*, per the standard bracketing convention).[30]

(36) a. Ahmad maa-šrib-š (wa)laa ʔahwa walaa šaay
Ahmad NEG-drink.PERF.3SGM-NEG (and)-no coffee and-no tea
'Ahmad drank neither coffee nor tea.'

b. Ahmad miš ħa-yi-šrab (wa)laa ʔahwa walaa šaay
Ahmad NEG FUT-IPFV-drink.3SGM (and)-no coffee and-no tea
'Ahmad will drink neither coffee nor tea.'

30. Notice that while either *laa* or *walaa* can be used to mark the first disjunct, only *walaa* can be used before the second disjunct, as the ungrammaticality of (i) below shows.

(i) *Ahmad maa-šrib-š (wa)laa ʔahwa laa šaay
Ahmad NEG-drink.PERF.3SGM-NEG (and)no coffee and-no tea

What is relevant to the topic of this paper is that, in addition to using the regular *maa…š* and *miš* patterns in ND contexts, as in (36), a marked pattern in these constructions obtains when the *laa* or *walaa* itself is used as the sentential negation maker, as illustrated in (37).[31]

(37) a. Ahmad (wa)laa širib (walaa) ʔahwa walaa šaay
Ahmad NEG drink.PERF.3SGM and-no coffee and-no tea
'Ahmad drank neither coffee nor tea.'

b. Ahmad (wa)laa ħa-yi-šrab (walaa) ʔahwa walaa šaay
Ahmad NEG FUT-IPFV-drink.3SGM and-no coffee and-no tea
'Ahmad will drink neither coffee nor tea.'

Notice that when *(wa)laa* is used in this function, the verb may not appear with the *-š* enclitic, as (38) shows.

(38) *Ahmad (wa)laa šrib-š (walaa) ʔahwa walaa šaay
Ahmad NEG drink.PERF.3SGM-NEG and-no coffee and-no tea
'Ahmad drank neither coffee nor tea.'

A very similar pattern of negation also appears in negative concord structures, where the *negative concord item* (NCI) *walaa* can be used to mark negation.[32] Compare (39a), a typical NC sentence, with (39b), where the NCI *walaa* appears as the sentential negation marker in pre-predicate position, and again crucially without the *-š* enclitic appearing on the verb, as shown by the ungrammaticality of (39c).

(39) a. ʔanaa maa-fhim-t-i-š walaa ħaaga min
I NEG-understand.PERF.1SG-EV-NEG no thing from
kalaam-u-h
speech-EV-his
'I did not understand anything from his speech.'

b. ʔanaa walaa fhim-t ħaaga min kalaam-u-h
I NEG understand.PERF.1SG thing from speech-EV-his
'I did not understand anything from his speech.'

c. *ʔanaa walaa fhim-t-i-š ħaaga min
I NEG understand.PERF.1SG-EV-NEG thing from
kalaam-u-h
speech-EV-his

31. I gloss *(wa)laa* as Neg when used to mark sentential negation, and in its literal meaning 'and-no,' when used to mark negative disjunction, just to distinguish the two functions of the term.

32. For a discussion of NC in both Egyptian and Moroccan Arabic, see Ouali and Soltan (2014).

Given that an account of these facts requires a full study of negative disjunction and negative concord in CEA, I will only speculate here on how this marked pattern comes to emerge in these particular contexts, hoping that more research on these constructions will provide us with a less speculative analysis.

It has been proposed that NC should be analyzed as a case of syntactic agreement in terms of feature licensing in the minimalist sense (Zeijlstra 2004, 2008; Watanabe 2004; Kuno 2007; Ouali and Soltan 2014). Let us also assume that ND involves a similar operation, especially in languages like CEA where the ND operator and the NCI are (almost) identical. This means that ND *(wa)laa* and NC *walaa* are licensed via Agree with the head encoding semantic negation, Pol in our split-Neg analysis. If nothing more than long-distance Agree takes place, the usual negation patterns will emerge, giving rise to structures such as (36) and (39a).

The question that arises here is: How do we derive sentences like (37) and (39b)? Recall that such sentences are felt to be more emphatic in their negation than their counterparts in (36) and (39a). Suppose that in such emphatic contexts, Pol carries an uninterpretable focus feature, call it [uFoc], which has to get licensed by a matching interpretable feature [iFoc] on the ND marker *(wa)laa* or the NCI *walaa* via an Agree relation, as in (40).[33]

(40) $[_{\text{PolP}}$ Pol$_{\text{[uFoc]}}$ $[_{\text{NegP}}$ Neg $[_{\text{TP}}$ T $[_{v\text{P}}$ v $[_{\text{VP}}$ … *(wa)laa*$_{\text{[iFoc]}}$ …]]]]]
Agree

Suppose further that, at the PF interface, a Pol-Neg complex marked with a checked [uFoc] is spelled-out as a copy of the Agreed-with Goal. This would account for the observed negation pattern in structures such as (37) and (39b), including the fact that the *-š* enclitic may not surface with this pattern.

While a number of questions remain to be answered for the syntax of ND and NC in general, all of which are beyond the scope of this paper, it is worth pointing out again that the particular pattern of negation discussed in this section is rather marked in the language, and, therefore, it is plausible to expect its derivation to involve some 'marked' operation like Goal-copying. Whether such an operation can be independently motivated in this language and others is an empirical question that will hopefully be elucidated by further research within Agree-based approaches.

33. I assume here that the ND and NC items also carry a [uNeg] feature by virtue of them being polarity-sensitive items, which is licensed in the process of Agree-ing with Pol.

7. Conclusions

In this paper, I have proposed an analysis of the distribution of negation patterns in CEA in terms of head movement taking place in the mapping from the syntax to the morphology, where the relevant notions are affixality, hosting heads, and adjacency. I have also provided empirical evidence showing that placing Neg above T in the clausal hierarchy allows us to account for attested patterns of negation that are problematic under a low-Neg analysis. I have also shown that by splitting the negative domain into Pol and Neg, we are able not only to derive the distribution of negation patterns in CEA, but also to formulate a rule (which can potentially be reduced to the effects of a general interface principle) to target *-š* for deletion in certain NPI contexts, but not in others, hence allowing us to explain away a morphosyntactic puzzle from CEA negation contexts. In addition, I have discussed the empirical implications of the proposed analysis for variation among Arabic dialects in negation strategies, showing that a split-Neg analysis is compatible with trends in the historical development of negation as defined by the Jespersen's cycle. Finally, I illustrated a marked negation pattern occasionally used in negative disjunction and negative concord structures in CEA, offering a tentative answer as to why the grammar allows such a pattern. While a set of questions about the morphosyntax of negation in Arabic dialects have been answered here, a new set of questions also emerge, and will hopefully continue to garner more interest among Arabic scholars in future research.

References

Abu Haidar, F. (1979). *A study of the spoken Arabic of Baskinta*. Leiden: Brill.

Aoun, J., Benmamoun, E., & Choueiri, L. (2010). *The syntax of Arabic*. The UK: Cambridge University Press.

Bahloul, R. M. (1996). Negation in French and Tunisian Arabic. In M. Eid (Ed.), *Perspectives on Arabic Linguistics Vol. VIII* (pp. 67–83). Amsterdam: John Benjamins.

Benmamoun, E. (1997). Licensing of negative polarity items in Moroccan Arabic. *Natural Language and Linguistic Theory*, 15, 263–287.

Benmamoun, E. (2000). *The feature structure of functional categories: A comparative study of Arabic dialects*. Oxford: Oxford University Press.

Benmamoun, E. (2006). Licensing configurations: The puzzle of head negative polarity items. *Linguistic Inquiry*, 37, 141–149.

Boeckx, C., & Stjepanović, S. (2001). Head-ing toward PF. *Linguistic Inquiry*, 32, 345–355.

Brustad, K. (2000). *The syntax of spoken Arabic: A comparative study of Moroccan, Egyptian, Syrian, and Kuwaiti dialects*. Washington, DC: Georgetown University Press.

Chomsky, N. (1995). *The minimalist program*. Cambridge, Mass.: MIT Press.

Chomsky, N. (2000). Minimalist inquiries: The framework. In M. Roger, D. Michaels, & J. Uriagereka (Eds.). *Step by step: Essays on minimalist syntax in honor of Howard Lasnik* (pp. 89–156). Cambridge, Mass.: MIT Press.
Chomsky, N. (2001). Derivation by phase. In M. Kenstowicz (Ed.), *Ken Hale: A life in language* (pp. 1–52). Cambridge, Mass.: MIT Press.
Diesing, M., & Jelinek, E. (1995). Distributing arguments. *Natural Language Semantics*, 3(2), 123–176.
Eid, M. (1993). Negation and predicate heads. In M. Eid, & G. Iverson (Eds.), *Principles and prediction: The analysis of natural language* (pp. 135–151). Amsterdam: John Benjamins.
Gelderen, E. van. (2008). Negative cycles. *Linguistic Typology*, 12, 195–243.
Gelderen, E. van. (2011). *The linguistic cycle: Language change and the language faculty*. Oxford: Oxford University Press.
Giannakidou, Anastasia. 1998. *Polarity sensitivity as (Non)veridical dependency*. Amsterdam: John Benjamins.
Haegeman, L., & Lohndal, T. (2010). Negative concord and (multiple) Agree: A case study of West Flemish. *Linguistic Inquiry*, 41, 181–211.
Hiraiwa, K. (2001). Multiple Agree and the defective intervention constraint. In O. Matushansky, & E. Gurzoni (Eds.), *The Proceedings of the MIT-Harvard Joint Conference: MIT Working Papers in Linguistics*, 40, 67–80.
Hoyt, F. (2007). An Arabic Wackernagel clitic? The morphosyntax of negation in Palestinian Arabic. In M. Mughazy (Ed.) *Perspectives on Arabic Linguistics Vol. XX* (pp. 105–134). Amsterdam: John Benjamins.
Jespersen, O. (1917). Negation in English and other languages. Reprinted in *Selected Writings of Otto Jespersen* (1962), (pp. 3–151). George Allen and Unwin Ldt.
Khalafallah, A. (1969). *A descriptive grammar of Sa'idi Egyptian Colloquial Arabic*. The Hague: Mouton.
Kuno, M. (2007). *Focusing on negative concord and negative polarity: Variations and relations*. Ph.D. dissertation, Harvard University.
Ladusaw, William. 1979. *Polarity sensitivity as inherent scope relations*. Ph.D. dissertation, University of Texas at Austin.
Laka, I. (1990). *Negation in syntax: On the nature of functional categories and projections*. Ph.D. dissertation, MIT.
Lucas, C. (2010). Negative -š in Palestinian (and Cairene) Arabic: Present and possible past. *Brill's Annual of Afroasiatic Languages and Linguistics*, 2, 165–201.
Lucas, C. (2013). Negation in the history of Arabic and Afro-Asiatic. In D. Willis, C. Lucas, & A. Breitbarth (Eds.), *The History of Negation in the Languages of Europe and the Mediterranean: Volume 1: Case Studies* (pp. 399–452). Oxford: Oxford University Press.
McCarthy, J. (1979). *Formal problems in semitic phonology and morphology*. PhD dissertation, MIT.
Moritz, L., & Valois, D. (1994). Pied-piping and specifier-head agreement. *Linguistic Inquiry*, 25, 667–707.
Omar, M. (1967). *The acquisition of Egyptian Arabic as a native language*. Washington, DC: Georgetown University Press.
Ouali, H., & Soltan, U. (2014). On negative concord in Egyptian and Moroccan Arabic. In S. Farwaneh, & H. Ouali (Eds.), *Perspectives on Arabic Linguistics Vol. XXIV–XXV* (pp. 159–180). Amsterdam: John Benjamins.

Ouhalla, J. (1990). Sentential negation, relativized minimality and the aspectual status of auxiliaries. *The Linguistic Review*, 7, 183–231.

Ouhalla, J. (1991). *Functional categories and parametric variation*. London: Routledge.

Ouhalla, J. (2002). The structure and logical form of negative sentences in Arabic. In J. Ouhalla, & U. Shlonsky (Eds.), *Themes in Arabic and Hebrew syntax* (pp. 299–320). Dordrecht: Kluwer Academic Publishers.

Palva. H. (1972). *Studies in the Arabic dialect of the semi-nomadic l-aarma tribe*. Göteborg: Acta Universitatis Gothorburgensis.

Penka, D. (2011). *Negative indefinites*. Oxford: Oxford University Press.

Pollock, J. (1989). Verb movement, universal grammar, and the structure of IP. *Linguistic Inquiry*, 20, 365–424.

Roberts, I. (2010). *Agreement and head movement: Clitics, incorporation, and defective goals*. Cambridge, Mass.: MIT Press.

Shlonsky, U. (1997). *Clause structure and word order in Hebrew and Arabic: An essay in comparative Semitic syntax*. Oxford: Oxford University Press.

Soltan, U. (2007). *On formal feature licensing in minimalism: Aspects of Standard Arabic morphosyntax*. PhD dissertation, University of Maryland, College Park.

Soltan, U. (2011). On issues of Arabic syntax: An essay in syntactic argumentation. *Brill's Annual of Afroasiatic Languages and Linguistics*, 3, 236–280.

Soltan, U. (2012). Morphosyntactic effects of NPI-licensing in Cairene Egyptian Arabic: The puzzle of *-š* disappearance resolved. In C. Jaehoon, E. A. Hogue, J. Punske, D. Tat, J. Schertz, and A. Trueman (Eds.), *Proceedings of the 29th West Coast Conference on Formal Linguistics* (pp. 241–249). Somerville, MA: Cascadilla Proceedings Project. http://www.lingref.com/cpp/wccfl/29/paper2708.pdf

Soltan, U. (2014). On the Distribution and licensing of polarity-sensitive items in Egyptian Arabic: The Cases of *ʔayy* and *walaa*. In S. Farwaneh, & H. Ouali (Eds.) *Perspectives on Arabic Linguistics Vol. XXIV–XXV* (pp. 181–206). Amsterdam: John Benjamins.

Travis, L. (1984). *Parameters and effects of word order variation*. Ph.D. dissertation, MIT.

Watanabe, A. (2004). The genesis of negative concord: Syntax and morphology of negative doubling. *Linguistic Inquiry*, 35, 559–612.

Zanuttini, R. (1997). *Negation and clausal structure: A comparative study of Romance languages*. Oxford: Oxford University Press.

Zeijlstra, H. (2004). *Sentential negation and negative concord*. Ph.D. Dissertation, University of Amsterdam.

Zeijlstra, H. (2008). Negative concord is syntactic agreement. Ms., University of Amsterdam.

Multiple agreement in Arabic

Hamid Ouali*
University of Wisconsin-Milwaukee

This article analyzes multiple subject verb agreement in complex tense clauses in light of the Feature Inheritance (FI) approach (Chomsky 2008, 2013). After establishing that these complex tense clauses are bi-clausal with two TP projections and one CP, I argue that they present a challenge to FI according to which C is the locus of φ-features. I propose an analysis where I maintain that T is lexically specified for φ-features and show how this view can account for all the multiple agreement facts.

Keywords: Agreement, Multiple Agreement, Agree, Complex Tense, Feature Inheritance

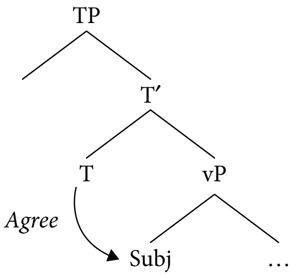
1. Introduction

Agreement has been a central concept in the syntactic theory of human language. Within Government and Binding (GB) framework, agreement was a result of the structural relation of Government; within Principles & Parameters and early Minimalism it was a result of Spec-Head relations; and within recent Minimalism, agreement is a result of an Agree operation.

Under the Agree analysis, the head bearing agreement features establishes an Agree relation with an argument. Subject-verb agreement in English, for example, results from T, which bears φ-features, establishing an Agree relation with the subject as schematized below:

* I would like to thank Karen Froud and Reem Khamis-Dakwar for all their help and for their patience during the editorial process of this volume. I also thank Abbas Benmamoun, Usama Soltan, and the audience at the Arabic Linguistics Symposium at Columbia University for very useful comments and suggestions. Needless to say that I bear sole responsibility for any errors.

(1)



Chomsky (2007, 2008, and 2013) has proposed, for conceptual and empirical reasons, that T is not the locus of φ-features but inherits these features from C. Richards (2007) provides further conceptual arguments for Feature Inheritance under the Phase theory (see also Miyagawa, 2010 and Ouali, 2008 and 2011). The Feature Inheritance approach seems to account for simple finite clauses and finite embedded clauses in English, which presumably have a single C and a single T. However, there are contexts in Arabic which seem to present a puzzle for this approach. These are cases of complex tense clauses that involve a copula and a main verb, both of which inflect for subject agreement as shown in (2):[1]

(2) *Kan-**a*** *l-walad-u* *j-aktub-u* *r-risalat-a*
BE.PERF-**3s** the-boy-NOM **3s**-write.IMPERF-IND the-letter-ACC
'the boy was writing the letter'

Complex tense clauses, such as (1) exhibit subject-verb agreement on both the copula and the main verb but seem to involve structurally two T heads and one C. The goal of this paper is to examine the clause structure in complex tense constructions and its implication on the C-T feature inheritance. I will argue that the structure of clauses denoting complex tense involves two T heads and one C, and that agreement in Arabic does not involve feature inheritance from C to T as was proposed for English (Chomsky, 2008, 2013). I will argue that T is inherently specified for φ-features and demonstrate that agreement in these clauses is result of a multi-Agree relation between the matrix T, the lower T, and the DP subject. This paper is organized as follows: section two discusses the Feature Inheritance approach and the challenges it faces regarding the Arabic facts, section three presents the properties of complex tense clauses, section four details an analysis of multiple agreement found in complex tense clauses in Arabic , and section five concludes.

1. I use complex tense clauses to refer sentences with verb complex forms that involve the copula *kana* and a main verb in Arabic. The copula can be in either the perfective form or imperfective form and can combine with a main verb in either perfective form or imperfective. These combinations can yield different tenses that span from past progressive as in (2) to compound tenses such as future in the past.

2. Feature Inheritance and multiple agreement in Arabic

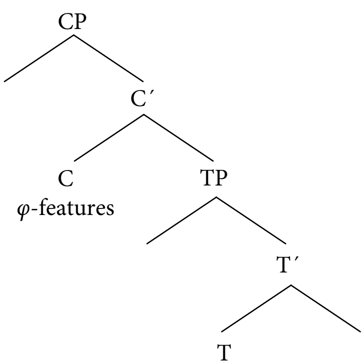
Chomsky (2007, 2008, and 2013) proposes that Agreement features are associated with C, the phase head, and that T inherits these features in the course of the derivation prior to establishing an Agree relation with the subject (see also Carstens, 2003 and Richards, 2007 among many others). Chomsky (2013) writes:

> ... there is good reason to suppose that the φ-features of T are in fact inherited from C and ... the tense feature as well. The system is simplified if features of an LI cannot move independently of the feature bundle to which they belong. That would entail that all the features of C should be inherited by T, including not only tense (as is overt) but also Q. (p.15)

Chomsky (2008) also writes:

> It seems problematic for T to fail to define a phase boundary along with C, since on the surface it seems to be T, not C, that is the locus of the φ-features that are involved in the Nominative-agreement system, and raising of the external argument subject or unaccusative/passive object to SPEC-T. There is, however, antecedent reason to suspect otherwise.... The antecedent reason is that for T, φ-features and Tense appear to be derivative from C. In the lexicon, T lacks these features. T manifests them if and only if it is selected by C... if not, it is a raising (or ECM) infinitival, lacking φ-features and tense. So it makes sense to assume that Agree- and Tense-features are inherited from C, the phase head. (p. 143)

(3)



There are conceptual reasons that necessitate Chomsky's C-T feature inheritance proposal and empirical facts that lend support to it. Regarding the conceptual reasons, Richards (2007) argues that grammatical features that are responsible for computations such as movement show up solely on the phase heads C/v. It is therefore conceptually reasonable to assume that C, being the upper phase head, is the locus of φ-features.

Empirical evidence for feature inheritance comes from ECM constructions such as (3) in English (Chomsky, 2008), where nonfinite embedded clauses can never begin with a complementizer because complementizers are associated with finiteness:

(4) Bill wants (*that) her to win

Also, in some West Germanic languages and dialects such as Limburgian (below) and West Flemish, C inflects for agreement (Carstens, 2003, Haegeman, 1993, and recently Haegeman & von Koppen, 2012).

(5) *Ich denk de- s doow Marie ontmoet- s.* (Limburgian)
I think that-2SG you.SG Marie meet-2SG
'I think that you will meet Marie' (Haegeman & von Koppen, 2012, p. 441)

In English, lack of subject-verb agreement indicates the absence of C in the derivation, as is the case in nonfinite clauses. Conversely, subject-verb agreement indicates the presence of C in the derivation hence the C-T φ-features yielding agreement, as is the case in finite clauses. In Arabic, however, this is not always true. That is, if we examine the sentence in (6), a case of multiple agreement from MA, one could argue that the higher T in the structure of the clause is selected by C and therefore inherits the φ-features from it, prior to establishing an Agree relation with the subject:

Agree

(6) $[_{CP}$ C T*ɣa* ***j-kun-u*** *l-wlad* T *ka* ***j-akl-u***
FUT 3-BE.IMPERF-p the-boys T ASP 3-EAT.IMPERF-p
'the boys will be eating'

This would explain the full subject-verb agreement on the copula *kana*. The question that we would need to answer is: How do we get agreement on the main verb? This agreement is presumably a result of an Agree relation between the lower T and the subject.[2] However, since the lower T is not selected by C, it cannot inherit the φ-features. In fact, T cannot be selected by C, since overt Complementizers are prohibited in that context as shown by the ungrammaticality of (7):

2. The subject could either be the lexical argument *l-wlad* "the boys", assuming that it is first merged as an argument of the main verb in lower vP:

a. ɣa **j**-kun-**u** l-wlad ka [vP ~~l-wlad~~ **j**-akl-**u**]

or be a null pronoun if the lexical argument is merged higher in the structure:

b. ɣa **j**-kun-**u** l-wlad ka [vP pro **j**-akl-**u**]

In Section 3, I will argue for the latter.

(7) **ɣa* ***j-kun-u*** *l-wlad bəlli ka* ***j-aklu-u***
FUT 3-BE.IMPERF-p the-boys Comp ASP 3-EAT.IMPERF-p

This leads us to conclude that C-T φ-feature inheritance must be parameterized. In English-type languages, C is inherently specified for φ-features, and T acquires these features through inheritance in the course of the derivation. In Arabic-type languages, however, T is lexically specified for φ-features. How can we account for multiple agreement in complex tense clauses? I will argue that these clauses are bi-clausal and their structures have two Tense projections. Both T heads establish an Agree relation with the closest DP. Before detailing this analysis, we will.

3. Properties of complex tense clauses

In both SA and the Arabic dialects, the simple past tense is expressed by using the perfective form of the verb, the present tense (continuous or habitual) by using the imperfective form, and the future tense by using the imperfective form combined with a future marker.

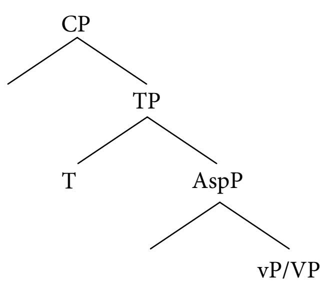
(8) *katab-a ʕalijj-un r-risalat-a* (SA)
write.PERF-3s ali-NOM the-letter-ACC
'Ali wrote the letter'

(9) *ja-ktub-u ʕalijj-un r-risalat-a* (SA)
3M-write.IMPERF-IND ali-NOM the-letter-ACC
'Ali is writing the letter'

(10) *sa ja-ktub-u ʕalijj-un r-risalat-a* (SA)
FUT 3M-write.IMPERF-IND ali-NOM the-letter-ACC
'Ali will write the letter'

There is almost a consensus that the clause structure of the sentences above involves one Tense projection (Benmamoun, 2000 and Soltan, 2007 among many others).[3]

(11)



3. See Fassi Fehri (2012) for a different analysis.

Complex tenses, such as the past perfect for example, are expressed by combining the copula *kana* and a main verb as illustrated by the SA example in (12) and the MA example in (13):

(12) *kan-a katab-a r-risalat-a lamma daxal-tu* (SA)
BE.PERF-**3s** write.PERF-3s the-letter-ACC when enter.PERF-**1s**
'He had written the letter when I entered' (Fassi Fehri, 2004, p. 238)

(13) *kan-u kla-w* (MA)
BE.PERF-3PL EAT.PERF-3PL
'They had eaten'

As we can see in (12) and (13) above, the past perfect is expressed by using the perfective form of *kana* combined with a main perfective verb. The past progressive and habitual past are expressed by combining the copula *kana* in the perfective form and a main verb which must be in the imperfective form, as illustrated by (14) from SA, (15) from JA, and (16) from MA below:_

(14) *ka:n-a l-walad-u j-aktub-u r-risa:lat-a* (SA)
BE.PERF-3s the-boy-NOM 3s-write.IMPERF-IND the-letter-ACC
'the boy was writing the letter' *Continuous*

(15) *ka:n l-walad b-j-əktub r-risa:lə* (JA)
BE.PERF-3s the-boy-boy ASP-3-write.IMPERF the-letter
'the boy was writing the letter'

(16) *ka:n l-wəld ka-j-ktəb r-risa:la* (MA)
BE.PERF-3s the-boy-boy ASP-3sm-write.IMPERF the-letter
'the boy was writing the letter'

Future in the past is denoted by the perfective form of *ka:na* and a main verb in the imperfective form combined with the future marker, as illustrated by the MA example in (17):

(17) *ka:n-u ɣa j-akl-u*
BE.PERF-3p FUT 3-EAT.IMPERF-p
'They were going to eat'

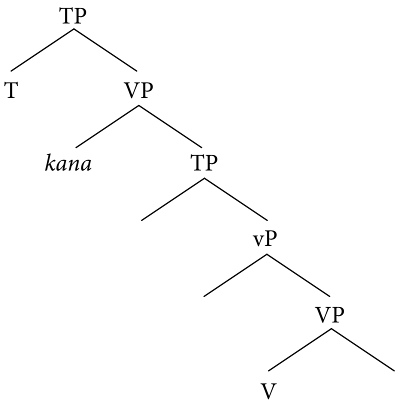
These facts have been described and discussed extensively in the literature. Haak (2006, pp. 219–220) provides a full paradigm of this verb complex. What is lacking is an analysis of agreement in these clauses in light of recent developments in the theory of agreement in Minimalism.

As we can see in all the examples listed above and the example in (18) below, both the auxiliary and the main verb are marked for tense and inflected for agreement. For example the auxiliary *kana* is preceded by the future maker *sa* denoting future, and the main verb is in the perfective form denoting past.

(18) *sa-ja-ku:n-u* *ʕali katab-a* *r-risa:lat-a* *ɣadan* (SA)
FUT-3SM-BE. IMPERF-IND ali write.PERF.3s the-letter-ACC tomorrow
'Ali will have already written the letter tomorrow'

Agreement with the subject is also marked on both the auxiliary and the main verb, a fact that I refer to as multiple agreement. These facts are attested in all Arabic dialects and raise a question about the structure of complex tense clauses. Ouali and Fortin (2007) have argued that these complex tense sentences are bi-clausal with two Tense projections in the structure as shown in (19), a position I take here as well (see also Soltan, 2007, 2011, and Fassi Fehri, 2004, 2012).

(19)



Having two T projections explains the complex tense paradigm in Arabic where the auxiliary can be marked for past, present, or future and combine with a main verb marked for any of these tenses. The multiple agreement facts also follow from the assumption that both the lower T and the higher T enter into an agreement relation with the subject, as will be detailed in the next section.

These facts, however, raise a challenge to the Feature Inheritance approach (Chomsky, 2005, 2007, 2008, 2013), according to which the subject-verb agreement (in finite clauses) is a result of T inheriting φ-features from C and then entering into an Agree relation with the subject. C is specified for φ-features in the lexicon but T is not. The next section discusses this topic further.

4. Multiple agreement and the subject positions

Considering the subject position in complex tense clauses, one could argue that both the higher T and the lower T establish an Agree relation with the same thematic subject. A Multiple Agree relation has been proposed in the literature where a probe F can establish multiple (Simultaneous) Agree relations with more than one goal as schematized in (20) (see e.g. Hiraiwa, 2001, 2005; Henderson, 2007).

(20) $[_{CP}\ T_{\varphi}\ \ VP\ \ [_{TP}\ T_{\varphi}\ \ [_{vP}\ \ DP_{\varphi}\ V\ ]]]$

For complex tense clauses with *kana* such as (21), one could argue that matrix T Agrees with both the lower T and the DP subject as shown in (22):

(21) *ɣa j-kun-**u** kla-**w** l-wlad*
FUT 3-BE.IMP-PL EAT.IMP-3PL the-boys
'the boys will have eaten'

(22) [$_{CP}$ C T *ɣa* [$_{VP}$ ***j-kun-u*** [$_{TP}$ *T kla-**w*** [$_{vP}$ *l-wlad* ~~*kla-w*~~]]]]
Agree Agree

This would account for examples such as (21) where the DP subject is in a post-verbal position. However, the picture is not as simple as it seems. The Subject can occupy different positions in such constructions as illustrated in (23), (24), and (25):

(23) *ɣa j-kun-**u** ka j-akl-**u** l-wlad*
FUT 3-BE.IMPERF-p ASP 3-EAT.IMPERF-p the-boys
'the boys will be eating'

(24) *ɣa j-kun-**u** l-wlad ka j-akl-**u***
FUT 3-BE.IMPERF-p the-boys ASP 3-EAT.IMPERF-PL
'the boys will be eating'

(25) *l-wlad ɣa j-kun-**u** ka j-akl-**u***
the-boys FUT 3-BE.IMPERF-p ASP 3-EAT.IMPERF-p
'the boys will be eating'

In SA, it is known that the position of the subject affects subject-verb agreement. For example, in (26) below where the subject precedes the main verb and follows the auxiliary, the main verb must be marked for full agreement and the auxiliary must be marked for person and gender agreement only; otherwise the sentence would be ungrammatical, as in (27).

(26) *ka:n-a l-ʔawla:d-u j-aktub-u:-na r-rasa: ʔil-a*
BE.PERF-3s the-boys-NOM 3-write.IMPERF-p-IND the-letter-ACC
'the boys were writing/ used to write letters'

(27) **ka:n-u: l-ʔawla:d-u j-aktub-u:-na r-rasa:ʔil-a*
BE.PERF-3PS the-boys-NOM 3-write.IMPERF-p-IND the-letters-ACC
'the boys were writing/ used to write letters'

When the subject precedes both the auxiliary and the main verb, both verbs must be marked for full subject-verb agreement as shown in (28); otherwise the sentence would be ungrammatical as shown in (29).

(28) *l-ʔawla:d-u* *ka:n-**u:*** *j-aktub-u:-na* *r-rasa: ʔil-a*
the-boys-NOM BE.PERF-3PM 3-write.IMPERF-P-IND the-letter-ACC
'the boys used to write/were writing letters'

(29) **l-ʔawla:d-u* *ka:n-**a*** *j-aktub-u:-na* *r-rasa: ʔil-a*
the-boys-NOM BE.PERF-3SM **3**-write.IMPERF-P-IND the-letters-ACC
'the boys used to write/were writing letters'

The position of subjects in simple SVO vs. VSO sentences is highly debated in the Arabic syntax literature (see for example Fassi Fehri, 1982, 1988, 1993, 2012; Mohammad, 1990, 2000; Benmamoun, 2000; Aoun, Benmamoun & Chouieiri, 2010; Eid, 1991; Bahloul & Harbert, 1993; Aoun, Benmamoun & Sportiche, 1994; Shlonsky, 1997; and recently Soltan, 2007, 2011 – to cite just a few). The analyses of the subject position in Arabic could be summarized in two major approaches: the so-called Movement Analysis and the Topic/Left Dislocation Analysis. According to the movement analysis, SVO and VSO sentences have the same underlying structure. SVO sentences are derived my moving the subject out of *v*P to Spec, TP. The Subject enters in a Spec-Head agreement relationship with T, which then results in full subject-verb agreement. This is schematized in (30):

(30) $[_{TP}\ \mathrm{DP}_{Subj}\ \ [vP\ \ \cancel{\mathrm{DP}_{Subj}}\ \mathrm{V}\ \mathrm{DP}_{Obj}]]$ (full agreement)

The VSO order is derived by moving the verb to T, whereas the subject stays *in situ* as schematized in (31):

(31) $[_{TP}\ pro_{Exp}\ \ \mathrm{V}\ [v\mathrm{P}\ \ \mathrm{DP}_{Subj}\ \cancel{\mathrm{V}}\ \mathrm{DP}_{Obj}]]$ (partial agreement)

Partial agreement is a result of the Spec-Head relation between T and the null expletive *pro.*

According to the Topic/Left Dislocation Analysis, the derivations of SVO and VSO involve two different underlying structures. In SVO sentences, the thematic subject is a null *pro* base-generated in Spec-*v*P. The full DP on the other hand is merged as a Topic. This analysis is schematized in (32):

(33) $[_{TP}\ \mathrm{DP}_{Top}\ \ [v\mathrm{P}\ \ pro_{Subj}\ \ \mathrm{V}\ \mathrm{DP}_{Obj}]]$ (full agreement in SA)

The underlying structure of VSO involves an overt DP subject, which stays *in situ*. The verb moves to a higher T as schematized in (33):

(33) [$_{TP}$ $_{T}$ V [*v*P DP$_{Subj}$ ~~V~~ DP$_{Obj}$]] (partial agreement in SA)

For detailed reviews of the Movement analysis and the Left Dislocation/Topic analysis see Benmamoun (2000), Aoun, Benmamoun & Chouieiri (2010), and Soltan (2007) among others.

Soltan (2007, 2011) adopts the second analysis and argues that the subject-verb agreement in sentences such as (34) is a result of an Agree relation between T and the thematic subject *pro*. The DP *l-ʔawla:d-u* 'the boys' is not the thematic subject but a Topic merged directly in the Topic position, which he argues could be Spec,TP as illustrated in (35):

(34) *l-ʔawla:d-u j-aktub-u: r-rasa:ʔil-a*
the-boys-NOM 3-write.IMPERF-PM the-letters-ACC
'the boys are writing /write the letters'

(35) [$_{TP}$ *l-ʔawla:d-u* T [$_{vP}$ *pro* *j-aktub-u:* *r-rasa: ʔil-a*]]
Agree

According to Soltan (2007, 2011), there are two main advantages of such an analysis. First, it accounts for the topic interpretation of the preverbal DP. Second, it accounts for the case form of the preverbal DP when embedded under the complementizer *ʔinna,* for example, as in (36):

(36) *ʔinna l-ʔawla:d-a j-aktub-u: r-rasa:ʔil-a*
Comp the-boys-ACC 3-write.IMPERF-p the-letters-ACC
'the boys are writing /write letters'

Soltan's analysis makes the important prediction that if full agreement is associated with a null pronoun, then it would also be expected with a regular pronoun. That is exactly what we find in SA as shown in (37) and (38), where full subject-verb agreement is obligatory with overt personal pronouns:

(37) *katab-u: hum r-rasa:ʔil-a*
write.PERF-3p they the-letters-ACC
'they wrote the letters'

(38) **katab-a hum r-rasa:ʔil-a*
write.PERF-3SM they the-letters-ACC

In (39), as discussed previously, only partial agreement is obligatory with full DPs.

(40) *katab-a l-ʔawla:d-u r-rasa: ʔil-a*
write.PERF-3SM the-boys-NOM the-letters-ACC
'the boys wrote the letters'

Why is agreement with full DPs partial? There are no convincing answers to this question in the literature, but some stipulations have been put forward. For example, Soltan (2007, 2011) suggests that partial agreement could be a result of the properties of the T head. He suggests that the T head itself is impoverished. Benmamoun (2000) and Aoun, Benmamoun & Choueri (2010) suggest that it could be a post-syntactic process. In other words, the full agreement gets lost once the derivation reaches the Morpho-Phonology component. A couple of possibilities could also be added to these previous stipulations. One possibility is that DPs are opaque and their φ-features are not visible to higher probes. Another possibility is that DPs have a complex structure where each φ-feature heads its own projection. Agree establishes a relation between T and the outer layer of the DP, which could be Gender Phrase (GenP). All of these will remain speculative answers. Further research is needed to determine which of them is on the right track, but that is beyond the scope of this paper.

Returning now to the complex tense clauses, I argue, extending Soltan's (2007, 2011) analysis, that the underlying structure involves merging a null *pro* in Spec, *v*P as the thematic subject. The full DP is externally merged in Spec, TP. Given the bi-clausal structure of these sentences, the lower T establishes an Agree relation with the closest Goal, which is the thematic subject *pro*. The higher T establishes an Agree relation with the full DP. This is illustrated in (40):

(40) [T .~~*ka:na*~~ [$_{\text{VP}}$ ka:na [$_{\text{TP}}$ */-ʔawla:d-u* [T [$_{v\text{P}}$ *pro j-aktub-uu-na r-rasa: ʔil-a*]]]]]
Agree Agree
BE.PERF.3SM the-boys-NOM 3-write.IMPERF-3PM-IND the-letters-ACC
'the boys were writing/used to write the letters'

How can we extend this analysis to sentences where the subject is in a post-verbal position as in (41) or pre-copula position as in (42)?

(41) *ka:n-**a*** *j-aktub-u* *l-ʔawla:d-u* *r-rasa: ʔil-a*
BE.PERF-3PM 3-write.IMPERF-P-IND the-boys-NOM the-letter-ACC
'the boys used to write/were writing letters'

(42) *ʔal-ʔawlad-u* *Kan-**uu*** *j-aktub-uu-na* *r-rasa: ʔil-a*
the-boys-NOM BE.PERF-3PM 3-write.IMPERF-P-IND the-letter-ACC
'the boys used to write/were writing letters'

I argue that the derivations of examples (41) and (42) involve multiple-Agree with the subject as schematized in (43) and (44) respectively:

(43) [$_{TP}$T *ka:n-**a*** [$_{VP}$ ~~*ka:na*~~ [$_{TP}$ T *j-aktub-u* [$_{vP}$ *l-ʔawla:d-u* j-aktub-u *r-rasa:ʔil-a]*]]]

Agree

Agree

(44) [$_{TP}$ *ʔal-ʔawla:d-u* T ~~*Kan-u:*~~ [$_{VP}$ Kan-u: [$_{TP}$ T... ~~*j-aktubu:na*~~ [*pro* j-aktubu:na *r-rasa:ʔil-a*]]]]

Agree

Agree

In (43) both T heads establish an Agree relation with the lexical thematic subject *l-ʔawla:d-u*, and in (44) the two T heads establish an Agree relation with the thematic subject *pro*. This analysis inherits the same problem of why we get partial subject agreement with the overt lexical post-verbal subjects and full agreement with the null pro in the same position. I do not have any solution to this problem.

The analysis can be extended to MA (and presumably other dialects). Sentences such as (45) from MA, are derived in the same fashion as (40). This is shown in (46):

(45) ɣa j-kun-u l-wlad kla-w
FUT 3-BE.IMP-PL the-boys EAT.imp-3PL
'the boys will have eaten'

(46) [$_{TP}$ ɣa [$_{VP}$ j-kun-u [l-wlad T kla-w [$_{vP}$ *pro* kla-w]]]]

Agree Agree

To conclude this section, I will discuss briefly the properties of complementizers, mainly null complemetizers, in Arabic. I have argued that Feature Inheritance is not active in deriving Arabic clauses. T in Arabic is lexically specified for Tense and φ-features. The question then arises: How about complementizers in this language? If C in English is inherently specified for φ-features, what about C in Arabic? I argue that at least null complementizers are not specified for φ-features. Null complementizers do not enter in an Agree relation with any argument and therefore no feature interpretability issues arise. In a sentence such as (39) above [repeated as (47) below], null C does not bear any φ-features, but the two T heads do:

(47) [C [T$_{φ}$ [$_{VP}$ *ka:na* [$_{TP}$ *ʔa-wlawla:d-u* [T$_{φ}$ *j-aktub-u:* [$_{vP}$ *pro r-rasa:ʔil-a*]]]]]]

Agree Agree

BE.PERF.3SM the-boys-NOM 3-write.IMPERF-3PM the-letters-ACC
'the boys were writing/used to write the letters'

5. Conclusion

In this paper I argued that complex tense clauses are structurally bi-clausal involving two TP projections. I also argued that Arabic agreement facts in these clauses seem to indicate that C-T Feature inheritance must be parameterized. In languages such as English, C is inherently specified for φ-features, and T inherits these features from C in the course of the derivation prior to establishing an Agree relation with the DP subject. In languages such as Arabic, I argued that null C is not specified φ-features. T is inherently specified for φ-features and therefore does not acquire these features by inheritance. This explains why agreement is manifested in contexts where there could be no complementizers. I adopted Soltan's (2007, 2011) analysis of simple SVO and VSO sentences and argued that the agreement facts in complex tense clauses are a result of two Agree relations, the first between the lower T and the thematic subject *pro* and the second between the higher T and the full DP. In cases where the subject is in post-verbal position, both the higher T and lower T establish a simultaneous Agree relation with the subject.

References

Aoun, J., Benmamoun, E., & Sportiche, D. (1994). Agreement, conjunction, and word order in some varieties of Arabic. *Linguistic Inquiry*, 25, 195–220.

Aoun, J., Benmamoun E., & Choueiri, L. (2010). *The syntax of Arabic*. Cambridge, MA: Cambridge University Press.

Bahloul, M., & Harbert, W. (1993). Agreement asymmetries in Arabic. In J. Mead (Ed.), *The Eleventh West Coast Conference on formal linguistics* (pp. 15–31). Stanford, CT: Center for the Study of Language and Information.

Benmamoun, E. (2000). *The feature structure of functional categories: A comparative study of Arabic dialects*. NY: Oxford University Press.

Carstens, V. (2003). Rethinking complementizer agreement: Agree with a case-checked goal. *Linguistic Inquiry*, 34, 393–412.

Chomsky, N. (2005) Three factors in language design. *Linguistic Inquiry*, 36(1), 1–22.

Chomsky, N. (2007). Approaching UG from below. In U. Sauerland, & H.-M. Gärtner (Eds.), *Interfaces + recursion = language? Chomsky's minimalism and the view from syntax-semantics* (pp. 1–29). Berlin: Mouton de Gruyter.

Chomsky, N. (2008). On phases. In R. Freidin, C. Otero, & M. L. Zubizarreta (Eds.), *Foundational issues in linguistic theory* (pp. 133–166). Cambridge, MA: MIT Press.

Chomsky, N. (2013). Problems of projection. *Lingua*, 18, 1–35.

Eid, M. (1991). Verbless sentences in Arabic and Hebrew. In B. Comrie, & M. Eid (Eds.), *Perspectives on Arabic linguistics, Vol. III: Papers from the Third Annual Symposium on Arabic linguistics, Salt Lake City, Utah 1989* (pp. 31–61). Amsterdam: John Benjamins.

Fassi Fehri, A. (1982). *Linguistique Arabe: forme et interprétation*. Rabat, Morocco: Publication de la Faculté des lettres et Sciences Humaines.

Fassi Fehri, A. (1988). Agreement, binding, and coherence. In M. Barlow, & C. A. Ferguson (Eds.), *Agreement in natural language: Approaches, theories, descriptions* (pp.107–158). Stanford: Center for the Study of Language and Information.

Fassi Fehri, A. (1993). *Issues in the structure of Arabic clauses and words*. Dordrecht: Kluwer.

Fassi Fehri, A. (2004). Temporal/aspectual interaction and variation across Arabic heights. In J. Gueron, & J. Lecame (Eds.), *The syntax of time* (pp. 235–257). Cambridge MA: The MIT Press.

Fassi Fehri, A. (2012). *Key features and parameters in Arabic grammar*. Amsterdam & Philadelphia: John Benjamins. DOI: 10.1075/la.182

Haak, M. (2006). Auxiliaries. In V. Kees, M. Eid, A. Elgibali, M. Woidich, & A. Zaborski (Eds.), *Encyclopedia of Arabic languages and linguistics* (Vols. I-V, pp. 216–221). Leiden: Brill.

Haegeman, L. (1993). Some speculations on argument shift, clitics and crossing in West Flemish. *Linguistische Berichte. Sonderheft*, 5(2p.), 131–160.

Haegeman, L., & von Koppen, M. (2012). Complementizer agreement and the relation between C^0 and T^0. *Linguistic Inquiry*, 43, 441–454.

Henderson, B. (2007). Multiple Agreement and Inversion in Bantu. *Syntax* 9 (3), 275–289.

Hiraiwa, K. (2001). Multiple agree and the defect intervention constraint. In O. Matushansky, & E. Gurzoni (Eds.), *The proceedings of the HUMIT 2000, MIT Working Papers in linguistics*, 40, 67–80. Cambridge, MA: MITWPL.

Hiraiwa, K. (2005). *Dimensions of symmetry in syntax: Agreement and clausal architecture*. Ph.D. dissertation, MIT.

Miyagawa, S. (2010). *Why agree? why move? Unifying agreement-based and discourse-configurational languages*. Cambridge, MA: MIT Press.

Mohammad, M. (1990). The problem of subject-verb agreement in Arabic: Towards a solution. In M. Eid (Ed.), *Perspectives on Arabic linguistics, Vol. I: Papers from the First Annual Symposium on Arabic Linguistics, Salt Lake City, Utah 1987* (pp. 95–125). Amsterdam: John Benjamins.

Mohammad, M. (2000). *Word order, agreement and pronominalization in standard and Palestinian Arabic*. Amsterdam: John Benjamins. DOI: 10.1075/cilt.181

Ouali, H. (2008). On C-to-T -feature transfer: The nature of agreement and anti- agreement in Berber. In R. D' Alessandro, G. H. Hrafnbjargarson, & S. Fischer (Eds.), *Agreement restrictions* (pp. 159–180). Berlin: Mouton de Gruyter.

Ouali, H. (2011). *Agreement, pronominal clitics and negation in Tamazight Berber: A unified analysis*. NY: Continuum.

Ouali, H., & Fortin, C. (2007). The syntax of complex tense in Moroccan Arabic. In E. Benmamoun (Ed.), *Perspectives on Arabic linguistics XIV* (pp. 175–189). Amsterdam: John Benjamins.

Richards, M. D. (2007). On feature inheritance: An argument from the phase impenetrability condition. *Linguistic Inquiry*, 38, 563–572.

Shlonsky, U. (1997). *Clause structure and word order in Hebrew and Arabic: An essay in comparative semitic syntax*. Oxford: Oxford University Press.

Soltan, U. (2007). *On formal feature licensing in minimalism: Aspects of standard Arabic morphosyntax*. Ph.D. dissertation, University of Maryland, College Park, MD.

Soltan, U. (2011). On issues of Arabic syntax: An essay in syntactic argumentation. *Brill's Annual of Afroasiatic Language and Linguistics*, 3, 236–280.

Cyclic AGREE derives restrictions on cliticization in classical Arabic*

Martin Walkow
Massachusetts Institute of Technology

Person based restrictions on clitic combinations serve as testing grounds for theories of syntactic locality and the means of avoiding them as windows into last resort mechanisms. Clitic restrictions in the verbal domain in Classical Arabic can be derived by Cyclic AGREE, rather than (defective) intervention. This offers a unified, syntactic analysis of the ultrastrong Person Case Constraint and restrictions on combinations of third person clitics. The alternative structures used when cliticization is blocked are the PF realization of independently established syntactic relations, not a last resort mechanism. Several properties of person restrictions are shown to follow from the causative structure of double object verbs in Classical Arabic.

Keywords: Person Case Constraint, restrictions on third person, locality of AGREE, cliticization, causatives.

1. Person based restrictions and the workings of AGREE

Person based restrictions on combinations of clitics or agreement markers have assumed great prominence as testing grounds for theories of the operation agree, the only non-structure building operation assumed in the framework of Chomsky (2000, 2001). Anagnostopoulou (2003, 2005) and Béjar (2003) show that such restrictions arise when two goals are in the domain of one probe and syntactic locality restricts the probe's ability to AGREE with both of them. Two major proposals have emerged of how syntactic locality leads to person based restrictions. On the one hand there is intervention based locality. In a structure [Probe > x > y], where > is c-command, x may interfere with AGREE between the probe and y, if x has the right features. On the other hand, there is Cyclic AGREE (Řezáč, 2003; Béjar, 2003; Béjar and Řezáč 2009), where probes AGREE with one goal when they are first merged and gain access to the second one through the expansion of the phrase marker via Merge. As opposed to intervention, where the features of the higher goal bleed AGREE with the lower one, here, the lower goal can bleed AGREE with the higher one. Variation in person restrictions has served as a testing ground for the two proposals. Languages differ with respect to which combinations of person on the two arguments they ban. Intervention based proposals have mostly been applied to variation of the Person Case Constraint (PCC), a restriction on combinations of two internal argument clitics or agreement markers (Bonet 1991, 1994). Anagnostopoulou (2005) and Nevins (2007) locate this variation in the operation AGREE itself. Nevins (2007) in particular shows that multiple Multiple agree, an operation that allows AGREE between one probe and multiple goals in one step, can be parameterized to account for four different kinds of PCC. In this context, he presents an analysis of the clitic restrictions in Classical Arabic (CA) (e.g. Fassi Fehri, 1988), which he calls the *ultrastrong* PCC. Cyclic AGREE analyses of variation have focused on restrictions between subjects and objects (Béjar, 2003; Béjar and Řezáč, 2009). Variation is located more conservatively in the properties of the functional lexicon, in particular the feature specification of probes. Walkow (2011) shows, though, that the logic of cyclic expansion can also derive the so called *strong* PCC, which bans local person direct objects in the presence of all indirect objects. This paper contributes to this debate in three respects. The first concerns the scope of what syntactic analyses of person restrictions can and should explain. I show that CA has restrictions on combinations of third person pronouns in addition to the PCC, leading to the overall picture of person restrictions in (1).

(1) The Classical Arabic Person Constraint: Two pronouns cliticize if
   a. The syntactically higher one is more local than the lower one,
   b. where: $1 \underset{\text{local}}{>} 2 \underset{\text{local}}{>} 3$

Restrictions on combinations of third person pronouns are typically argued to be morphological, rather than syntactic, and are sidelined in discussions of the PCC (Anagnostopoulou, 2003; Nevins, 2007; Řezáč, 2008). Continuing the argument in Walkow (2011, 2012a,b) on Catalan and Spanish, the parallelism between the PCC and restrictions on third person are shown to follow naturally, if third person is represented in the syntax and participates in the same interactions of AGREE as other person categories. Secondly, I present a Cyclic AGREE analysis of (1) demonstrating the wider applicability of Cyclic AGREE to the PCC and its ability to deliver an account of variation based in the properties of the functional lexicon rather than the parameterization of syntactic operations. Thirdly, I address how the syntactic structure that gives rise to (1) relates to the strategies used when cliticization of both arguments is blocked. Previous work on alternative strategies for realizing banned clitic combinations has focused on how the observed repair strategies avoid the PCC (Anagnostopoulou, 2003; Řezáč, 2007, 2008, 2011), and how the last-resort character of these strategies arises (Řezáč, 2007, 2011). I argue that the alternate realization of banned clitic combinations in CA neither avoids the PCC nor is the result of a last resort mechanism. Rather, it is the derivational byproduct of AGREE relations that happen independently, but do not lead to cliticization. Core aspects of the analysis are motivated by the fact that (1) arises in causative structures.

The paper is organized as follows. Section 2 introduces the data about cliticization in CA and motivates (1). Section 3 shows that a central class of predicates that allow clitic combinations are causatives. Section 4 presents the analysis of (1) for causatives, and Section 5 shows how this explanation may extend to non-causative contexts that allow clitic combinations. Section 6 shows how the analysis here compares to the Multiple AGREE analysis.

2. Restrictions on cliticization in classical arabic

The clitic restrictions in CA are an instance of a family of restrictions on marked person specifications on arguments in low syntactic positions known as the Person Case Constraint or PCC. After a brief introduction of the system of clitics and free pronouns in CA, Section 2.1 will present the restrictions on combinations containing local person pronouns as they have been discussed in previous analyses (Fassi Fehri, 1988; Nevins 2007).

Section 2.2 shows that the full picture of person restrictions in CA includes both the *ultrastrong* PCC and the ban on combinations of third person pronouns. Section 2.3 formulates two questions about the alternative realization of banned clitic combinations: Why is it always the direct object that is realized as a free pronoun, and why is it realized as a free pronoun, rather than a PP? The data are taken from the grammar by Amr ibn Uthmān Sibawayh written in the late 8th century, the seminal book on Arabic grammar. It will be cited as 'Sibawayh (1881)' with page numbers from Derenbourg's (1881) critical edition, and occasionally accompanied by the corresponding page numbers in Jahn's (1900) translation to allow easier access.

Table 1. Inventory of pronominal clitics in Classical Arabic (Wright, 1874, vol. 1: 100).

		1:	**2:**	**3:**
SG:	M:	*-iː*$_{gen}$/*-niː*$_{acc}$	*–ka*	*–hu*
	F:		*–ki*	–hæː
Dual:		n/a	*–kumaː*	*–humaː*
PL:	M:	*–naː*	*–kum*	*–hum*
	F:		*–kunːa*	*–humːa*

Object pronouns in CA come in two morphological forms: clitic pronouns, which attach to a host, and free pronouns, which do not. Clitic pronouns morphologically distinguish three persons, two genders and three numbers, Table 1. Case distinctions are only present in the first person singular. The genitive is -[iː]. The accusative -[niː] has an additional /n/ (*nun al wiqaaya* in the traditional grammar, e.g. Wright, 1874, vol. 1, p. 101).

Free pronouns on the other hand consist of the genitive forms of the clitics plus the accusative marker [ʔɪjːaː]-, as illustrated in example (2).

The choice between clitic and free pronouns is not free. Similar to what Cardinaletti and Starke (1999) observe for a variety of Indo-European languages, clitics must be used unless their use is blocked (e.g. Sibawayh, 1881, §206ff; Howell, 1880, pt *i*, vol 2, p. 253; de Sacy, 1905, p. 377). The use of clitics is blocked, for example, when a pronominal object is fronted to the left of the verb as in (2), when it is separated from the verb as in the second position in a conjunction (3) (pronouns in initial conjuncts can cliticize, e.g. Reckendorf 1895, p. 395), or when a clitic combination would violate (1). In the examples here and below, the relevant pronouns will be underlined.

(2) ʔɪjːaː-ka naʕbudu
ACC-2SG 1PL-worship
'Thee we worship'

(3) fa allah-u
PRT Allah-NOM
ja-rʕaː [abaː ħarb-in wa ʔɪjːaː-naː]
3SG.M-protect father.ACC war-GEN and ACC-1PL
'Thus Allah protects a warrior and us'
(Sibawayh, 1881, p. 332, §207, translation after Jahn, 1900, vol. 2, p. 95)

Against this backdrop, (1) is a person- based alternation between clitics and free pronouns.

The Person case constraint

Clitic combinations arise with predicates that take two non-nominative arguments. A core class of contexts are verbs like [ʔaʕtˤaː], 'give,' in (4), that take two accusative arguments. Section 5 will discuss some other contexts.

(4) ʔaʕtˤaɪ-tu zaɪd-an dɪrham-an
gave-1SG Zaid-ACC.IND Dirham-ACC.IN
'I gave Zaid a Dirham' (Howell, 1880, pt *i*, vol 2, p. 536)

For now I will refer to arguments like [zaɪd-an] as the *first accusative* and those like [dɪrham-an] as the *second accusative*. Section 3 will show that they are the causee and direct object of a causative predicate. When both accusative arguments are pronouns, Sibawayh (1881, p. 335) reports, the possibility of clitic combinations depends on the person specification of the pronouns: a less local person pronoun must not precede a more local one, where first person is more local than second person, and second person is more local than third:

(5) Sibawayh's Generalization:
a. $^{*}\mathrm{CL}_1 >_{\text{local}} \mathrm{CL}_2$,
b. where: $1 >_{\text{local}} 2 >_{\text{local}} 3$

The effect of (5) is illustrated in (6) and (7). Person combinations where cliticization of both accusatives is banned, as in (7), are realized by introducing the second accusative as a free pronoun (see Sibawayh, 1881, p. 336; Jahn, 1900, pt *i*, p. 98; Wright, 1874, vol. 1. p. 104; Howell, 1880, pt *i*, vol 2. p. 540; Reckendorf, 1895, p. 394; Brockelmann, 1960, p. 59).

(6) Possible clitic combinations:
a. ʔaʕtˤaː -nɪː -ka
gave.3SG-CL. 1SG -CL.2SG.M
'He gave me you'
b. ʔaʕtˤaː -nɪː -hɪ
gave.3SG -CL. 1SG -CL.2SG.M
'He gave me him/it'
c. ʔaʕtˤaː -ka -hu
gave.3SG-CL. 1SG -CL.3SG.M
'He gave you him/it'
(Sibawayh, 1881, p. 336; Wright, 1874, vol 1. p. 103)

(7) Banned clitic combinations:

a. ʔaʕtˤaː -ka {*-nɪː / -ʔɪjːaː-ja }
gave.3SG -CL.2SG.M {-CL.1SG.M / ACC-1SG.M}
'He gave me to you' (Sibawayh, 1881, p. 335/6)

b. aʕtˤaː -huː {*-nɪː / -ʔɪjːaː-ja }
gave.3SG-CL.2SG.M {-CL.1SG.M / ACC-1SG}
'He gave me to him' (Sibawayh 1881, p. 335/6; Wright 1874, vol 1, p. 104)

c. aʕtˤaː -huː {*-ka / -ʔɪjːaː-ka }
gave.3SG -CL.3SG.M {-CL.2SG.M / -ACC-2SG.M}
'He gave him you' (Sibawayh, 1881, p. 336; de Sacy, 1905; p. 378)

The restriction on clitic combinations in (6/7) is not unique to CA. This exact pattern has also been reported for some speakers of Spanish (Perlmutter, 1971) and Catalan (Bonet, 1991, p. 179). Nevins (2007) calls it the *ultrastrong* PCC. I will refer to it simply as the PCC, since no other varieties of PCC will be discussed here (see Walkow, 2013 on the strong PCC in CA).

Combinations of third person pronouns

Sibawayh reports that the pattern of cliticization with two third person pronouns is not as categorical as that of combinations with local person pro nouns. Some combinations of third person clitics are possible, (8a), and some attested examples are reported in grammars, e.g. (8b) (also de Sacy, 1905, p. 379).

(8) Possible combinations of two third person clitics:

a. ʔaʕtˤaː {-huː -haː / -huː -haː }
gave.3SG {-CL.3SG.F -CL.3SG.M / -CL.3SG.F -CL.3SG.M}
'He gave {her to him/ him to her}' (Sibawayh, 1881, p. 336)

b. ʔasmɪʕ -humuː -hu
CAUS.hear.IMP -CL.3PL.M -CL.3SG.M
'Make them hear it!' (Reckendorf, 1895. p. 394)

This is consistent with the generalization in (5), Sibawayh notes, because neither pronoun is more local than the other. At the same time, he reports that even the possible combinations like (8) are usually avoided. Like the banned clitic combinations with local person in (7), combinations of two third person pronouns are typically realized by using a free pronoun for the second accusative, (9).

(9) Realizing Combinations of third person pronouns:

a. ʔaʕtˤaː -hu ʔɪjːaː-hu
gave.3SG -CL.3SG.M ACC-3SG.M
'He gave him it' (Sibawayh, 1881. p. 336; Wright, 1874, vol 1. p. 104)

b. zaʊw⟨ː⟩adʒa -hu ʔɪjːaː-haː
marry(CAUS).3SG.M -3SG.M ACC-CL.3SG .F
'He made him marry her' (Brockelmann, 1960, p. 59)

c. ʔa-nzɪlu -hu ʔɪjːaː-haː
CAUS-inhabit.3SG.M -CL.3PL.M ACC-3SG.F
'He made them inhabit it' (Reckendorf, 1895, p. 394)

Wright (1874) furthermore reports that combinations of two masculine singular clitics as in (10) are categorically banned.

(10) *ʔaʕtˤaː -hu -hu
gave.3SG -CL.3SG.M -CL.3SG.M
'He gave him to him' (Wright, 1874, vol 1, p. 103)

Combinations of third person pronouns resemble those that violate (5). Cliticization of both arguments is (typically) avoided. Instead, the second accusative is realized as a free pronoun. When combinations of third person pronouns do surface, the two pronouns tend to differ in their gender/number-specifications. I would like to propose that the restriction in (5) and the restrictions on third person pronouns are part of the same larger restriction in (1). It is not just that the first accusative must not be less local than the second, but that the first accusative has to be more local than the second.

Syntactic analyses of the PCC typically sideline restrictions on combinations of third person pronouns arguing that these are morphological (e.g. Anagnostopoulou, 2003; Nevins, 2007; Řezáč, 2008). This view is supported by the fact that restrictions on combinations of third person clitics in Romance and Basque force a change in the morphology of one of the clitics, rather than the use of a free pronoun as does the PCC. This difference between how restrictions on combinations of third and local person are circumvented is absent in CA. When cliticization is impossible, the second accusative is always realized as a free pronoun. This parallelism of how the grammar avoids restrictions on third and local person and the fact that all person restrictions fall under the same generalization in (1) suggests that all of (1) should follow from the same processes in the same module of grammar (for a more general argument to this point, see Walkow, 2012a).

In light of (1), what is to be made of possible third person clitic combinations like (8)? I see two interpretations of their grammaticality that maintain (1), but the sparsity of data does not allow us to distinguish between the two at this point. The first possibility is that (8) is the correct generalization about person, but that its effect is obscured in the realm of third person by the effect of number and gender. Phenomena of this kind are found in some dialects of Catalan (Walkow, 2012a, §2.4.3.3/4). Non-Valencian Standard Catalan, for example, imposes restrictions on combinations of third person direct and indirect object clitics (DO and IO) as long as IO is singular. When IO is plural, no restrictions arise. The effect of number on clitic restrictions, however, is limited to combinations of third person clitics (e.g. Walkow, 2012a, §4.3.3.1). CA may illustrate another such interaction of non-person features with cliticization that is limited to third person pronouns. A second possibility is to adopt the proposal that there are marked and un- marked instances of third person (e.g. Adger and Harbour, 2007; Béjar, 2003; Béjar and Řezáč, 2009; Ormazabal and Romero, 2007; Walkow, 2012a). It is often observed that some third person arguments behave syntactically like local person ones. More specifically, third person arguments that are introduced as the specifiers of applicative heads or *v* often behave like marked third person. Section 3 will argue that the first accusative is the subject of a causativized predicate, making it a prime candidate for being marked third person. Combinations of two third person clitics would then be possible when the second accusative is an unmarked third person and the first one is a marked one that shares some property with local person. Such an analysis would weaken the letter of (1), but is entirely consistent with how 'more local' is implemented syntactically in Section 4 and how marked vs. unmarked third person is typically implemented syntactically (in particular Walkow, 2012a, §4.2). Either of these lines of analysis would assimilate the CA data in (8) to known phenomena of restrictions on agreement and cliticization while maintaining (1) as the fundamental generalization about the role of person in restricting clitic combinations in CA.

Alternate strategies for realizing banned clitic combinations

Another set of questions concerns the alternative structure used when clitic combinations are blocked. As shown in (7/9), the second accusative is realized as a free pronoun in such contexts, irrespective of the particular person combination. CA differs in this respect from languages like Spanish and Catalan, where banned clitic combinations are avoided by changing the properties of IO when it is third person, but can be avoided by changing the properties of DO when IO is local person (Bonet, 1991). CA's consistent choice of the second accusative as the argument to change when when cliticization is blocked needs an explanation. The second question concerns what morphosyntactic change the alternative realization consists of. A remarkable property of PCC repairs in many languages is that they use morphosyntactic strategies that are not otherwise available (Bonet, 1994; Řezáč, 2007, 2011). For example, Spanish requires clitic doubling of strong DO pronouns, unless such pronouns are used to avoid the PCC. Similarly, PCC repairs on indirect objects in French have been argued to introduce the recipient as a PP rather than a dative, which is not otherwise possible (Postal, 1990; Řezáč, 2007). CA fits this pattern to the extent that free pronouns are not freely available (§2). The question then is why CA chooses the free pronoun over other strategies like introducing the theme argument with a preposition (possible with some double accusative verbs – e.g., Lane, 1863, vol. 5, p. 2084 for [ʔaʕtˤaː], 'give') when realizing pronoun combinations where cliticization is impossible.

Summary

This section has outlined the CA phenomena that will be analyzed in the remainder of this paper. Clitic combinations in verbal contexts observe the restriction in (1) that subsumes both the PCC, (5), and restrictions on third person. Person combinations where cliticization of both arguments is impossible are consistently realized by introducing the second accusative as a free pronoun. The next section shows that a core set of verbs where clitic restrictions arise are causative, and Section 4 shows that this derives (1) and the facts discussed in Section 2.3.

3. The structure of causative double accusative verbs

The majority of verbs that take two accusative objects in CA share morphological, semantic and syntactic characteristics that indicate that they are causatives rather than double object constructions as they are found in English or Romance (e.g. Pylkkänen 2002, Cuervo 2003). This section presents evidence that these verbs have the underlying structure in (11) with three major properties: (i) Double accusative verbs are causatives, (ii) v^{caus} takes (at least) v^{ag} as its complement, and (iii) head movement raises V to v^{ag} and v^{caus} . Head-movement is shown here and below with dashed arrows.

(11) [S^{causer} [**V. v^{ag}. v^{caus}** [**DP^{causee}** [**V.v^{ag}** [**V DP** ^{DO}]]]]]
[ACC] [ACC]

Under this analysis, the first and second accusative are the subject and object of a causativized verb. This structure will be central to how (1) arises and why structures like (7) and (9) surface when it is violated. Section 5 discusses non-causative contexts where clitic combinations arise and briefly shows how the analysis motivated on the specifics of causatives could extend to these contexts.

The relationship between double accusative verbs and causativity has long been noted in grammars of CA. Wright (1874, vol. 2, p. 47) divides the verbs that take two accusative complements into "all causatives of the [CCːC] and [ʔaCCC] verbal forms" on the one hand and, roughly speaking, verbs that take small clause complements. That is, double accusative verbs are derived verbs bearing the CCːC- and ʔaCCC- morphemes. Many uses of these morphemes are described as causative (e.g. Brockelmann, 1960, p. 139; Howell, 1880, pt *iii* pp. 265–270; Reckendorf, 1895, pp. 44–6; Wright, 1874, vol. 2, p. 48). Relatedly, Howell (1880, pt *iii*, p. 103) and Reckendorf (1895, p. 113) observe that the thematic relation between the first and the second accusative often mirrors the relation between the subject and the object of the verbs that the double accusative verbs are derived from. This is illustrated by pairs of derived verbs their bases like [samɪʕa], 'hear,' vs. [ʔa-smaʕa], 'make hear' (8c) and [tˤaʕama], 'eat,' vs. [ʔa-tˤʕama], 'feed' (8d), for ʔaCCC, and [ʕalɪma], 'know,' vs. [ʕalːama], 'teach' (14) below, for CCːC. The give-verb [ʔa-ʕtˤaː] in (6/7) is derived from the verb [ʕatˤaː] meaning "reach with the hands (to take)" (Lane, 1863, vol. 5, p. 2084).

Syntactic evidence for a causative structure rather than a double object one comes from binding. Baker (1988, pp. 210ff) observes that causatives and double object constructions differ in their binding possibilities. Objects in double object constructions can only corefer with the subject if they are reflexive, (12a). Non-reflexive pronouns that are objects in causatives can corefer with the matrix subject, (12b). English data with periphrastic causatives appear here for convenience. Baker demonstrates these facts for morphological causatives. Example (13) shows that CA double accusative verbs pattern with causatives rather than double object constructions.

(12) a. John$_i$ showed Mary {him$_{*i}$/ himself$_i$.}
b. John$_i$ made Mary see {him$_{i\surd}$/ *himself$_i$.}

(13) ʔa-qbadˤ-tu$_i$ -<u>ka</u> <u>ʔijːaː-ja</u>^{i}
CAUS-take.3SG -CL.2SG.M ACC.1SG.M
'I made you take me' (de Sacy, 1905, p. 378)

The datum in (13) also adds evidence that causatives in CA involve head movement. Baker (1988) argues that the formation of causatives can involve head movement of V or phrasal movement of VP. These two options, he shows, affect the binding behavior of the direct objects in causatives. In VP-movement causatives, the direct object inside the VP is moved across the causee and placed in the same binding domain as the causer. Such causatives only allow coreference of the causer and the direct object when the direct object is a reflexive. Causatives where only V moves do not change binding relations. Such causatives show the behavior in (13), where the causer and the direct object can co-refer while the direct object is a non-reflexive pronoun.

Finally, there is evidence that causative heads in CA can take complements as large as v^{ag}. Pylkkänen (2002, §3.4) shows that causatives differ on how big the complement of the causative head is. It can be the only the verbal root, the VP containing the verb and its internal argument, or a structure that contains the projection that introduces the verb's external argument (v^{ag}P here). Only languages of the third type allow causatives of transitive verbs that result in a structure with three arguments like in (11). As mentioned above, many double accusative verbs are transparently related to transitives, suggesting that they are of the third type. The structure in (11) also makes a claim about the syntactic relation of the causee and the DO: the former is syntactically higher than the latter. No direct evidence for c-command between these two arguments is available. Support for a position of the causee above the DO comes from clitic order in nominalizations. Nominalizations allow combinations of subject and object clitics to a limited extent (§5) as in (14). In such combinations, the clitic of the subject, here –[iː], precedes that of the object, -[hɪ].

(14) ħub: -iː -hɪ
love.M-CL.1SG.GEN -CL.3SG.M
'My love of him' (Wright, 1874; vol. 2, p. 59)

The leftward position of the subject clitic mirrors its higher syntactic position. In the structure in (11), the clitic order of causee and DO would mirror the underlying syntactic structure in the same way.

In summary, the three major features of the structure in (11) can be empirically supported. This supports the phrasing of (1) in terms of the syntactic relation between the two arguments.

Deriving the person case constraint in classical Arabic

This section presents a Cyclic AGREE based analysis of the data in Section 2. Section 4.1 introduces the AGREE based system of case and cliticization, and the necessary assumptions about the syntactic representation of person. Section 4.2 shows how that system combined with the syntactic structure in (11) derives (1). Section 4.3 shows how the output of the syntactic computation is realized morphologically as either a clitic combination or a free pronoun for the second accusative.

AGREE, cliticization and person restrictions

I adopt a system of case and agreement following Chomsky (2000, 2001), where the operation AGREE, (15), establishes a relation between the unvalued φ-features of probes ([$u\varphi$]) and the valued ones of goals ([φ]) that results in the valuation of the probe's $u\varphi$– and the goal's case-features.

(15) Matching is a relation that holds of a probe P and a goal G. Not every Matching pair induces Agree. To do so, G must (at least) be in the domain D(P) of P and satisfy locality conditions. The simplest assumptions for the probe-goal system are shown [below:]
a. Matching is feature identity.
b. D(P) is the sister of P.
c. Locality reduces to "closest c-command." (Chomsky 2000, p. 122)

The identity of the probe, T or *v*, determines whether the case on the goal is nominative or accusative. Valued features on a probe can no longer enter AGREE-relations and become *inactive*. The focus of the discussion here will be on objects. As an extension of the idea that *v* assigns case to object, I assume that both of the *v* projections in (11) do so. This leads to the relations of AGREE and case assignment illustrated in Table 2.

In addition to case and agreement, AGREE also drives movement and cliticization (e.g. Anagnostopoulou, 2003, 2005; Béjar and Řezáč, 2003; Preminger, 2011a). The probe responsible for cliticization is often assumed to be *v* making cliticization a byproduct of case assignment (Anagnostopoulou, 2003; Béjar and Řezáč, 2003). Within this framework, restrictions on cliticization are restrictions on successful AGREE. Syntactic analyses of person restrictions (Anagnostopoulou, 2003, 2005; Béjar, 2003; Béjar and Řezáč, 2003, 2009) have shown that they arise when two arguments compete to AGREE with the same agreement head. For the clitic restrictions under discussion here this is *v*. When both arguments AGREE, both cliticize. When one argument fails to AGREE, it hast to be licensed by some other head leading to a different morphosyntactic realization. Otherwise the derivation crashes (Řezáč, 2007, 2011). Central to this type of analysis is the idea that probes carry multiple features that enter AGREE independently of one another. In particular, Béjar (2003) and Béjar and Řezáč (2003, 2009) argue that

Table 2. AGREE relations in a double accusative construction.

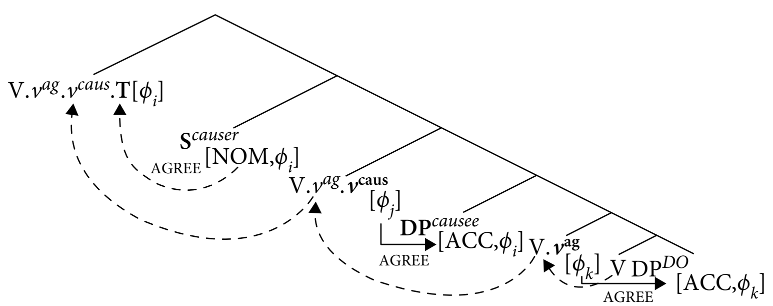


Table 3. Person categories as bundles of privative features (Béjar and Řezáč, 2009, p. 42).

Person:	**3rd**	**2nd**	**1st**
Feature Specification:	[π]	[π] [PART(ICIPANT)]	[π] [PART(ICIPANT)] [SPEA(KER)]

person itself is not a monolithic probe, but consists of multiple independent features. Their proposal builds on Harley and Ritter's (2002) work on the structure of pronoun inventories. Harley and Ritter observe that pronoun inventories across languages show implicational relationships between person categories. They model these by morphological structures where more local person categories ('1/2') subsume the structure of third person ones ('3'). Béjar and Řezáč argue that the implicational relations among person categories that are visible in the morphology are also present in their syntactic representation. They model this by representing person categories as sets of privative features where subset relations reflect the implicational relationships between person categories, Table 3. The feature specification of more local person categories subsumes the features of less local ones. Third person is the person category with the fewest features, represented only as [π]. First and second person have an additional [PART(ICIPANT)] that sets local person categories apart from third. First and second person are further distinguished from one another by the feature [SPEA(KER)].

The representations in Table 3 allow a restatement of the 'more local' part of (1) in terms of subset relations among features. A person category *X* is 'more local' than person category *Y*, if *X*'s person features are a superset of *Y*'s. I will write Φ(*X*) to mean *the person features of X*. The CA Person Constraint can now be restated as in (16).

(16) The Classical Arabic Person Constraint:
Two pronouns *x* and *y*, where $x > y$, cliticize if
a. $\Phi(x) \supset \Phi(y)$
b. where $\Phi(1) \supset \Phi(2) \supset \Phi(3)$

The feature bundles in Table 3 represent person on goals. Probes have unvalued variants of these features (on valuation in a system with privative features see Béjar 2003). How many of these features are present on the probe, $\begin{bmatrix} u\pi \\ u\text{PART} \\ u\text{SPEA} \end{bmatrix}$ vs. $\begin{bmatrix} u\pi \\ u\text{PART} \end{bmatrix}$ vs. [π], determines how finely the probes distinguish between person categories (see Béjar, 2003; Béjar and Řezáč, 2009; Walkow, 2013 for the strong vs. ultrastrong

PCC in CA). For example, a probe that is specified as $\begin{bmatrix} u\pi \\ u\text{PART} \\ u\text{SPEA} \end{bmatrix}$ can distinguish between first and second person, because first person values all of its features, but second person leaves [*u*SPEA] active. A probe that is only specified for on the other hand is fully valued after AGREEING with either first or second person, and only distinguishes between third and local person. CA distinguishes between all three person categories, (5). Accordingly, I assume that its probes are specified as.

The more complex syntactic representation of person requires a refinement of the notions of MATCHING and valuation. MATCHING as defined in (15a) requires that the features of probe and goal be identical. In the system here, this would disallow AGREE between a probe specified as [π, PART, SPEA] and third or second person arguments that would only be specified as [π] and [π, PART], respectively (Béjar, 2003, §2.6). Instead, Béjar and Řezáč (2009) define MATCHING in terms of subset relations between the features ('segments') of the probe and the goal.

(17) Match Requirement:
For a probe segment [*u*F], a subset [*u*F′] of [*u*F] must match.
(Béjar and Řezáč, 2009, p. 45)

A probe has MATCHED a goal, if at least one of its features has MATCHED. In the example above, a probe that is specified as $\begin{bmatrix} u\pi \\ u\text{PART} \\ u\text{SPEA} \end{bmatrix}$ can match a goal specified as $\begin{bmatrix} \pi \\ \text{PART} \end{bmatrix}$ because the [*u*π]- and [*u*PART]-features of the probe have MATCHED the [π]- and [PART]-features of the goal. Activity, valuation and inactivity are refined as in (18).

(18) Assumptions for AGREE
a. Each feature that seeks to AGREE is active upon being inserted into the derivation.
b. When a feature [*u*F] matches with a goal [F′], AGREE copies the feature structure containing [F′] (i.e., all features that entail [F′]) to [F]; this constitutes valuing.
c. An active feature that is locally related to a nonactive feature (i.e., a feature that stands in the configuration created by [(18b)]) is no longer active. (Béjar and Řezáč, 2009, p. 45)

Condition (18a) regulates activity. All φ-features enter the derivation active for AGREE. This differs from (17) in that there, φ-features on arguments are not *per se* active, but are visible to AGREE due to the unvalued case features that they are associated with. Béjar and Řezáč move away from independent case features and try to reduce the effects of the case filter to the need for φ-agreement. I will not take a position on the necessity or desirability of case features here. For the core proposal about double accusative verbs, all that is necessary is that accusative case is the result of a relation between an argument and *v*, and that person AGREE regulates cliticization. Condition (18b) regulates valuation. Valuation is still copying of features, but it is not limited to the features that have been MATCHED. For example, when a probe agrees with multiple goals, it can happen that a probe has the unvalued features [*u*PART, *u*SPEA] and MATCHES a goal specified as. [π, PART]-features of probe and goal, but valuation will also copy the goal's [π]-feature, because it is contained in the feature structure that has been MATCHED. Inactivity, (18c), works much as before: Any feature that has MATCHED and AGREED becomes inactive. In addition, features that have contributed to valuation via being contained in a feature structure that has MATCHED and AGREED also become inactive. So in the example just discussed, the goal's [π]-feature, which is copied in valuation due to its relation to a [PART]-feature that AGREES, also becomes inactive.

The final question for this section is how AGREE between one probe with multiple goals comes about derivationally. Řezáč (2003) and Béjar and Řezáč (2009) argue that this happens in a two step process they call *cyclic expansion*. The first step of probing happens when the probe is first merged. In accordance with (15), it can probe anything in its complement in this step. A second opportunity for probing arises in the next cycle of the derivation from the interaction of projection and the merger of the specifier. When the head and its complement are merged, a label for the resulting projection is created. According to Chomsky (2000), this label is identical to the head. The label of the projection accordingly contains the probe with its unvalued features. When the specifier is merged, it is the sister of this new projection and thereby in its c-command domain. This allows probing into the specifier. Cyclic Expansion derives the probe's ability to AGREE with higher arguments from the cyclic construction of syntactic structure. Under the logic of (in)activity, only features of the probe that did not AGREE in the first step can probe in the second one. Combined with the feature structures in Table 3, this means the second AGREE is possible when the first goal has a subset of the features on the probe, and the second goal has a superset of the first goal's features. Put differently, one probe can AGREE with two goals when the second goal is more local than the first. Whenever the first goal has the same or a superset of the features of the second goal, the second step of AGREE will be impossible.

Comparing this derivational logic to (1′) and (11) suggests that the probe responsible for clitic restrictions in CA is the one on v^{ag}. It is the only probe that is local enough to both DO and the causee to AGREE with them via Cyclic Expansion.

The syntax of person restrictions

I begin with the derivation of possible clitic combinations where the causee is more local than the direct object. Due to the structure of person categories in Table 3, any such derivation will leave active features on v^{ag} after AGREE with the object that the causee can value. Thus v^{ag} can AGREE with both arguments allowing them to cliticize.

This is illustrated in Table 4 with the derivation for (6a), where a second person object and a first person causee cliticize. The second person of the object is represented as $\begin{bmatrix} \pi \\ \text{PART} \end{bmatrix}$, the first person causee as $\begin{bmatrix} \pi \\ \textit{PART} \\ \textit{SPEA} \end{bmatrix}$. In Step 1, V has already undergone movement to v^{ag}. v^{ag}'s [$u\pi$]- and [u PART]- features AGREE with the corresponding features on the object. The object's φ -features are copied to the probe (not shown), rendering [(u)π] and [(u)PART] on probe and goal inactive (indicated by crossouts in Step 2). The probe's [uSPEA]-feature on the other hand remains unvalued and active. In Step 2, v^{ag}'s features have been projected to the label of the projection consisting of it and VP. When the causee has been merged with this projection, it becomes the sister of $v \begin{bmatrix} \pi \\ \textit{PART} \\ u\textit{SPEA} \end{bmatrix}$, allowing v^{ag}'s [uSPEA]-feature to probe the cause and AGREE with its [SPEA]-feature. In accordance with (18), this leads to the copying of all of the causee's features to the probe, and their deactivation. In this person combination, v^{ag} has AGREED with both arguments allowing both of them to cliticize.

Two more remarks are in place about unvalued features on probes. In combinations of third person direct objects and second person indirect ones, v^{ag} will have an unvalued [uSPEA]-feature left at the end of the derivation. More generally, in all possible clitic combinations, the probe on v^{caus} does not AGREE at all. Both of these potential concerns are put to rest by recent work arguing that probes have rather

Table 4. Derivation for *ʔaʕtˤaː -nɪː -ka*, 'He gave you to me' (6a).

Step 1: v^{ag} AGREES with 2-DO.	Step 2: v^{ag} AGREES with 1-causee.

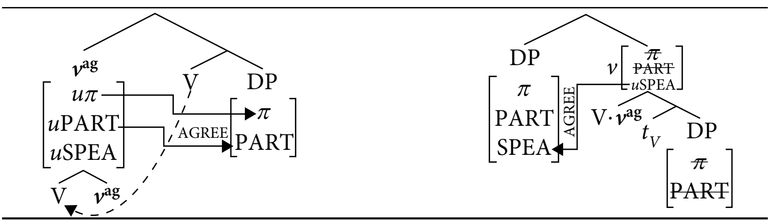
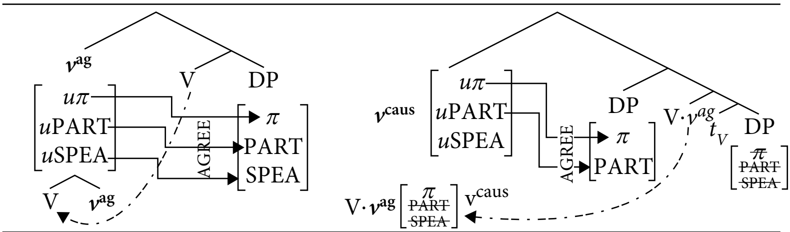


Table 5. Derivation for ʔaʕtʕaː -ka ʔɪjːaː-ja, 'He gave me to you' (7a).

Step 1: v^{ag} AGREES with 1-DO.	Step 2: v^{ag} AGREES with 2-causee.



modest licensing needs. Béjar and Řezáč (2009) propose that probes are licensed if at least one of their features has successfully AGREED. Preminger (2010, 2011a) more generally argues that a probe's failure to AGREE successfully never leads to ungrammaticality. Within that proposal, v^{caus}'s failure to AGREE is entirely unproblematic.

Let us now turn to clitic combinations where cliticization of both pronouns is not possible. In all clitic combinations where the direct object is not less local than the causee, the direct object will deactivate any features on v^{ag} that the causee could value. As a result v^{ag} can never AGREE with both the direct object and the causee. This is illustrated in Table 5 for example (7a), a combination of a first person direct object with a second person causee. This time the direct object has all three person features, while the causee has only [π, PART]. In Step 1, v^{ag} AGREES with DO in all of its features leaving it entirely inactive. When v^{ag}'s specifier is merged and v^{ag} becomes the label of the new structure, no second AGREE is possible, because *vag* is inactive. In Step 2, v^{caus} is merged and probes the causee. The two AGREE in all of the causee's features. Since neither probe has AGREED with both arguments, they cannot both cliticize.

The ban on combinations of third person pronouns, where DO and causee have the same features, is derived in the same way. DO deactivates v^{ag}'s [$u\pi$]. Since this is the only feature the causee has, AGREE between it and v^{ag} is impossible even though v^{ag} has active features left. The analysis for restrictions on clitic combinations involving local person pronouns thus naturally extends to those between third person ones.

The morphological realization of pronoun combinations

Section 4.2 showed under which conditions one probe can AGREE with two goals. This section turns to how the syntactic structures in Tables 4 and 5 are mapped to the morphological forms in (6/7), in particular why person combinations that disallow cliticization are realized by cliticizing the causee and using a free pronoun for the direct object. I identify two factors that govern cliticization in CA: (i) Cliticization spells out the syntactic dependencies of the highest valued probe in the *v*-domain and (ii) Cliticization and realization as a strong pronoun are two different ways of spelling out the syntactic relation between the probe and the goal: Clitics are the morphological realization of the goal's properties on the probe, while free pronouns are the realization of the probe's properties on the goal. The proposal is phrased in a late insertion model (e.g. Distributed Morphology, Halle and Marantz, 1993) where the nodes manipulated by syntax do not contain phonological material and such material is inserted in the post syntactic component, sensitive to the output of the syntactic computation.

The syntactic dependencies in the *v* domain are spelled out as two clitics when v^{ag} has AGREED with both DO and the causee, in Table 4, and v^{caus} does not AGREE at all. This means that PF interprets a complex head of the form '$v^{caus}_{u\varnothing} \cdot v^{ag}\begin{bmatrix}\pi \\ \textsc{part} \\ (u)\textsc{spea}\end{bmatrix}.\mathrm{v}$' by mapping the syntactic dependencies of v^{ag} with both DO and the cause into the pair of clitics like *-niː-ka*. The unvalued probe on v^{caus} is ignored. The major difference between this and person combinations where only the causee cliticizes is that the complex head in the *v*P domain contains two valued probes as in Table 5: $v^{caus}\begin{bmatrix}\pi \\ \textsc{part} \\ u\textsc{spea}\end{bmatrix}.v^{ag}\begin{bmatrix}\pi \\ \textsc{part} \\ \textsc{spea}\end{bmatrix}.V$. This syntactic structure is spelled out by cliticizing only the causee. Taken together, this suggests that when faced with a complex head with multiple valued probes, PF spells out the dependencies of only the highest, valued probe in the *v*-domain as a clitic. I have no explanation why this would be so, but a similar pattern is reported for the phenomenon of double agreement in Dutch dialects by van Craenenbroeck and van Koppen (2002). The reason then why in the causee cliticises rather than DO when clitic combinations are impossible is that the interaction of two probes and head movement bleeds the realization of v^{ag}'s dependencies as cliticization.

The use of a free pronoun for DO when it cannot cliticize can be understood in analogy to PCC repairs in Romance languages. PCC re-pairs in Romance languages often consist of a preposition or case marker a plus a non-nominative pronoun (e.g. Spanish and Catalan: Bonet 1991; French: Řezáč 2008, p. 98; Italian: Cardinaletti and Starke, 1999, p. 169). The pronoun forms with *a* also appear when pronouns are topicalized or coordinated (Cardinaletti and Starke 1999). Řezáč (2007) suggests that *a* in PCC repairs is the morphological realization of a probe that is added to avoid a crashing derivation and license an argument that has failed to AGREE with *v*. This last-resort addition of a probe accounts for the fact that PCC repair strategies are not freely available elsewhere and sometimes morphosyntactically unusual (§2.3). [ʔɪjːaː]- resembles *a* both in its distribution, see (2/3), and in its association with case. Analogous to Řezáč's (2007) proposal for *a*, I propose that [ʔɪjːaː]- is the realization of the probe that licenses the direct object. Unlike in his proposal, however, this probe is not added to avoid a crashing derivation, but is the independently present probe of v^{ag} spelled out on the object, rather than v^{ag} .

In conclusion, the realization of banned clitic combinations with strong pronouns can be attributed to an alternate way of spelling out syntactic dependencies that are independently established. The ungrammaticality of the clitic combinations in (7) then is not the result of a crashing derivation, but rather there is no derivation that leads to a clitic combination in these person combinations. A derivation that contains the relevant lexical items in the numeration will always lead to a morphological realization of the direct object as a free pronoun rather than a clitic. In this respect the proposal here resembles Preminger's (2011a) proposal about ungrammaticality in other agreement restrictions.

4. Clitic combinations in other contexts

In addition to the causative verbs discussed so far, there are at least four contexts where combinations of pronominal clitics are possible: (i) some underived verbs, (ii) nominalizations like (14), (iii) some verbs with small clause complements, and (iv) prepositional imperatives. Sibawayh (1881, p. 336) reports that clitic combinations of the second and third type are rather marginal, and the status of (iv) is somewhat unclear (Sibawayh, 1881, p. 334/§208). I will only discuss (i) and (ii), and outline how the analysis in Section 4.2 could extend to them.

There are some underived verbs that take two accusative complements and allow cliticization, e.g. (19). Wright (1874, vol 2, p. 47) reports that these are verbs with meanings similar to *fill, satisfy, give, deprive, forbid, ask* or *entreat*.

(19) Clitic clusters on non-causative double object verbs:

a. na-biːʕu-ka -hu
1PL-sell-CL.2SG.M -CL.3SG.M
'We sell you it' (Reckendorf, 1895, p. 111)

b. sal -niː -haː
ask for.IMP.M.SG -CL.1SG -CL.3SG.F
'Ask me (about) it!' (de Sacy, 1905, p. 379)

(20) dˤaraba zaid-an ʔa ðˤːahr-a
beat.3SG.M Zaid-ACC DEF.back-ACC
'He slapped Zaid (on) the back' (Reckendorf, 1895, p. 113)

In addition, normally transitive verbs like *dˤaraba*, 'beat,' in (20), can take two objects when one is a body part of the other. Reckendorf (1895, p. 113) reports that the construction in (20) typically involves an animate first accusative and that the whole individual denoted by the first accusative is affected by the action performed on its part. The construction is reminiscent of possessive datives in Romance languages that require a similar part-whole relation between two internal arguments (e.g. Spanish: Cuervo, 2003, §2.2.2/3). I was unable to find information on whether constructions like (19)/(20) are subject to (1). If they are, they could be integrated into the analysis here by treating them as true double object constructions similar to English, and adopting an applicative analysis along the lines of Pylkkänen (2002) and Cuervo (2003). In applicative double object constructions, the non-theme argument is introduced as the specifier of an applicative head (APPL) that can be in several positions below v (Cuervo 2003). To derive the clitic restrictions from the interactions of head movement and two probes, APPL^0 would have to bear a φ-probe just like v (e.g. Adger and Harbour, 2007) and undergo head movement. To fully motivate such an analysis one would have to show that constructions like those in (19/20) have the properties of associated with applicatives, which is beyond the scope of this paper.

Clitic combinations are also possible in nominalizations as in (14): [ħubː -iː -hɪ]. Sibawayh reports that they observe (5), but a strong pronoun is commonly used for the second clitic even in person combinations that obey (5) (Sibawayh, 1881, p. 333, §207; Jahn, 1900, pt *ii*, p. 95; also Howell, 1880, pt *i*, vol. 2, p. 542). The proposal here derives cliticization in nominalizations in the same way as in the verbal domain. Subjects and objects are introduced by v^{ag}, and v^{ag} 's probe can AGREE with both of them under the same conditions as elsewhere. The presence of (1) in nominalizations and with verbs follows without additional stipulation. An important difference between (14) and other clitic combinations, however, is case. The subject clitic in (14) bears genitive, rather than accusative as would be expected from case assignment by v^{ag}. The case of the object clitic is not morphologically marked, but the case assignment patterns with non-pronominal arguments in nominalizations show that genitive subjects appear together with accusative objects (Wright, 1874, vol 2., p. 58). If both subject and object AGREE with v^{ag} to cliticize and receive case, the two arguments should not have different case values. PCC between arguments with different cases is quite common. In Romance languages the PCC typically arises between dative and accusative arguments. A preliminary proposal for these contexts is that they are further evidence for separating restrictions on cliticization from case assignment. I have argued in previous work on Catalan (Walkow, 2012a,b) that the assignment of dative an accusative case is separated from person agreement. Preminger (2011a, b) makes a more general argument for separating person licensing from the case filter. Independently, there is a long line of work on deriving case from syntactic configurations of arguments rather than by assignment via functional heads (Marantz, 1991; McFadden, 2004; Bobaljik, 2008; Preminger, 2011a; Pesetsky, 2012). Extending this division of φ-AGREE and case to CA, one could say that the mechanism of φ -AGREE developed here accounts for the possibility of cliticization, but other aspects of the syntactic structure determine the case of the arguments.

A multiple AGREE analysis

The only extant analysis of the clitic restrictions in CA as restrictions on AGREE is in Nevins (2007) where it is part of a larger proposal about variation in person restrictions. The proposal differs from the one here in several points. It is built on Multiple AGREE, which allows one probe to AGREE with multiple goals in one step. Person restrictions arise from restrictions on the operation Multiple AGREE. Different kinds of PCC arise from how the restrictions on Multiple AGREE are parameterized. The system uses binary rather than privative features. I will begin by presenting the Multiple AGREE analysis for (5), and then go on to show how it falls short at accounting for restrictions on third person clitics.

The Multiple AGREE analysis relies on representations of person categories using two binary features [±PART(ICIPANT)] and [±AUTH(OR)]. The three person categories are represented as 1: [+PART, +AUTH], 2: [+PART, –AUTH] and 3: [–PART, –AUTH]. The binary status of the features is key to how Multiple AGREE is restricted. The Multiple AGREE proposal also differs in its assumptions about syntactic structure. The probe that is responsible for deriving clitic combinations is always above both arguments. In the syntactic structure argued for in Section 3, it would be $\left[v \begin{smallmatrix} caus \\ [u\phi] \end{smallmatrix} \left[\mathrm{DP}^{\mathrm{causee}} \ldots \mathrm{DP}^{\mathrm{DO}} \right] \right]$.

Person restrictions arise from two restrictions on Multiple AGREE called Contiguous Agree and Matched Values. Matched Values will not be discussed here as it does not contribute to the restrictions in CA. Informally, Contiguous Agree requires that when a probe P AGREES in some feature with an argument x in a structure P > y > x, y also has to have that feature. The feature sensitivity of Contiguous Agree is parameterized and the different parameterizations derive different patterns of PCC. Nevins (2007) proposes that Contiguous Agree in CA is parameterized to [+AUTH] and [+PART], leading to the definition of Contiguous Agree in CA in (21).

(21) Parameterization of Contiguous Agree in Classical Arabic:
For the value {+} of the features [±PART] and [±AUTH] on a Probe P,
If x is [+AUTH], ¬∃y, such that y > x and P > y and y is not [+AUTH]

and

If x is [+PART], ¬∃y, such that y > x and P > y and y is not [+PART]
There can be no interveners between P and x that are not valued {+}
for whichever of [+PART] and [+AUTH] x has.
(based on Nevins, 2007, pp. 291, 298)

Contiguous Agree as in (21) derives the clitic restrictions as follows. *-niː-ka*,1ACC-2ACC, satisfies (21), because the lower 2 has a [+PART] that is shared by the higher 1. **-huː-niː*, 3ACC-1ACC, and **-huː-ka*, 3ACC-2ACC, are ruled out, because the lower arguments' [+PART] and [+AUTH] features are not shared by the higher 3. Similarly, *-ka-niː, 2ACC-1ACC, is ruled out because the probe could AGREE with the lower 1 in both [+AUTH] and [+PART], but the higher 2 lacks [+AUTH]. Finally, *-niː-hɪ*, 1ACC-3ACC, and *-ka-hu*, 2ACC-3ACC, are ruled in, because they satisfy (21) vacuously. Contiguous Agree is operative only when the lower argument is specified for [+AUTH] or [+PART]. Since third person lacks both of these specifications, any combination that contains a 3-DO satisfies it and Multiple AGREE can target both internal arguments.

The multiple AGREE account does not deliver restrictions on third person pronouns. Since third person lacks the features that Contiguous Agree is parameterized for, combinations of third person pronouns vacuously satisfy it and can enter Multiple AGREE. The proposal predicts that combinations of third person pronouns should be generally possible, which is not the case. Nevins (2007) discusses restrictions on third person pro-nouns in Basque and Spanish, and attributes them to post-syntactic processes. As argued in Section 2.2, this type of explanation is not plausible for CA.

In conclusion, the Multiple AGREE system cannot account for the full set of person restrictions in (1). The Cyclic AGREE proposal presented here gives a unified analysis of (1) while conservatively locating variation in the functional lexicon.

References

Adger, D., & Harbour, D. (2007). The syntax and syncretisms of the person case constraint. *Syntax*, 10, 2–37. DOI: 10.1111/j.1467-9612.2007.00095.x

Anagnostopoulou, E. (2003). *The syntax of ditransitives: Evidence from clitics*. Berlin: Mouton de Gruyter.

Anagnostopoulou, E. (2005). Strong and weak person restrictions: A feature checking analysis. In L. Heggie, & F. Ordóñez (Eds.), *Clitic and affix combinations* (pp. 199–235). Amsterdam: John Benjamins. DOI: 10.1075/la.74.08ana

Baker, M. (1988). *Incorporation: A theory of grammatical function changing*. Chicago, IL: University of Chicago Press.

Béjar, S. (2003). Phi-syntax: A theory of agreement. Unpublished Ph.D. dissertation, University of Toronto.

Béjar, S., & Řezáč, M. (2003). Person licensing and the derivation of PCC effects. In A. T. Pérez-Leroux, & Y. Roberge (Eds.), *Romance linguistics: Theory and acquisition* (pp. 49–61). Amsterdam: John Benjamins. DOI: 10.1075/cilt.244.07bej

Béjar, S., & Řezáč, M. (2009). Cyclic agree. *Linguistic Inquiry*, 40(1), 35–73. DOI: 10.1162/ling.2009.40.1.35

Bobaljik, J. D. (2008). Where's phi? Agreement as a post-syntactic operation. In D. Harbour, D. Adger, & S. Béjar (Eds.), *Phi-theory: Phi features across interfaces and modules* (pp. 295–328). Oxford: Oxford University Press.

Bonet, E. (1991). Morphology after syntax: Pronominal clitics in Romance. Unpublished Ph.D. dissertation, MIT, Cambridge, MA.

Bonet, E. (1994). The person-case constraint: A morphological approach. In H. Harley, & C. Phillips (Eds.), *The morphology-syntax connection, number 22 in MIT working papers in linguistics* (pp. 33–52). Cambridge, MA: MIT Press.

Brockelmann, C. (1960). *Arabische Grammatik*. Leipzig: Verlag Enzyklopadie.

Cardinaletti, A., & Starke, M. (1999). The typology of structural deficiency: A case study of the three classes of pronouns. In H. van Riemsdijk (Ed.), *Clitics in the languages of Europe* (pp. 145–233). Berlin: Mouton de Gruyter.

Chomsky, N. (2000). Minimalist inquiries: The framework. In R. Martin, D. Michaels, & J. Uriagareka (Eds.), *Step by step: Essays on minimalism in honor of Howard Lasnik* (pp. 89–155). Cambridge, MA: MIT Press.

Chomsky, N. (2001). Derivation by phase. In M. Kenstowicz (Ed.), *Ken Hale: A life in language* (pp. 1–52). Cambridge, MA: MIT Press.

Cuervo, M. C. (2003). Datives at large. Unpublished PhD dissertation, MIT, Cambridge, MA.

de Sacy, A. I. S. (1905). *Grammaire arabe* (Vol. 2, 3rd edn). Société anonyme de l'imprimerie rapide, Tunis, (first edition 1810).

Fassi Fehri, A. (1988). Agreement in Arabic, binding and coherence. In M. Barlow, & C. A. Ferguson (eds.), *Agreement in natural language: Approaches, theories, descriptions* (pp. 107–158). Stanford, CT: Center for the Study of Language and Information.

Halle, M., and Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale, & S. Keyser (Eds.), *The view from building 20: Essays in linguistics in honor of Sylvain Bromberger* (pp. 111–176). Cambridge, MA: MIT Press.

Harley, H. & Ritter, E. (2002). Person and number in pronouns: A feature-geometric analysis. *Language*, 78(3), 482–526. DOI: 10.1353/lan.2002.0158

Howell, M. S. (1880). *A grammar of the classical Arabic language in four parts*. Allahabad: North-Western Provinces and Oudh Government Press.

Jahn, G. (1900). *Ŝıbawaihi's Buch ü ber die Grammatik* (Vol. 2). Berlin: Reuther & Reichard.

Lane, E. W. (1863). *An Arabic-English lexicon*. London: Williams and Norgate, London.

Marantz, A. (1991). Case and licensing. In G. Westphal, B. Ao, & H.-R. Chae (Eds.), *Proceedings of ESCOL '91* (pp. 234–253). Ithaca, NY: Cornell Linguistics Club.

McFadden, T. (2004). The position of morphological case in the derivation: A study on the syntax-morphology interface. Unpublished Ph.D. dissertation, University of Pennsylvania, Philadelphia.

Nevins, A. (2007). The representation of third person and its consequences for person-case effects. *Natural Language & Linguistic Theory*, 25, 273–313. DOI: 10.1007/s11049-006-9017-2

Ormazabal, J. & Romero, J. (2007). The object agreement constraint. *Natural Language & Linguistic Theory*, 25, 315–347. DOI: 10.1007/s11049-006-9010-9

Perlmutter, D. M. (1971). *Deep and surface structure constraints in syntax*. New York: Holt, Rinehart and Winston.

Pesetsky, D. (2012). Russian case morphology and the syntactic categories. Unpublished manuscript, MIT.

Postal, P. M. (1990). French indirect object demotion. In P. M.Postal, & B. D. Joseph (Eds.), *Studies in relational grammar 3* (pp. 104–200). Chicago, IL: University of Chicago Press.

Preminger, O. (2010). Failure to agree is not a failure: φ -Agreement with post-verbal subjects in Hebrew. In J. van Craenenbroeck, & J. Rooryck (Eds.), *Linguistic variation yearbook* (Vol. 9, pp. 241–278). Amsterdam: John Benjamins.

Preminger, O. (2011a). Agreement as a fallible operation. Unpublished Ph.D. dissertation, MIT.

Preminger, O. (2011b). Asymmetries between person and number in syntax: A commentary on Baker's SCOPA. *Natural Language & Linguistic Theory*, 29, 917–937. DOI: 10.1007/s11049-011-9155-z

Pylkkänen, L. (2002). Introducing arguments. Unpublished Ph.D. dissertation, MIT.

Reckendorf, H. (1895). *Die Syntaktischen Verhä ltnisse des Arabischen*. Leiden: 1895–1898.

Řezáč, M. (2003). The fine structure of cyclic agree. *Syntax*, 6(2), 156–182. DOI: 10.1111/1467-9612.00059

Řezáč, M. (2007). Escaping the person case constraint: Reference-set computation in the φ -system. *Linguistic Variation Yearbook*, 6, 97–138.

Řezáč, M. (2008). The syntax of eccentric agreement: The person case constraint and ansolutive displacement in Basque. *Natural Language & Linguistic Theory*, 26, 61–106. DOI: 10.1007/s11049-008-9032-6

Řezáč, M. (2011). *Phi-features and the modular architecture of language*. Dordrecht: Springer.

Sibawayh, A. i. U. (1881). *Le Livre de ŝıbawaihi, traité Arabe*, Volume 1. Paris: Imprimerie Nationale.

van Craenenbroeck, J., & van Koppen, M. (2002). The locality of agreement and the CP-domain. *Handout, GLOW 25*, April 9–11, 2002.

Walkow, M. (2011). Person effects and the representation of third person: an argument from Barcelońı Catalan. In J. Berns, H. Jacobs, & T. Scheer (Eds.), *Romance languages and linguistic theory: Selected papers from 'Going Romance' Nice 2009* (pp. 343–361). Amsterdam: John Benjamins. DOI: 10.1075/rllt.3.20wal

Walkow, M. (2012a). Goals, big and small. Unpublished Ph.D. dissertation, University of Massachusetts at Amherst.

Walkow, M. (2012b). The syntax of the person case constraint drives morphological impoverishment of clitics. In J. Choi, E. A. Hogue, J. Punske, D. Tat, J. Schertz, & A. Trueman (Eds.), *Proceedings of the poster session of 29th West Coast Conference on Formal Linguistics, volume 20 of Coyote Papers: Working Papers in Linguistics*. Phoenix, AZ: University of Arizona Linguistics Circle.

Walkow, M. (2013). Locating variation in person restrictions. *University of Pennsylvania Working Papers in Linguistics*, 19(1), article 28.
Wright, W. (1874). *A grammar of the Arabic language*. Cambridge: Cambridge University Press.

Acknowledgment

I would like to thank Rajesh Bhatt, Seth Cable and Kyle Johnson for their feedback on earlier stages on this work as well as the audiences at the Arabic Linguistic Symposium 2012, in particular Elabbas Benmamoun, Mohammad Mohammad and Usama Soltan, and at the Penn Linguistics Colloquium 2012, in particular Anthony Kroch. All remaining faults are my own.

Phonology

Secondary stress exist in Cairene Arabic?

Rajaa Aquil

In this paper I demonstrate that secondary stress in Cairene Arabic (CA) exists, and that is by invoking Optimality Theoretic (OT) constraints of FootBinarity, Align-Head/R, Trochaic, Wsp, *Clash, and *Lapse. There is no dispute in the literature over the place of primary stress, on one of the rightmost three syllables. However, very few studies support the presence of secondary stress (Harms, 1981; Welden, 1980), while more research suggest that secondary stress does not exist (Aljarah, 2008; De Lacy, 1998; Halle & Vergnaud, 1987; Hayes, 1995; Kenstowicz, 1994; McCarthy, 1979; Watson, 2002). Although I agree with research (Gairdner, 1926; Harms, 1981; Welden, 1980) that secondary stress exists in CA, I refute the data and analysis proposed by this research.

Keywords: Cairene Arabic, primary stress, secondary stress, rhythm, Optimality Theory, segmental rule-based, metrical phonology tree and grid.

1. Introduction

Cairene Arabic (henceforth referred to as CA) is classified in the literature as a stress language (Hayes, 1995). In some stress languages, the distance between a stressed syllable and another is equal. The distance is the number of syllables intervening between a stressed syllable and another stressed syllable, and is usually limited to one or two syllables (Selkirk, 1984).

The literature, to my knowledge, does not explicitly indicate that CA stress is rhythmic; nor has CA secondary stress been tackled, except by Welden (1980) and Harms (1981). The working definition of rhythm in this paper is the distribution of primary and secondary stress within a word and the distance between a stressed syllable and another. Secondary stress in Arabic has long been debated in the literature. No known studies make the claim that Arabic in general has secondary stress. Very few studies (*cf.* Harms, 1981; Mitchell, 1956; Welden, 1980) contend that CA does, while Halle and Vergnaud (1987) Hayes (1995) and McCarthy (1984, reported in Halle and Vergnaud, 1987, personal communication, p. 60 footnote 5)[1] maintain that CA does not have secondary stress. Welden (1980), for example, states that secondary stress is present in long words of six or more syllables, specifically words that have affixes added to them.

In this paper, I motivate the notion of rhythm in CA by invoking the presence of secondary stress in CA. Through an Optimality Theoretic (OT) representation, I demonstrate that the constraint TROCHAIC secures that the default rhythmic pattern is followed in words of four or more syllables, particularly when the appropriate conditions are present. This occurs, for example, when an [L L] syllable starts a word of four syllables or more (such as [bàrabánd] 'fluent', [ʕàsaliyyáaya] 'kind of candy', or in long words that have affixes added to them (such as [mà.ka.ta.bit.hal.náaʃ] 'she didn't write it for us'). Also, by appealing to The Principle of Rhythmic Provisions as in Anti-clash Provision, and Anti-lapse Provision (Selkirk, 1984), in conjunction with Optimality Theory stress constraints, I demonstrate that it is plausible that CA has a secondary stress motivated by bimoracity (enforced by FOOTBINARITY), and stress-related constraints ALIGN-HEAD/R, TROCHAIC, WSP, *CLASH, and *LAPSE.

The point of this paper is purely theoretical. There is no intention in this paper to discuss surveys of native speakers' judgments, acoustic evidence, empirical studies or experiments with synthesized words. The aim is to compare earlier theories, e.g., rule-based, metrical phonology and OT in accounting for secondary stress. The inclusion of OT constraints is primarily to show that secondary stress can be motivated. A detailed OT study, comparing analyses to demonstrate that the analysis which predicts secondary stress is less complex and plausible than one that does not, is saved for future papers, as is a study based on "diagnosing Stress Patterns" following Hayes's (1995) paradigm.

2. Stress in CA

Primary stress

There is no dispute over the place of primary stress in CA, namely, it is on one of the rightmost three syllables (Gairdner, 1926; Halle & Vergnaud, 1987; Harms, 1981; Harrell, 1957; Hayes, 1995; McCarthy, 1979; Mitchell, 1956, 1960; Welden, 1980). The description of stress placement in CA was first formulated

1. Halle and Vergnaud (1987) reported on p. 60 footnote 5, that McCarthy (1984) personally communicated that his phonetic studies of CA speech did not reveal any trace of secondary stresses.

by McCarthy (1979), and cited by Halle and Vergnaud (1987), then Hayes (1995). It is as follows (Halle & Vergnaud, 1987, p. 61).

1. a. Stress the penult whether light or heavy.
   - i. bána[2] 'he built'
   - ii. fíhim 'he understood' (3rd SG, m)
   - iii. mibáħbaħ 'easy going'

   b. Stress the final if superheavy.
   - i. barabánd 'one who talks very fast'
   - ii. mutaʕalliqáat 'belongings'
   - iii. baraníitˤ 'hats'

   c. Stress the antepenult or the penult, whichever is separated by an even number of syllables from the immediately preceding heavy syllable or the beginning of the word (where zero separation is counted as even).[3]
   - i. ʃágara 'a tree'
   - ii. ʔibtádaʕ 'invented'
   - iii. sabahlála 'haphazardly'

Formulations of primary stress have had different representations, e.g., rule-based stress assignment (Chomsky & Halle, 1986, 1991), metrical phonological trees (Liberman & Prince, 1977), or metrical phonological grids (Halle & Vergnaud, 1987).

2. Segmental Rule-Based (Welden, 1980, p. 20).
   S is either a heavy syllable (H) or a light syllable (L)
   $S_{<L>} \rightarrow$ [+stress]/ ⟨# (X H)⟩ ____ (⟨L⟩ L) ##
   - i. The rules abbreviate the following environments. Stress the antepenult if it is a light syllable, right after a heavy syllable or starting the word, as in (a) and (b).
   - ii. Or stress the penult as in (c) or the final elsewhere as in (d).
     - a. L → [+stress]/ # X H ___ L L ## ʔimbásˤatˤu
       'they became happy'
     - b. L → [+stress]/ # ___ L L ## kátabu
       'they wrote'

2. These data are from the author, who is a native speaker of CA. The data in this paper are the author's except for Welden's (1980).

3. Generalizations in (1c) are particularly important because they demonstrate how important a heavy syllable is in demarking the site of primary stress. In fact, in the absence of a heavy syllable in the word, the edge of the word or zero separation from the edge or beginning of the word is very important. This correctly explains why a word like [šágara] has the primary stress on the antepenult and not the penult.

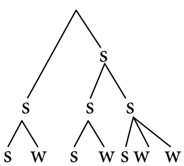
c. S → [+stress]/ ___ L ## rámа
'he threw'
maktába
'library'

d. S → [stress]/ ___ ## ʃáaf
'he saw'
katabúu<h>
'they wrote it'
kitáab
'book'

I would like to acknowledge the anonymous reviewer of the paper who drew my attention to the fact that when Welden's formula in (2) is unpacked, it would produce [ʃaafú], [ʃagaraatú], [ʔistafadná], or perhaps [ʃáafú], [ʃʕagráatú], and [ʔistafádná], all of which are ungrammatical.

The following metrical phonological tree and metrical phonological grid structures (3) demonstrate the representation of the primary stress in the word [ʕasaliyyáaya] 'kind of candy'.

3. a. Metrical Phonology Tree



ʕa.sa.liy.yáa.ya

b. Metrical Phonology Grid

2 (*) Prosodic word
1 (*.)(*.)(*.) Foot
0 ** ** ** *

μ μ μμ μμ μ Mora
ʕa.sa.liy.yáa.ya

In metrical tree theory of stress, nodes of the tree are labeled strong S and weak W to mark relative prominence. In (3a), light syllables (e.g., [ʕa.sa]) are formed into a foot and in accordance with stress being Trochaic; prominence is on the left syllable represented in the structure as S. The rimes of the heavy syllables (e.g., [liy] and [yáa]) are branched to S and W. The remaining stray material (e.g., [ya]) is gathered and connected to the tree. The right node is strong (i.e., carries stress) only if it branches. As seen in the structure, the right node branches and, accordingly, is strong and carries the main stress of the word.

Metrical phonological grid theory of stress represents stress as a hierarchically organized rhythmic structure where the grid depicts the temporal structure represented as beats. The grid shows a sequence of beats vertically aligned in grid columns. Beats are equally spaced in time but vary in strength according to their height in the grid columns. Rows in the grid refer to layers 0–2. Following Halle & Vergnaud (1987, pp. 35–47) in structure (3b), each mora [μ]is a weight-bearing unit; therefore, it gets a beat. Beats are represented by asterisks in layers 0–2. Stressed syllables get an extra beat in layer 1, and finally, due to End Rule Right, the rightmost foot gets the extra beat in layer 2.

Based on the literature on primary stress of CA, some constraints are found to be at play. These constraints are: (1) FOOT BINARITY (FTBIN), which stipulates that a foot be bimoraic (McCarthy, 1979; Watson, 2002), (2)TROCHAIC (TR), which stipulates that a left syllable in a bisyllabic word be stressed, (Hayes, 1995; McCarthy, 1979), and (3) ALIGN HEAD RIGHT (ALIGNHEAD/R), which stipulates that primary stress is on one of the rightmost three syllables (McCarthy, 1979).

An Optimality Theoretic analysis of primary stress in CA has been formulated. However, I find the point of contention in these studies to be the data being analyzed. Formulations by Aljarah (2008) and De Lacy (1998) are based on data from Cairene Classical Arabic, which is a variety of Classical Arabic pronounced by Azhar-trained Egyptians or Cairene Arabic native speakers. In my opinion, analysis of primary stress of CA based on data that is Classical Arabic falls short of capturing the prosodic changes and alternations that take place in CA. Data in (Welden, 1977, p. 6) are Classical Arabic words: (ka.ta)(BA).ha 'he wrote it', (ka.ta) BAA).ha, 'they wrote it', (mak)(TUU)bun, 'a letter', (mar)(suu)(MAA)tun, 'they are drawn'. The same applies to the data used in De Lacy (1998, footnote 1, p.1), where he states that the data are from Mitchell (1960) – a study on words as pronounced by Azhar-trained Egyptians of Classical Arabic.

There are studies in the literature that show that there is a difference between CA and Classical Arabic, particularly in prosodic structures, for example vowel shortening in closed syllables and vowel lengthening before suffixes (Harrell, 1957; Mitchell, 1960; Watson, 2002).

4. MSA input CA output[4]
   a. saaħiba sаħ.ba
      'friend' (f.SG.)
   b. taaxudi tax.di
      'you take' (f.SG.)

5. Cairene Arabic v̌ ~ v: alternation[5]
   ʔábu ʔabúuya
   'father' 'my father'

In this section, I translate the literature on CA primary stress into an OT representation (see Tableau 1). The tableau illustrates the constraints mentioned above, as

4. The following data are from (Harrell, 1957, p. 30).

5. The following data are from (Watson, 2002 cited in McCarthy 2005, p. 1).

well as the following constraints. (1) WEIGHT TO STRESS (WSP), which stipulates that a heavy syllable must carry stress, (2) PARSE SYLLABLE (PARSEσ), which stipulates that all syllables be parsed into a higher prosodic hierarchy, namely the foot, and (3) ALIGN ALL FEET LEFT (AFL), which stipulates that the formation of all feet must be on the left side of the word. I will not delve into a detailed OT analysis for primary stress; however, a paper on secondary stress would not be complete if an account of primary stress is not provided. Therefore, I present OT tableaux of primary stress in a subsequent section.[6] At this juncture of the paper and based on the literature, it is sufficient to posit that FTBIN, ALIGNHEAD/R, and TR, are highly ranked. The interaction between these constraints is discussed in 3.2 below.

Secondary stress

As mentioned earlier, there is no dispute over primary stress of CA; however, there is one over secondary stress. Only a small number of studies (Harms, 1981; Welden, 1980) contend that CA has secondary stress, while others (Halle and Vergnaud, (1987); Hayes, (1995) maintain that it does not. Gairdner (1926, p. 137) in a footnote implied the presence of secondary stresses: "Where more than one accent appears to fall on a single word-group, it will be found that the last is the strongest, and the rest secondary."

Tableau 1. Constraints involved in primary stress.[7]

Optimal forms	FTBIN	ALIGNHEAD/R	TR	WSP	PARSEσ	AFL
☞(ʕasa) (liy) (yáa)ya						
☞(bara) (níi)<tˤ>[8]						
☞mu(sal) (sála)				*	*	**
☞(fíhi)<m>						
☞(ʔid) (dal) (ʕádi)				*		***
☞(šága)ra					*	
☞(mùta) (ʕal)li(qáa)<t>					*	**

6. For a detailed analysis of primary stress on CA spoken Arabic data, I refer the interested reader to (Aquil, 2012).

7. In OT the use of dotted lines to show that there is no ranking between the constraints yet.

8. Following (Kiprasky, 2003: p. 156) and (Vaux, 2004) I consider < tˤ> in [baraníi tˤ],] 'hats', the <m> in [fihim], 'he understood' and the < tˤ> in [mùtaʕalliqáat] 'belongings' to be extra-syllabic, extra prosodic, or extra metrical (Hayes, 1995), hence, not calculated metrically. This renders the final syllable light and not heavy.

As established, primary stress in CA, as discussed in the previous section, is predictable and derived by phonological rules and constraints, rather than free and lexically determined. Phonological constraints are related to foot form and directionality, syllable weight, and edge of the word (i.e. rightmost). Furthermore, I posit that Trochaic constraint, which controls the rhythm of the language, is highly ranked. Although the literature never explicitly states that rhythm in CA is alternating, the interaction between stress-related constraints such as Align head/r, Trochaic, Wsp, and *Lapse, *Clash (discussed below) demonstrates that CA stress is a rhythmically bounded-system, where the distance between a stressed syllable and another is determined by some prosodic constituent, and that is one mora. By invoking OT constraints (Prince & Smolensky, 1993, 2004), the Principle of Rhythmic Provisions, and the Anti-lapse Provision (Selkirk, 1984) I motivate the presence of secondary stress in CA below (see Section 3).

Segmental rule-based theory

Few studies have investigated secondary stress in CA; those include work by Welden (1980) following the rule-based theory, and Harms (1981) following the metrical tree theory. Both Welden's (1980) and Harms's (1981) formulations are based on a different set of data than the one I have in this paper.

Welden (1980) theorized that CA has a rhythmic secondary stress. According to her analysis, secondary stress is iterative and works backward toward the left of the word counting from the primary stress syllable and onward.

6. Rhythmic Stress (RS) (left-iterative) (Gairdner, 1926, p. 35)
   S → [+stress]/ ⟨## (S S)0⟩ _____⟨L⟩S S
   [+ stress]

The rule abbreviates the following environments. (Gairdner, 1926, p. 36)

7. a. S → [+stress]/ ##_____L S S
   [+ stress]
   ## SS _____L S S
   [+ stress]
   ## SS SS _____L S S
   [+ stress]

   b. S → [+stress]/ _____ S S
   [+ stress]

According to the analysis, the rule accounts for environments in which the preceding syllables are even-numbered as well as odd-numbered. In some environments both (7 a and b) account for the secondary stress in the word. For example, (7a) applies with three syllables as in (8a), but (7b) applies in (8b).

8. Three syllables before primary stress
   a. S̀L̆S̆X́ — màkătăbúuʃ
      'they (pl) didn't write'
   b. S̆H̀S̆X́ — mĭstàʕmĭlíin
      'they're using'

When it comes to words with five syllables before the primary stress both rules (7a and b) are applied but in a different order as in (9a and b).

9. Five syllables before primary stress
   a. SSSLSX́ — matărgimtuhalíiʃ
      'you didn't translate it for me'
8. a. S̀L̆S̆X́ — gìmtŭhălíiʃ
   b. S̀S̆̀X̀ — màtărgìmtŭhălíiʃ
   b. SLSHSX — mastalamithalháaʃ
      'she didn't receive it for her'
8. b. H̀S̆X́ — mìthălháaʃ
   a. S̀L̆S̆X̀ — màstălămìthălháaʃ

According to the analysis, environment (7b) accounts for the words in (10a and b).

10. a. Two syllables
       L̀L̆X́ — sànătéen
       'two years'
       H̀H̆X́ — ʔìstăʕmílu
       'they used'
       L̀H̆X́ — kàtăbtáha
       'I/you wrote it'
       H̀L̆X́ — kàllĭmítha
       'she called her'
    b. Four syllables
       L̀L̆L̀L̆X — màkătàbŭháaʃ
       'they didn't write it'
       L̀H̆L̀H̆X — bìtĭrmìhŭmláha
       'she is throwing them to her'
       H̀H̆H̀H̆X — bìttărgìmhŭmláha
       'she is translating them to her'

As observed, Welden's (1980) analysis does not capture rules of primary and secondary stress assignment in one formula. Because rules are ordered, clarity and ease of explanation could be compromised. In addition, I refute the presence of a secondary stress on the first L syllable of the following words: /màkătàbŭháaʃ/ and /bìtĭrmìhŭmláha/. According to OT, these syllables are not parsed because they are not bimoraic. Bimoracity of a foot is a requirement the language faithfully obeys.[9]

We note here that Welden's (1980) analysis does not make any specific differentiation between heavy and light syllables when the preceding syllables are even numbered or when there are three syllables. However, the analysis makes a distinction between the light and heavy syllables in words of six syllables. Following Allen (1969), she proposes that a light syllable L has one light element (e) and a heavy syllable H has two light elements (e.e), however she does not specify what these elements are or which prosodic constituents they belong to. She also incorporates some informal metrical principles justifying for secondary stress, which are:

11. Metrical principles
    a. Maximize trochaic rhythm
    b. SˈSˈ (two stressed σ) is not permitted
    c. When the number of syllables preceding a stressed syllable is odd, stress falls on the syllable containing the ante-penultimate light element.

Principle (11) does not specify whether the antepenult syllable is the antepenult of the word or the antepenult syllable counting leftward from the primary stressed syllable. Also, according to Welden (1980), the rule in (6) and the principles in (11) should account for secondary stress in example (13). Welden (1980) counts the syllables from the final or primary stressed syllables towards the left, and therefore claims that the syllable [gim] takes secondary stress because it is the third after the primary stressed syllable (liiš). She then applies the iterative aspect, meaning that counting starts over again after the first secondary stressed.

12. L̀ H̆ H̀ L̆ L̆ X́
    màtărgìmtŭhăliišʃ
    'you (plu) didn't translate it for me'

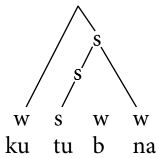
Metrical tree theory

Harms (1981) formulates Welden's (1980) data and segmental rule-based analysis into a metrical phonological tree and counts for secondary stress by positing that the construction of the tree is leftward, as in (13) below.

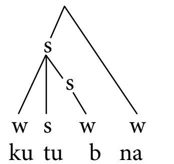
9. For more on this, see Watson (2002, p. 72) and Broselow (1992, p. 10).

13. Backward Metrical Approach

Right Branching

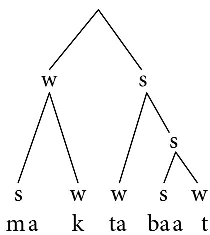


Left Branching

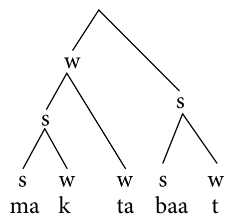


In this formulation, no claims are made regarding syllable internal structure. Heavy syllables and light syllables are grouped as a stress group. According to Harms, primary stress can be described using either tree, i.e., light branching or left branching. In fact, he asserts that a left-branching tree gives a more correct secondary stress placement, as in (14).

14. Right Branching



Left Branching



Backward counting (as in Welden's segmental rule-based theory) and Harms' metrical tree are interesting, but I find them implausible if a language learner is taken into consideration. How could a learner know in advance that the primary stress is on the final syllable and then count backward for secondary stress? Also, as observed, neither formulations make a distinction between the light and heavy syllables in terms of attracting stress; nor do they pay attention to internal syllable weight or bimoracity. Within the analyses, accounting for primary and secondary stress cannot be integrated into a singular schema or formulation, therefore compromising economy and clarity of explanation.

Optimality Theoretic formulation based on constraints ranking, on the other hand, demonstrates a representation of one schema for both primary and secondary stress, a property Welden's (1980) and Harms' (1981) formulations lack. My contribution agrees with (Harrell, 1957), considering word internal heavy syllables as the dock of secondary stress, but also does not differ from Welden's (1980) in terms of the trochaic default stress as an explanation for secondary stress. I agree that the Trochaic principle is at play but disagree that internal heavy syllables are not prominent. Additionally, as a native speaker of CA I dispute the secondary stress on some of the words in Welden's study; primarily, the stress on the first syllable illustrated in (15). I have checked with a number of native speakers of CA and they also confirmed that the initial syllable does not have any prominence.[10] I show subsequently that the initial syllable is not parsed because FOOTBINARITY (FTBIN) is an undominated inviolable constraint while ALIGNFEET/LEFT (AFL) is ranked low and violated in CA.

15. **Welden (1980) data** — **But rather**
    a. màkatabúuʃ — makàtabúuʃ
       'they didn't write'
    b. bìtirmìhumláha — bitìrmihùmláha
       'she is throwing them to her'
    c. màtargimtuhalíiʃ — matàrgimtuhalíiʃ
       'you (plu) didn't translate it (fem) for me'

3. Optimality theoretic formulation

The importance of the mora

The mora is a very important prosodic constituent in CA. Following Hayes's (1995) Moraic Trochee model, a single heavy syllable forms metrical units of moraic weight that is equal to that of two light syllables, as in (16).

16. Moraic Trochee (x .) (x)[11]
    ˘ ˘ or –

The mora is also important in the Minimal Word Requirement (Kenstowicz, 1994). Well established phonological analyses in the literature demonstrate that some languages require content words to consist of some minimal size, often two syllables or two moras (Kenstowicz, 1994). In CA, a monosyllabic content word must be superheavy, CVVC or CVCC, because a final consonant does not add to the weight of a syllable, so only superheavy syllables reach the minimum size of two moras; for example, [bonn] 'coffee beans', [diib] 'wolf', and [bint] 'girl'. Also following (Watson, 2002), I posit that feet in CA should be bimoraic, and a degenerate foot (of one mora) must be forbidden.

I also appeal to the Metrical Phonological Theory Grid representation (Liberman, 1975; Liberman & Prince, 1977) where, since each mora is a weight-bearing unit, it gets a beat.

10. This is an intuitive impressionistic inquiry. I read to four native speakers of CA the words in data (16) and asked them to tell me which syllable they hear as the most prominent. Then I read the words again and asked them whether the first syllable is prominent or not.

11. Following Hayes (1995) notations, a light (CV) stressed syllable is / ˘ /. A light (CV) unstressed syllable is a dot /. /, whereas a stressed syllable is /x/ whether it is a heavy syllable (CVC or CVV) / – / or a light (CV) stressed syllable / ˘ /.

```
17.  2    (`     `      *     )  ┐  Prosodic word level
     1    (*)   (*)    (*)       ┤  Foot level
     0    **    **     ** *      ┘  Mora level

          μμ    μμ     μμ  μ
          ʕàsa  lìy    yáa ya
```

The grid demonstrates that, on the mora level, each mora receives a beat. However, on higher levels between every beat and another, there is an element (i.e., a mora) that does not have a higher asterisk on the higher layer. In [ʕàsalìyyáaya] neither the second mora, [a] nor the mora of coda [y] in [lìy], or the mora for [a] in the penult [yáa] have asterisks on higher levels. These moras are represented by dots. Moras are units of weight; therefore, I posit that they constitute the distances between an asterisk and another, i.e., between a stressed syllable and another. Distances are controlled by principles. One of these is the Principle of Rhythmic Alternation (Selkirk, (1984) which stipulates an anti-clash as well as an-anti lapse provision.

18. The Principle of Rhythmic Provision
    a. Anti-clash Provision
       Every strong position[12] on a metrical level *n* should be followed by at least one weak position on that level.
    b. Anti-lapse Provision
       Any weak position on a metrical level *n* may be preceded by at most one weak position on that level.

Based on the Principle of Rhythmic Alternation by Selkirk (1984), I posit that the weak position intervening between two strong positions is a weight-bearing unit – i.e., a mora.

Finally, to secure prosodic hierarchy, I invoke the Weak Layering theory of Ito & Mester, 1992 and propose that the foot dominates both the mora and the syllable in CA, as in structure (19A), contrary to the Strict Layering in (19B).

19. Weak Layering

```
A.        Foot                    B.     Foot
         /    \                           |
        /      \                       Syllable
     Mora    Syllable                     |
                                         Mora
```

By following both the Principle of Rhythmic Alternation and the Weak Layering premise, I claim that the distance between a stressed syllable and another in CA is

12. Strong positions are represented by asterisks and weak positions are represented by dots in the grid.

a mora, as illustrated in the grid, and that the presence of this distance provides the suitable environment for a secondary stress to exist in CA.

Constraints of stress in CA

Primary stress

In OT parlance, the account for stress in CA, as mentioned above, is explained through the interaction and competition between stress-related constraints, e.g., TROCHAIC, WEIGHT TO STRESS PRINCIPLE, and ALIGN-HEAD/RIGHT. For primary stress, the competition is won by the un-dominated constraint of FOOTBINARITY, ALIGN-HEAD/RIGHT followed by TROCHAIC.[13]

Tableau 2. FTBIN>> PARSEσ, PARSE-SG.[14]

/mutaʕalliqaat/ 'belongings'	FTBIN	PARSE-σ	PARSE-SG[15]
☞a. (muta) (ʕal)li(qáa)<t>		*	*
b. (muta) (ʕal) (li) (qáat)	*W	L	L
c. (múta) (ʕal)li(qaat)	**W	*	

13. For space consideration, I present only tableaux (2–10) illustrating the constraints of primary stress in CA and their interaction, however I summarize the text under each tableau in notes below. (See Aquil, 2012 for the detailed analysis).

14. This tableau is a combination tableau adapted from McCarthy (2008, pp. 46–47). According to McCarthy, the tableau ensures that the first two requirements of a valid ranking are met, which are constraint conflict and a winner. The cells with W and L show that these constraints compete over the choice of the winner. For the winner to win the constraint with the W must be higher than the one with the L. According to McCarthy: "The comparative or combination format is best for the ranking problem" (2008, p. 48). McCarthy advocates the combination tableau since it includes violations as well as W and L annotations of the comparative tableau.

15. PARSE SG, specifies that all the segments of a syllable should be parsed. The tableau shows that there is a direct ranking between FTBIN and PARSE-σ. No ranking is evident between PARSE-σ and PARSE SG, as the two are in a stringency relationship, i.e., every violation of PARSE-σ is also a violation of PARSE SG, but the reverse is not the case. In the tableau, candidate (a) is the optimal candidate because it does not violate FTBIN. Candidates (b & c) lose because they violate FTBIN. Candidate (b) satisfies PARSE-σ by parsing all the syllables, and it also satisfies PARSE SG by parsing the last consonant in the word. However, by satisfying these two constraints, candidate (b) violates FTBIN when it parses a degenerate foot that is not of two moras (i.e., [li]) and when it parses the last consonant. By parsing the last consonant, the final syllable becomes trimoraic.

Tableau 3. FTBIN >> TROCHAIC (TR), PARSEσ[16]

/banaa/ ←/banaa<h>/ 'he built it'	FTBIN	TR	PARSEσ
☞a. ba (náa)			*
b. (ba) (náa)	*W	*W	L
c. (banáa)	*W	*W	L
d. (bánaa)	*W	L	L

Tableau 4: FTBIN, ALIGN-HEAD/R>> TROCHAIC (TR).[17]

/ʔistiratijiyya/ 'strategy'	FTBIN	ALIGN-HEAD/R	TR
☞ a. (ʔis) (tira)ti(jíy)ya			
b. (ʔis) (tira)ti(jiy) (yá)	*W		
c.(ʔís) (tira)ti(jiy)ya		*W	
d.(ʔis) (tíra) (tijiy)ya	*W	*W	L

Some CA words of LLL syllable structure have the primary stress placed on the penult rather than the antepenult – e.g., [libísa] 'underwear' or borrowed words such as [gatóo] 'pastry'. In CA, ranking ALIGN FOOT L (AFL) low, while ranking the undominated Footbinary (FTBIN) high accounts for stress placement in the mentioned words. The ranking of these constraints also explains the stress placement on the following words: [nisíi ← nisíi<h>] 'he forgot it' or ɦmi¯íi ← mi¯íi<h>] 'he walked it'.

Consider Tableau (5) for [ʃágara] and Tableau (10) for [nisíi].

16. This tableau demonstrates that FTBIN dominates both PARSE-σ and TROCHAIC. The most optimal candidate is (a) since it obeys FTBIN TROCHAIC and PARSE-σ. Candidate (b) loses because it parses a degenerate foot (i.e., [mi]), whereas candidates (c) and (d) lose because a foot exceeds two moras. Candidate (d) loses in spite of the fact that it follows the default stress pattern of disyllabic words, namely trochaic.

17. The winner in this tableau is (a) because it obeys FTBIN, ALIGN-HEAD/R and TROCHAIC . Other candidates incur crucial violations with the relevant constraints. For example, candidate (b) does not obey FTBIN, because it contains a degenerate foot [yá]. Candidate (c) on the other hand, loses because it violates ALIGN-HEAD/R by having the stress fall on the first syllable [ʔís] and not on one of the three rightmost ones. Candidate (d) loses because it violates both FTBIN and ALIGN-HEAD/R, even though it satisfies TROCHAIC.

Tableau 5. ALIGN-FOOT-L (AFL) >> ALIGN-FOOT-R (AFR).[18]

/šagara/ 'tree'	FTBIN	TR	AFL	AFR	PARSEσ
☞a. (šága)ra				*	*
b. (šága) (ra)	*W		*W	L	L
c. (ša) (gára)	*W		*W	L	L
d. ša(gára)			*W	L	*
e. ša(gará)		*W	*W	L	*

Tableau 6. PARSEσ >> AFL.[19]

ʕàsaliyyáaya 'kind of candy'	PARSEσ	AFL
☞ a.(ʕàsa) (liy) (yáa)ya	*	**
b. (ʕàsa) liyyaaya	***W	L

Tableau 7. WSP >> Parseσ.[20]

/ʔiftitaħiyya/ 'inauguration'	WSP	PARSEσ
☞ a. (ʔif) (tita) (ħíy)ya		*
b. (ʔif) (títa) (ħiy) (ya)	*W	L

18. The Tableau provides evidence that foot construction in CA is aligned at the left edge of the word. Candidate (a) wins, as it obeys high-ranked constraints namely, FTBIN, TROACHIAC, and it does not violate AFL. Other candidates (b, c, and e) violate AFL and the highly ranked FTBIN and TROACHIAC constraints. Candidate (d) does not obey AFL. It constructs the foot on the right and has the stress on the left syllable of the final disyllabic syllable.

19. Candidate (a) is the optimal one because it minimally violates PARSE- σ. Candidate (b) loses to candidate (a) because, by satisfying AFL and aligning all feet to the left edge of the word, three violations of PARSE σ occur.

20. The optimal candidate is (a) because it does not violate WSP, whereas candidate (b) does. Candidate (a) obeys WSP while minimally violating PARSE σ, a low-ranked constraint. Although candidate (b) fulfills PARSE-σ by parsing all syllables of the word, it loses because it does not obey WSP; stress is assigned to a light syllable (i.e., [títa]).

Tableau 8. Align-head/r >> Wsp.[21]

/ʔiddalʕadi/ ‘term used by women in reference to men as dudes’	Align-head/r	Wsp
☞ a. (ʔid) (dal) (ʕádi)		*
b. (ʔid) (dál) (ʕadi)	*W	L

Tableau 9. Ftbin, Align-head/r>> Trochaic (Tr).[22]

/ʔistiratijiyya/ ‘strategy’	Ftbin	Align-head/r	Tr
☞ a. (ʔis) (tira)ti(jíy)ya			
b. (ʔis) (tira)ti(jiy) (yá)	*W		
c. (ʔís) (tira)ti(jiy)ya		*W	
d. (ʔis) (tíra) (tijiy)ya	*W	*W	L

Secondary stress

As for secondary stress in CA, I explain it through the interaction between primary and secondary stress-related constraints, namely, Trochaic, Wsp, and Align-head/right interacting with two other stress-related constraints mentioned in the literature, *Lapse and *Clash.

20. Secondary stress constraints
    *Clash
    Avoid stress on adjacent syllables, (Prince & Smolensky, 1993& 2004).
    *Lapse
    Prohibits stress lapse, (Prince & Smolensky, 1993& 2004).

21. Candidate (a) wins, although it does not fulfill Wsp, since stress falls on a light syllable (i.e.,[ʕádi] instead of the preceding antepenultimate heavy syllable [dal]. Candidate (b) obeys Wsp, but violates a higher-ranked constraint, Align-Head/R, and hence loses to the winner (a). The tableau illustrates the domination of Align-Head/R over Wsp.

22. The winner is (a) because it obeys Ftbin, Align-Head/R and Trochaic . Other candidates incur crucial violations with the relevant constraints. For example, candidate (b) does not obey Ftbin, because it contains a degenerate foot [yá]. Candidate (c), on the other hand, loses because it violates Align-Head/R by having the stress fall on the first syllable [ʔís] and not on one of the three rightmost ones. Candidate (d) loses because it violates both FtBin and Align-Head/R, even though it satisfies Trochaic. Based on the principles of combination tableaux, both Ftbin and Align-Head/R dominate Trochaic because there are two (Ws) on the left of the (L) which is in the Trochaic column of candidate (d).

Tableau 10. Primary Stress constraint hierarchy.[23]

FTBIN, ALIGN-HEAD/R >>TR >>WSP >> PARSEσ,PARSE SG >>AFL

Optimal forms	FTBIN	ALIGN-HEAD/R	TR	WSP	PARSEσ	PARSE-SG	AFL
☞ (bàra) (níi)< tʕ>						*	
☞ mu(sal) (sála)				*	*		**
☞ (fíhi)<m>						*	
☞ (ʔis) (tira)ti(jíy)ya					**	**	***
☞ (ʔid) (dal) (ʕádi)				*			***
☞ (ʔif) (tita) (ħíy)ya						*	***
☞ (šága)ra					*		
☞ sa(bah) (lála)					*		****
☞ ni (síi)					*		*
☞ (mùta) (ʕal)li(qáa)<t>					*	*	**

To capture the importance of the mora as a weight unit in CA, I parameterize the *LAPSE constraint to be more sensitive to the mora than the syllable, hence, *LAPSEμ.

21. Lapseμ
    Prohibits stress lapse, (Prince & Smolensky, 1993& 2004)
22. *Lapseμ >> *Clash

Note the low ranking of *CLASH in tableaux (11) to (15). Tableau (11) demonstrates the domination of *LAPSEμ over *CLASH. Note the low ranking of *CLASH in Tableaux 11 to 15. Tableau 11 demonstrates the domination of *LAPSEμ over *CLASH.

Candidate (a) in Tableau (11) wins even though it violates *Clash, as there are two stressed syllables adjacent to each other (i.e. (lìy) (yáa)). On the other hand, when candidate (b) satisfies *CLASH by un-stressing one of the syllables, it loses because it violates *LAPSE μ.

23. Trochaic >>*Clash
    *CLASH is ranked low in relation with TROCHAIC. Consider Tableau (12), which demonstrates Trochaic dominating *Clash.

---

23. In summary, a direct ranking is found between the ALIGN-HEAD/R and TROCHAIC constraints and between WSP and PARSE-σ. But the ranking relationship between ALIGN-HEAD/R and AFL still needs to be established. I propose that ALIGN-HEAD/R dominates AFL transitively. I propose the same for the interaction between WSP and AFL by means of transitivity.

Tableau 11. *Lapseμ >> *Clash.

ʕàsaliyyáaya 'kind of candy'	*Lapse μ	*Clash
☞ a. (ʕàsa) (lìy) (yáa)ya \`μμ \`μμ ´μμ		*
b. (ʕàsa) (liy) (yáa)ya \`μμ \`μμ ´μμ	*W	L

Tableau 12. TR>>*Clash.

/sabahlala/ 'haphazardly'	Tr	*Clash
☞ a. sa(bàh)´(lála) \`σ ´σσ		*
b. sa(bàh) ´(lalá) \`σ σ ´σ	*W	L

Here Trochaic and *Clash are in direct competition and ranking. Candidate (a) wins even though it violates *Clash by having two stressed syllables adjacent to each other (e.g., bàhlála). However, when candidate (b) rectifies this violation by distancing one of the (LL) stresses, it crucially violates a higher-ranked constraint, Trochaic.

As for interaction between Wsp and *Clash, we observe that Wsp also dominates *Clash.

Tableau (13) confirms the domination of Wsp over *clash. Wsp prefers candidate (a), which is the winner. By comparison, candidate (b) loses because it violates Wsp. Candidate (b) obeys *clash by imposing a distance between the two adjacent stressed syllables; only the first syllable (ʔìd) is stressed while the antepenult does not have a stress, although it is heavy.

Tableau 13. Wsp >>*Clash.

/ʔiddalʕadi/ 'term used by women in reference to men as dudes'	Wsp	*Clash
☞a. (ʔìd) (dàl) (ʕádi) \`σ \` σ ´σ σ		**
b. (ʔìd) (dal) (ʕádi) \`σ σ ´σ σ	*W	L

Tableau 14. Secondary stress.
FTBIN, ALIGN-HEAD/R>>TR>>WSP>>*LAPSEμ>>*CLASH

Optimal forms **Rule based vs. OT constraints**	**Ftbin**	**ALIGN-HEAD/R**	**TR**	**WSP**	***LAPSEμ**	***CLASH**
☞a. makàtabùu<ʃ>						
b. màkatabùu<ʃ>	*		*			
☞ a. matàrgìmtùhalíi<ʃ>						***
b. màtargìmtuhalíi<ʃ>	*		*	*		
☞a. katàbtáha						*
b. kàtabtáha	*			*		
☞a. bitìrmihùmláha						*
b. bìtirmìhumláha	**			**		

If we filter data from (Gairdner, 1926; Welden 1980) through the proposed OT constraints we find that the data incurs many violations to inviolable constraints such as FOOTBINARIY, as well as violations to high-ranking constraints like TROCHAIC and WSP. See Tableau (14).

Tableau 15. Summary tableau.[24]
FTBIN, ALIGN-HEAD/R>>TR>>WSP>>PARSEσ, PARSE SG >>AFL; *LAPSEμ >>*CLASH

Optimal forms	**Ftbin**	**ALIGN HEAD/R**	**TR**	**WSP**	**PARSEσ**	**AFL**	***LAPSEμ**	***CLASH**
☞ (bàra) (níi)<tˤ>								
☞ mu(sàl) (sála)					*	**		*
☞ (fíhi)<m>								
☞ (ʔìs) (tìra)ti(jíy)ya					**	***		*
☞ (ʔìd) (dàl) (ʕádi)						***		**
☞ (ʔìf) (tìta) (ħíy)ya						***		**
☞ li(bísa)					*	*		
☞ (šága)ra					*			
☞ sa(bàh) (lála)					*	****		*
☞ ni (síi)					*	*		
☞ (mùta) (ʕàl) li(qáa)<t>					*	**		*
☞ ma(kàta)búuʃ					*	****		
☞ bi(tìr)mi(hùm) (láha)					**	**** **		

24. Since I have not yet come across data in CA which can illustrate the interaction between the following constraints PARSE σ, PARSE SG >>AFL and *LAPSEμ I opted to use the convention of using two lines to separate the constraints.

All in all, by invoking OT constraints of stress, we observe that accounting for secondary stress in CA is not complex and does not propose any challenges to the theory. Moreover, the studies conducted on CA used mixed data (Classical Cairene Arabic), which I find confusing and not representative of CA prosodic information. Tableau (15) demonstrates the constraints relevant for stress placement in CA. Observe that I did not use the constraints and the ranking arrived at in the aforementioned studies (McCarthy, 1984; Welden, 1977) as their findings are irrelevant to the data in this paper.

Hence, the presence of secondary stress in CA can be theoretically motivated. We can motivate the presence of secondary stress through TROCHAIC and WSP constraints acting independently. The role of *LAPSE-μ postulated so far, motivates some rhythmic pattern in CA where the distance between one stressed syllable and another is a weak mora.

3. Conclusion

Theoretically, the existence of secondary stress in CA is well-supported. We account for secondary stress by employing externally motivated high-ranked constraints, TROCHAIC and WSP interacting with *LAPSE μ and *CLASH. However, such a claim needs to be validated phonetically, through acoustically and perceptually based studies. Such studies are warranted in order to prove or disclaim the presence of secondary stress in CA.

References

Aljarah, R. (2008). Cairene arabic word stress: A Constraint-based analysis. *Dirasat, Human and Social Sciences*, 35(3), 1–13.

Allen, W. S. (1969). The Latin accent. *Journal of Linguistics*, 5, 193–203.

Aquil, R. (2012). Optimality theoretic representation of stress in Cairene Arabic. *Open Journal of Modern Linguistics (OJML)*, 2(3), 85-89.

Broselow, E. (1992). Parametric variation in Arabic dialect phonology. In E. Broselow, M. Eid, & J. McCarthy (Eds.).

Chomsky, N., & Halle, M. (1986, 1991). *The sound pattern of English*. Cambridge, MA: The MIT Press.

De Lacy, P. (1998). Sympathetic stress. Retrieved from roa.rutgers.edu/files/294-0199/294-0199-LACY-0-0.pdf

Gairdner, W. H. T. (1926). *Egyptian colloquial Arabic: A conversation grammar*. (2nd edition ed.). London: Oxford University Press.

Halle, M., & Vergnaud, J. R. (1987). *An Essay on Stress*. Cambridge, Massachusetts: The MIT Press.

Harms, R. (1981). A backwards metrical approach to Cairo Arabic stress. *Linguistic Analysis*, 7(4), 429–451.
Harrell, R. S. (1957). *The phonology of colloquial Egyptian Arabic*. New York: American Council of Learned Societies.
Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.
Kenstowicz, M. (1994). *Phonology in generative grammar*. Cambridge, MA: Blackwell.
Kiparsky, P. (2003). Syllables and moras in Arabic. In C. Fery, & R. V. D. Vijver (Eds.), *The syllable in Optimality Theory*. New York, NY: Cambridge University Press.
Liberman, A. M. (1975). *The intonational system of English*. Ph.D., MIT, Cambridge: MA.
Liberman, A. M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8, 249–336.
McCarthy, J. (1979). On stress and syllabification. *Linguistic Inquiry*, 10, 443–466.
McCarthy, J. (1984). [Personal Communication].
McCarthy, J. (2005). The length of stem-final vowels in colloquial Arabic. In M. T. Alhawary, & E. Benmamoun (Eds.), *Perspectives on Arabic Linguistics (*Vol. *XVII-XVIII* (pp. 1–26). John Benjamins Publishing Company.
McCarthy, J. (2008) Doing optimality theory: applying theory to data. Malden, MA: Blackwell Publishing.
Mitchell, T. F. (1956). *An introduction to Egyptian colloquial Arabic* New York: Oxford University Press.
Mitchell, T. F. (1960). Prominence and syllabification in Arabic. *Bulletin of the School of Oriental and African Studies*, 23, 369–389.
Prince, A., & Smolensky, P. (1993). *Optimality theory: Constraint interaction in generative grammar*. University of Colorado (CU-CS-696-93). Boulder, Colorado.
Prince, A., & Smolensky, P. (2004). *Optimality Theory: Constraint interaction in generative grammar* (Revised edition ed.). Malden, MA: Blackwell.
Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, Massachusetts: The MIT Press.
Vaux, B. (2004, February 20, 2004). *The appendix*. Paper presented at the Symposium on phonological theory: Representations and architecture, CUNY.
Watson, J. C. E. (2002). *The phonology and morphology of Arabic*. Oxford: Oxford University Press.
Welden, A. (1977). *Prosodic aspects of Cairo Arabic phonology*. Dissertation. Linguistics. The University of Austin, Texas. Austin, Texas.
Welden, A. (1980). Stress in Cairo Arabic. *Studies in the Linguistic Sciences*, 10(2), 99–120.

Paradoxical paradigms! Evidence from Lebanese Arabic phonology

Youssef A. Haddad and Caroline Wiltshire

Examinations of Optimal Paradigms (McCarthy, 2005) suggest that only phonological outputs of inflectional paradigms obey Optimal Paradigm constraints, whereby each member of a given paradigm strives to resemble the phonological structure of each other member of that paradigm. Words that are derived from "base" words, on the other hand, are claimed to obey Base-Output constraints which require all members of the paradigm to conform to the structure of one member selected as a Base. In this study, we examine Lebanese Arabic paradigms of sound regular, sound geminate, weak hollow, and weak defective verbs. Lebanese Arabic verbal paradigms show that while Base-Output constraints do play a role in determining the phonological structure of related words derived from a base, Optimal Paradigm constraints also participate in such morphology. The outcome is paradoxical paradigms in which members strive to resemble a Base while at the same time trying to resemble each other phonologically.

Keywords: Optimal Paradigms constraints, base-output constraints, Lebanese Arabic.

Introduction

McCarthy (2005), building on Benua's (1997) Transderivational Correspondence Theory and on Kenstowicz's (1996) Uniform Exponence, puts forth a theory of Optimal Paradigms (OP) in which he argues that output-output correspondence is a necessary part of Optimality Theory (Prince & Smolensky, 1993, 2004). More specifically, McCarthy argues that individual members in inflectional paradigms take into account the phonological structure of all paradigm members and that each member strives to conform to the structure of the majority even if this means that the phonological form of some members will end up being opaque and/or violating some input-output constraints. For example, the English word *lightening* is pronounced with a syllabic *n* – and not like *lightning* – in conformity with the co-members of its paradigm, namely *lighten, lightens,* and *lightened* (McCarthy, 2005, p. 170).

While illustrated with examples from inflectional paradigms, the above output-output constraint is claimed not to apply to words derived from a base, which are called "derivational paradigms" (McCarthy, 2005, p. 174). Derivational paradigms seem to work differently in that the relation among the members is asymmetric: All members conform to the structure of one member selected as a base, while the base need not conform to the structure of any other member. To illustrate, in Palestinian Arabic, a high vowel that would appear as unstressed in a non-final syllable normally undergoes syncope. For example, /fihimna/ 'we understood', a member of the inflectional paradigm of the verb /fihim/ 'he understood', surfaces as [f<i>.'him.na].[1] Surprisingly, a segmentally identical input with a different morphological structure, /fihim-na/ 'he understood us', is realized as [fi.'him.na]; that is, without syncope. The latter is a member of the derivational – rather than the inflectional – paradigm of /fihim/ 'he understood'. Several researchers argue that this is the case because members of such derivational paradigms obey a base-identity constraint that requires all members to resemble a morphologically related base (e.g. Kenstowicz, 1996; Kager, 1999). In the case of [fi.'him.na] 'he understood us', no syncope takes place in order for the verb to conform to the structure of the base ['fi.him] 'he understood'. Kager (1999, p. 216 [15]) calls this constraint Head-Max (B/O):

(1) Head-Max (B/O)
Every segment in the base prosodic head has a correspondent in the output.

Note that [f<i>.'him.na] 'we understood', which is also morphologically related to ['fi.him], does not have to obey this constraint because it is a member of the inflectional rather than the derivational paradigm of ['fi.him]. The relation among the members of an inflectional paradigm is symmetrical: There is no base; every member tries to resemble the structure of the majority.

The focus of this paper is on Lebanese Arabic (LA) verbs that take on dative and accusative pronominal clitics.[2] Accusative and dative clitics are level-two morphemes introduced at the "word level" of lexical phonology (Kiparsky, 2002).[3] Thus, these clitics are attached to a word that counts as a base.

1. The angled brackets indicate that the segment is not pronounced in the surface form.

2. Unless otherwise specified, the LA data come from the variety of suburban Beirut of which one of the authors is a native speaker.

3. Subject agreement is also level-two morphology, but it is inserted at the stem level rather than at the word level of lexical phonology.

The paradigms under examination seem to obey the aforementioned base-output constraint Head-Max (B/O) with respect to syncope; that is, all members of the paradigm strive to be structurally similar to a base. However, we witness an unexpected case of majority rule with regard to the assignment of stress. In this case, the base form does not determine the output, and an Optimal Paradigm constraint that we will call OP-Iden-Stress is obeyed. As (2) indicates, the constraint requires each member of the paradigm to resemble the majority of the members with respect to stress assignment.

(2) OP-Iden-Stress
Each member of a paradigm has stress on the same syllable as each other member of that paradigm.

We will consider four types of verbs: sound regular, sound geminate, weak hollow, and weak defective. Sound verbs are based on triconsonantal roots $C_1C_2C_3$, none of which is a semi-vowel /w/ or /j/. Sound regular (hereafter sound) verbs are based on roots in which C_2 and C_3 are distinct; [kasar] 'he broke' is an example, as it is based on the root K-S-R. Sound geminate (hereafter geminate) verbs are based on roots in which C_2 and C_3 are identical; for example, [madd] 'he stretched' is based on the root M-D-D.

Weak verbs are based on triconsonantal roots, one consonant of which is a semi-vowel /w/ or /j/. Weak hollow (hereafter hollow) verb roots have a semi-vowel as C_2, for example N-W-M. The semi-vowel is not realized in the output of pattern I verbs; thus, /nawam/ – or /newem/ as it would probably be pronounced if realized faithfully in LA – surfaces as [ne:m] 'he slept'. Weak defective (hereafter defective) verb roots have a semi-vowel as C_3, for example Ħ-K-J. Again, the semi-vowel is not part of the output of the verbs in question, resulting in [ħike:] or [ħaka:] 'he spoke/said'. (For a detailed overview of the different types of verbs in Standard Arabic see Ryding, 2008, chapter 22 or Alhawary, 2011, chapter 13).[4] Since the location of stress will be involved in the analysis, we briefly introduce the generalizations here. Stress falls on the ultimate syllable if superheavy, on the penultimate if heavy, and otherwise on the antepenultimate, thereby landing two or three moras from the right edge of the word. While

4. There are other types of verbs that we do not discuss separately because their behavior is identical to the behavior of one or more of the types we discuss here. These are hamzated, assimilated, and doubly weak verbs. Hamzated verbs are based on roots that contain a glottal stop as one of its consonants. Weak assimilated verbs are based on roots that contain a semi-vowel as C_1. Neither the glottal stop nor a semi-vowel as C_1 adds an extra dimension to the behavior of the paradigm of verbs under examination. Doubly weak verbs are based on roots that contain two semi-vowels; they can be assimilated defective or hollow defective. Both types may be treated as otherwise defective verbs for the purposes of this paper.

these generalizations could clearly be stated in terms of competing constraints in OT, for ease of presentation we will use a single cover constraint we will call StressLocation.

(3) StressLocation:
Stress falls on the ultimate syllable if superheavy, on the penultimate if heavy, and otherwise on the antepenultimate.

In the following section we will first introduce the derivational paradigms of verbs with dative clitics followed by a presentation of the derivational paradigms of verbs with accusative clitics. These analyses show that these derivational paradigms strive to obey not only the base-output constraint Head-Max (B/O), as expected, but also the OP constraint OP-Iden-Stress, which is unexpected based on cross-linguistic investigations (e.g., McCarthy, 2005). Following these analyses, we present the analysis of sound verbs plus accusative clitics and attempt to provide a possible explanation to its unusual paradigm structure.

2. Verbs plus dative clitics

We start with the paradigm of geminate 3rd Sg Mas perfective verbs because it presents most clearly the satisfaction of both the OP constraint OP-Iden-Stress as well as the base-output constraint Head-Max (B/O). We next show that the sound and defective paradigms of perfective 3rd Sg Mas pattern I verbs satisfy OP-Iden-Stress vacuously, after which we generalize over verbs of different aspects (e.g., imperfective), agreement (e.g., 1st Sg), and patterns (e.g., pattern IX). The final section focuses on hollow verbs. These also satisfy OP-Iden-Stress; however, they do so indirectly.

Geminate verbs with dative clitics

Table (4) presents the paradigm of the verb 'to return (sth)' with dative pronominal clitics. The verb may take two surface forms in different paradigms: the degeminated form [rad] or the geminate [radd]. The former is usually realized in a pre-pausal position or if the verb is followed by CV, where the CV can be the onset of a new word or of a suffix; the latter is usually realized pre-vocalically, where the vowel can be in the input or epenthetic.

(4) Geminate verb *radd* 'he returned (sth)' + dative clitics

Input	Optimal Output	'returned sth'	Sub-optimal Output
a. /radd-l-i/	rad.'d-al.li	'~ to me'	* 'rad.d-a.li, * 'rad.-li
b. /radd-l-na/	rad.'d-al.na	'~ to us'	
c. /radd-l-ak/	rad.'d-al.lak	'~ to you (M)'	* 'rad.d-a.lak, * 'rad.lak
d. /radd-l-ik/	rad.'d-al.lik	'~ to you (F)'	* 'rad.d-a.lik, * 'rad.lik
e. /radd-l-kun/	rad.'d-al.kun	'~ to you all'	
f. /radd-l-o/	rad.'d-al.lo	'~ to him'	* 'rad.d-a.lo, * 'rad.lo
g. /radd-l-a/	rad.'d-al.la	'~ to her'	* 'rad.d-a.la, * 'rad.la
h. /radd-l-un/	rad.'d-al.lun	'~ to them'	* 'rad.d-a.lun, * 'rad.lun

The analysis reveals that all the members of the paradigm obey HEAD-MAX (B/O), as they match the base form [radd]. It should be noted, though, that the paradigm also strives, in an overkill fashion, to satisfy the OP constraint OP-IDEN-STRESS.

The relevant members of the paradigm are (4a), (4c), (4d), (4f), (4g), and (4h), in which the dative marker /-l-/ is realized as [–ll-]. To justify the input form of the dative marker as a single /-l/ rather than already being /-ll-/, we illustrate in (5a) below that the dative marker is normally realized faithfully as [–la-], with no consonantal epenthesis, in the case of a sound verb. Note that if the dative marker were underlyingly /-ll-/ instead of /-l-/, the surface form in (5a) would be *[kataballa] instead of [katabla]. In (5b), an epenthetic [a] is used to break the cluster /CCC/, while in (5c) we see that the epenthetic vowel may be [i], as in the imperative form. In both these cases, the /-la/ appears as geminated [–lla].[5]

(5) a. /katab/ + /-la/ → [katabla] 'he wrote to her'
b. /radd/ + /-la/ → [raddalla] 'he returned (sth) to her'
c. /ridd/ + /-la/ → [riddilla] '(you) return (sth) to her'

Second, we should justify that the final consonant of the verb /radd-/ is already a geminate, and is not being doubled, since the doubling of a consonant as a result of level-two morphology is not uncommon in Arabic, both standard and colloquial. For example, observe the LA paradigm of the preposition /min/ 'of/from' with the accusative clitics in the table in (6). The /n/ is doubled before clitics that begin with a vowel in (6a), (6c), (6d), (6f), (6g), and (6h), but no doubling takes place when the clitic begins with a consonant in (6b) and (6e). The reason is that level-two morphemes have a minimal-stem restriction; they cliticize to stems that

5. The epenthetic vowel matches the preceding vowel. In (5b), the epenthetic vowel is [a] because the preceding vowel is [a], while in (5c), the epenthetic vowel is [i] to match the preceding vowel, which is also [i].

are at least bimoraic (Watson, 2002, pp. 205–206). By realizing /n/ as [nn], /min-i/ may be realized as [minni]. This means that it is syllabified as [min.ni], a bimoraic stem + a clitic, instead of [mi.ni], which would be a mono-moraic stem + a clitic. This syllabification reflects the system's use of gemination as a repair strategy to satisfy the minimal-stem restriction, since the final /-n/ of the stem would otherwise become an onset without a mora.

(6) Preposition *min* 'of/from' + pronominal clitic

Input	Optimal Output	'from'	Sub-optimal Output
a. /min-i/	'min-n-i	'~ me'	* 'min-i
b. /min-na/	'min-na	'~ us'	
c. /min-ak/	'min-n-ak	'~ you (M)'	* 'min-ak
d. /min-ik/	'min-n-ik	'~ you (F)'	* 'min-ik
e. /min-kun/	'min-kun	'~ you all'	
f. /min-o/	'min-n-o	'~ him'	* 'min-o
g. /min-a/	'min-n-a	'~ her'	* 'min-a
h. /min-un/	'min-n-un	'~ them'	* 'min-un

It might be argued that the preposition 'of/from' is underlyingly /minn/ and that it undergoes degemination before clitics that begin with a consonant. Evidence that this is not the case comes from cases of juncture with a following word that begins with an (epenthetic) vowel. For example, /min + l-madrase/ 'from the school' is realized is [mi.n il.mad.ra.se] with an epenthetic vowel rather than *[min.nil.mad.ra.se]. Compare this to words with a true final geminate, like /ʔimm/ 'mother'. When used in a juncture position followed by a vowel, /ʔimm/ is realized as [ʔimm]; for example, /ʔimm + l-walad/ 'the mother of the child' surfaces as [ʔim.m il-wa.lad]. In both cases, the vowel is epenthetic to save an otherwise illegal onset/coda.

We maintain that providing a bimoraic stem for a level-two morpheme is not why /radd/ has a doubled final consonant; instead, the geminate is present in the input. Nor is it the motivation for the doubling of the dative marker in the examples in Table 1. As illustrated in (7a), the stem *radd* plus the dative clitic *-la* may not be realized faithfully due to constraints that prohibit complex codas and non-initial complex onsets. These constraints, however, may be satisfied in at least two ways: vowel epenthesis, (7b), and degemination, (7c), both of which are less than optimal. In (7b), vowel epenthesis breaks the consonant cluster, resulting in a three-syllable word. In (7c), degemination reduces the consonant cluster CCC into

CC, with one consonant being syllabified in the coda of the first syllable and the other in the onset of the second syllable.[6]

(7) a. /radd/ + /-la/ → *[rad.dla] OR *[radd.la]
b. /radd/ + /-la/ → *['rad.d-a.-la]
c. /radd/ + /-la/ → *['rad.-la]

Although vowel epenthesis will prove necessary in the optimal form, it is not sufficient; neither (7b) nor (7c) is considered acceptable. Rather, an apparently less than optimal output with both an epenthetic [a] and a doubled [l] surfaces as the grammatical form, [rad.'dal.la]. Two OT tableaux will perhaps make the overkill clearer. The tableaux will employ the following additional constraints:

(8) a. *COMPLEX – No consonant clusters in onsets/codas.
b. DEP IO (V) – No vowel insertion. Vowels in the output correspond to vowels in the input.
c. DEP IO (C) – No consonant insertion. Consonants in the output correspond to consonants in the input.
d. REALIZEMORPH – No morpheme deletion. All morphemes in the input must have some exponence in the output.

As the tableau in (9) shows, forms with a C-initial suffix behave unremarkably, and are stressed on the penultimate syllable:

(9) Geminate verb *radd* + dative clitic without overkill

Input /radd-l-na/ base prosodic head=[radd]	STRESS LOC	*COMPLEX	HEAD-MAX BO	REALIZE MORPH	DEPIO (V)
a. ☞ [rad.'dal.na]					*!
b. ['radd.lna]		*!			
c. ['rad.lna]		*!	*!		
d. ['rad.da]				*!	

The tableau in (9) illustrates an example with the minimal necessary changes (one Dep-IO(V) violation in the winning candidate (9a)) in order to satisfy higher ranked constraints. Candidates (9b) and (9c) lose because the sequence of consonants cannot be syllabified without violating the constraint *COMPLEX, which forbids having two consonants in onset or coda. Candidate (9d) solves that problem

6. Degemination of the type proposed in (7c) is common in many Arabic dialects (see, for example, Watson, 2002, pp. 210–211). For example, when /ʔimm/ 'mother' takes on a clitic that begins with a consonant, it undergoes degemination: /ʔimm + na/ 'our mother' surfaces as [ʔim<m>-na]. Compare to /ʔimm + ak/ 'your (Mas) mother' that is realized as [ʔimm-ak].

by deleting two consonants, including deleting the dative morpheme altogether, resulting in a REALIZEMORPH violation. With DEP-IO(V) ranking lower than these constraints, candidate (9a) wins.

Tableau (10) shows the same ranking, with the added low ranked DEP-IO(C), when applied to the input /radd-l-ik/. Note that the actual output form, (10a), violates both DEP-IO(V) and DEP-IO(C), although, as shown in (10b), it is possible for a candidate to satisfy the higher ranked constraints while violating only DEP-IO(V).

(10) Geminate verb *radd* + dative clitic with overkill

Input /radd-l-i/ base prosodic head=[radd]	STRESS LOC	*COM PLEX	HEAD-MAX BO	REALIZE MORPH	DE-PIO (V)	DE-PIO (C)
a. ☹ [rad.'dal.li]					*	*!
b. ☞ ['rad.da.li]					*	
c. [rad.'da.li]	*!				*	
d. ['rad.dli]/['radd.li]		*!				
e. ['rad.li]			*!			
f. ['rad.di]				*!		

Candidate (10c) is identical to candidate (10b) except for the location of stress, which in (10c) falls on the light penultimate syllable and violates STRESSLOCATION. Candidates (10d), (10e), and (10f) have the same kinds of violations as seen in the previous tableau in candidates (9b-d), and similarly cannot compete against the top candidates. Thus, the actual output (10a), indicated with a ☹, is harmonically bound, that is, beaten regardless of the constraint ranking, by the more conservative candidate in (10b), indicated with ☞. The question is: Why?

Closer examination of the paradigm in table (4) above shows that the overapplication of consonant insertion in the case of all but two members of the paradigm allows all the members to be realized with stress on the same syllable. That is, the overkill does not satisfy a base-output constraint or a single phonotactic constraint. Rather, it satisfies OP-IDEN-STRESS, whereby each member of the paradigm strives to be similar to every other member of the paradigm, and the outcome is a uniform paradigm in terms of the location of stress. The optimal form has stress on the heavy penultimate syllable [dal], while without the DEP-C violation, the penultimate syllable would be light and the first syllable would get stress. If an OP-IDEN-STRESS constraint is higher ranked than the DEP constraints limiting epenthesis, and if it is satisfied by the real winner and not satisfied by the more conservative candidate, then we have an explanation for the overkill, as shown in the tableau in (11).[7]

7. The violation of STRESSLOCATION in (11) will not be the whole story, as we will elaborate upon later.

(11) Optimal Paradigm effect on geminate verb *radd* + dative clitic (a snapshot)

Input /radd-l-i/ base prosodic head=[radd]	OPIden Stress	Stress Loc	*Comp	Head-Max BO	Realize Morph	DepIO (V)	DepIO (C)
a. ☞ [rad.'dal.li]						*	*
b. ['rad.da.li]	*!					*	
c. [rad.'da.li]		*!				*	
d. ['rad. dli]/['radd.li]	*!		*!				
e. ['rad.li]	*!			*!			
f. ['rad.di]	*!				*!		

Tableau (12) shows a fuller evaluation of the whole paradigm. In the winning set of candidates, (12a), OP-Iden-Stress is satisfied, since all members of the paradigm have the same location of stress while also satisfying the constraint Stress-Loc. However, if only vowel epenthesis without *l*-doubling applies, the paradigm will not be uniform with regard to stress, which makes it less than optimal, as (12b) illustrates with two members stressed on the penultimate syllable and six on the initial syllable. In (12c), stress will fall uniformly on the first syllable throughout the paradigm, but for two members of the paradigm (['rad.dal.na] and ['rad.dal.kun]), this location violates StressLoc because the penultimate syllables are heavy yet not stressed. Resolving that problem by deleting the dative suffix fatally runs afoul of the RealizeMorph constraint, in (12d).

(12) Optimal Paradigm of geminate verb *radd* + dative clitic (complete)

Input /radd+ l/+{i, na, ak, ik, kun, o, a, un} base prosodic head=[radd]	OPIden Stress	Stress Loc	Realze Morph	Head-Max BO	Dep IO(V)	Dep IO(C)
a. ☞ {rad.'dal.li, rad.'dal.na, rad.'dal. lak, rad.'dal.lik, rad.'dal.kun, rad.'dal.lo, rad.'dal.la, rad.'dal. lun}					**** ****	**** **
b. {'rad.da.li, rad.'dal.na, 'rad.da.lak, 'rad.da.lik, rad.'dal.kun, 'rad.da. lo, 'rad.da.la, 'rad.da.lun}	**!				**** ****	

c.	{'rad.da.li, 'rad.dal.na, 'rad.da.lak, 'rad.da.lik, 'rad.dal.kun, 'rad.da.lo, 'rad.da.la, 'rad.da.lun}		*!*			**** ****	
d.	{'rad.da.li, 'rad.da.na, 'rad.da.lak, 'rad.da.lik, 'rad.da.kun, 'rad.da.lo, 'rad.da.la, 'rad.da.lun}			*!*		**** ****	

Any attempt to have stress fall uniformly on the initial syllable results in problems for the two members of the paradigm whose affix begins with a consonant. We mentioned in the introduction that OP is about the pressure for the majority to win. In the case of *radd* 'to return (sth)', the majority does not win. Six members of the paradigm in (4) above experience two violations of Input-Output faithfulness constraints in order to be more like two members: (4b) *rad.'dal.na* 'he returned (sth) to us' and (4e) *rad.'dal.kun* 'he returned (sth) to you all'. While this outcome goes against majority rule, it is justified. The reason is that (4b) and (4e) do not have a way to shift the stress to the first syllable to be more like the more faithful alternatives of the other six members.

To elaborate, consider the underlying form of (4b) as presented in (13), along with the potential outputs in (13a-d). (13a) is not possible because it violates constraints on complex onsets and codas: *[radd.lna]. In (13b) degemination takes place; this is possible if the doubled consonant is followed by a consonant-initial clitic, which is the case here. Degemination does not solve the problem, however, because every possible output would still violate constraints on complex onsets or codas; e.g., *[rad.lna] – *[radl.na]. Finally, the deletion of two consonants in (13c) and (13d) should be able to solve the problem; however, the deletion of the doubled consonant in (13c) means doing away with two of the three consonants of the tri-radical root R-D-D, and the deletion of the dative marker in (13d) means that the clitic will now be confused with its accusative counterpart. This is why (13c) and (13d) are less than optimal.

(13) /radd+ lna/ Actual output: [rad.'dal.na]
  a. Faithful output: *[raddlna]
  b. Degeminated output: *[radlna]
  c. Stem consonant deletion: *['ral.na]
  d. Degemination + Clitic consonant deletion: *['rad.na]

Given that there is nothing that (4b) and (4e) can do to satisfy OP-Iden-Stress by stressing their initial syllables, the other six members in the table in (4) must be the ones to undergo changes in order to match the penultimate stress of the minority two members. Although this results in additional DepC violations for these six members of the paradigm, the satisfaction of paradigm uniformity in stress is achieved at the least overall cost.

Sound and defective verbs plus dative clitics

Both OP-Iden-Stress and Head-Max (B/O) are satisfied in two other paradigms: sound and defective 3rd Sg Mas perfective verbs plus dative clitics. However, satisfying OP-Iden-Stress in these paradigms happens vacuously as it follows from other high-ranking constraints. We begin with sound verbs plus dative clitics. As the paradigm of the verb ‘to hear’ in (14) illustrates, all the members of the paradigm – except the base, which we do not consider as a member of the paradigm as far as stress is concerned – are assigned stress on the same syllable [miʕ].[8]

(14) Sound verb *simiʕ* ‘he heard’[C_1iC_2iC_3] + dative clitics

Input	Optimal Output	‘heard’	Sub-optimal Output(s)
a. /simiʕ-li/	si.ˈmiʕ.-li	‘~ from me’	*s<i>.ˈmiʕ.-li
b. /simiʕ-lna/	si.ˈmi.ʕi-l.na	‘~ from us’	*s<i>.ˈmi.ʕi-l.na, *si.mi.ˈʕi-l.na
c. /simiʕ-lak/	si.ˈmiʕ.-lak	‘~ from you (M)’	*s<i>.ˈmiʕ.-lak
d. /simiʕ-lik/	si.ˈmiʕ.-lik	‘~ from you (F)’	*s<i>.ˈmiʕ.-lik
e. /simiʕ-lkun/	si.ˈmi.ʕi-l.kun	‘~ from you all’	*s<i>.ˈmi.ʕi-l.kun, *si. mi.ˈʕi-l.kun
f. /simiʕ-lo/	si.ˈmiʕ.-lo	‘~ from him’	*s<i>.ˈmiʕ.-lo
g. /simiʕ-la/	si.ˈmiʕ.-la	‘~ from her’	*s<i>.ˈmiʕ.-la
h. /simiʕ-lun/	si.ˈmiʕ.-lun	‘~ from them’	*s<i>.ˈmiʕ.-lun

In two instances, (14b) *si.ˈmi.ʕ i-l.na* ‘he heard from us’ and (14e) *si.ˈmi.ʕ i-l.kun* ‘he heard from you all’, the ʕ in [miʕ] is syllabified as the onset of the following syllable, turning [miʕ] into [mi] and, thus, a light syllable. In principle, stress should fall on the heavy syllable [ʕil], contrary to fact. According to Kager (1999, p. 222), the reason is that [i] in [ʕil] is epenthetic, and stress avoids syllables whose nuclei are epenthetic, even if they fall in the right location to be stressed (that is, in a heavy penultimate syllable).

(15) *StressEpenV:
Every vowel in the output prosodic head has a correspondent in the input (named Head-Dep(OI) in Kager, 1999, p. 222 [33])

8. ‘He heard from me’ in (14) means ‘he listened to me/he took my opinion into consideration’.

Note, that in the paradigm of geminate verbs, as illustrated by [radd] 'to return (sth)' in (4), stress falls on syllables with epenthetic nuclei: the *a* and *i* in [rad.'dal.li] 'he returned (sth) to me' and [rid.'dil.li] '(you.Mas) return (sth) to me' respectively are epenthetic. According to Kager (1999, pp. 240–243; see also Kenstowicz & Abdul-Karim, 1980), this is possible only when the preceding syllable is heavy, which is the case in *rad.'dal.li* but not in *si.'mi.ʕ i-l.na*. In other words, to obey the constraint in (15), stress may fall on the preceding syllable, but only if the preceding syllable is not heavy; otherwise, it must remain on the penultimate syllable. Thus, the constraint in (16) prevents stress from falling too far from the right edge of the word; i.e., more than three moras.

(16) *'σHσHσ
Do not stress a heavy antepenultimate syllable when the penultimate syllable is also closed/heavy.

The ranking of the stress constraints reflects the priorities: stress falls on a heavy penultimate syllable by StressLocation, unless that would stress an Epenthetic vowel (*StressEpenV), unless avoidance of an epenthetic vowel results in stress on a heavy antepenultimate syllable followed by a heavy penultimate one.

(17) Relative ranking of stress-related constraints:
*'σHσHσ >>*StressEpenV >> StressLocation

Thus the forms in the table in (14) follow the above regular pattern of stress in LA, as reflected in the above constraints and their ranking. Furthermore, note that the stressed vowel in the base is realized in every member of the paradigm even when it is unstressed and therefore should have been deleted. In this way, the paradigm satisfies Head-Max (B/O).

The analysis of sound verbs applies not only to stems in which the vocalic melody is [i-i] but also to stems in which the vocalic melody is [a-a], as the table in (18) illustrates. Regarding the base-output constraint, and the non-deletion of [a], it is worth noting that unlike Palestinian Arabic, which allows i-syncope but not a-syncope (see Kenstowicz & Abdul-Karim, 1980), the LA dialect analyzed here allows a-syncope, though it is optional. For example, /dafaʃna/ 'we pushed' may surface as [da.'faʃ.na] or [d<a>.'faʃ.na]. Still, the unstressed [a] is preserved in all the members of the derivational paradigm in (18) in order to satisfy Head-Max (B/O).

(18) Sound verb *dafaʃ* 'he pushed' [$C_1aC_2aC_3$] + dative clitics

Input	Optimal Output	'pushed (sth)'	Sub-optimal Output(s)
a. / dafaʃ-li/	da.'faʃ.-li	'~ for me'	*d<a>.'faʃ.-li
b. / dafaʃ-lna/	da.'fa.ʃi-l.na	'~ for us'	*d<a>.'fa. ʃi-l.na, *da.fa.'ʃi-l.na
c. / dafaʃ-lak/	da.'faʃ.-lak	'~ for you (M)'	*d<a>.'faʃ.-lak
d. / dafaʃ-lik/	da.'faʃ.-lik	'~ for you (F)'	*d<a>.'faʃ.-lik
e. / dafaʃ-lkun/	da.'fa.ʃi-l.kun	'~ for you all'	*d<a>.'fa. ʃi-l.kun, *da.fa.'ʃi-l.kun
f. / dafaʃ-lo/	da.'faʃ.-lo	'~ for him'	*d<a>.'faʃ.-lo
g. / dafaʃ-la/	da.'faʃ.-la	'~ for her'	*d<a>.'faʃ.-la
h. / dafaʃ-lun/	da.'faʃ.-lun	'~ for them'	*d<a>.'faʃ.-lun

OP-Iden-Stress is vacuously satisfied in defective verbs as well. These verbs are realized with a final long vowel, although the semi-vowel may surface in verbs with different agreement. For example, the root for the verb 'to speak/say' is Ħ K J. The perfective 3rd Sg Mas pattern I form is [ħike:] or [ħaka:] 'he spoke/said', whereas the 1st Pl counterpart is [ħakajna], [ħkajna], or [ħki:na] 'we spoke/said'. As the paradigm in (19) shows, the syllable preceding the dative clitic is always heavy CV:, sometimes even superheavy CV:C, and thus will always receive stress, satisfying OP-Iden-Stress and resulting in paradigm uniformity. By the same token, the stressed vowel in the base has a correspondent in each member of the paradigm although it is stressed in none. Deletion of the [i] would result in suboptimal outputs.

(19) Defective verb *ħike:* 'he spoke/said' + dative clitics

Input	Optimal Output	'spoke/said/related'
a. /ħike:-li/	ħi.'ke:.-li	'~ to me'
b. /ħike:-lna/	ħi.'ke:-l.na	'~ to us'
c. /ħike:-lak/	ħi.'ke:.-lak	'~ to you (M)'
d. /ħike:-lik/	ħi.'ke:.-lik	'~ to you (F)'
e. /ħike:-lkun/	ħi.'ke:-l.kun	'~ to you all'
f. /ħike:-lo/	ħi.'ke:.-lo	'~ to him'
g. /ħike:-la/	ħi.'ke:.-la	'~ to her'
h. /ħike:-lun/	ħi.'ke:.-lun	'~ to them'

In the following section, we generalize beyond perfective 3rd Sg Mas pattern I verbs.

Other verbs plus dative clitic

In this section, we show that our analysis applies beyond the narrow sliver of verbal morphology presented in the previous sections. Our analysis makes an important prediction: OP constraints are satisfied in an overkill fashion when the dative clitics are added to a paradigm in which the base ends with a superheavy syllable, such as [radd] 'he returned (sth)'. These same constraints are satisfied vacuously elsewhere. This prediction is borne out.

Observe the paradigms of pattern I perfective and imperfective verbs plus subject agreement in the tables in (20) through (22).[9]

(20) Perfective and imperfective geminate verb plus subject agreement

	Perfective		Imperfective
a. He	/radd/	**['radd]**	**['jridd]**
b. She	/raddit/	['rad.dit]	**['tridd]**
c. They	/raddu:/	['rad.du]	['jrid.du]
d. You (M)	/raddt/	**[rad.'de:t] [rad.'dajt]**	**['tridd]**
e. You (F)	/raddti:/	[rad.'daj.ti]	['trid.di]
f. You (PL)	/raddtu:/	[rad.'daj.tu]	['trid.du]
g. I	/raddt/	**[rad.'de:t] [rad.'dajt]**	**['ridd]**
h. We	/raddna:/	[rad.'daj.na]	**['nridd]**

(21) Perfective and imperfective sound verb plus subject agreement

	Perfective		Imperfective
a. He	/simiʕ/	['si.miʕ]	['jis.maʕ]
b. She	/simiʕit/	['sim.ʕit]	['tis.maʕ]
c. They	/simiʕu:/	['sim.ʕu]	['jis.ma. ʕu]
d. You (M)	**/simiʕt/**	**['smiʕt]**	['tis.maʕ]
e. You (F)	/simiʕti:/	['smiʕ.ti]	['tis.ma. ʕi]
f. You (PL)	/simiʕtu:/	['smiʕ.tu]	['tis.ma.ʕu]
g. I	**/simiʕt/**	**['smiʕt]**	['ʔis.maʕ]
h. We	/simiʕna:/	['smiʕ.na]	['nis.maʕ]

9. Members (20d) through (20h) in the perfective column are reanalyzed in dialects to behave like pattern II of defective verbs, rather than pattern I of geminate verbs as expected. This behavior is typical of geminate verbs in dialects (Watson, 2002, pp. 181) and explains the lack of stress uniformity detected here. The alternatives in (20d) and (20g) are available in the Lebanese dialect analyzed here as a result of optional monophthongization, whereby [aj] becomes [e:].

(22) Perfective and imperfective defective verb plus subject agreement

	Perfective		Imperfective
a. He	/ħikij/	[ˈħi.ke]	[ˈjiħ.ke]
b. She	/ħikijit/	[ˈħik.jit] [ˈħi.kit]	[ˈtiħ.ke]
c. They	/ħikiju:/	[ˈħik.ju] [ˈħi.ku]	[ˈjiħ.ku]
d. You (M)	**/ħikijt/**	**[ˈħki:t]**	[ˈtiħ.ke]
e. You (F)	/ħikijti:/	[ˈħki:.ti]	[ˈtiħ.ke]
f. You (PL)	/ħikijtu:/	[ˈħki:.tu]	[ˈtiħ.ku]
g. I	**/ħikijt/**	**[ˈħki:t]**	[ˈʔiħ.ke]
h. We	/ħikijna:/	[ˈħki:.na]	[ˈniħ.ke]

All the paradigm members in bold in (20) through (22) end with a superheavy syllable: CVCC or CV:C. These behave in the same way as [radd] 'to return' in (4) when they take on dative clitics. In the perfective paradigms in (20), (21), and (22), the bold members include (d) and (g) of each paradigm – that is, the members that display 2nd Sg Mas agreement and 1st Sg agreement. The two are homophonous. The [radd] paradigm has member (20a) as an additional case. To illustrate, verb [ħki:t] 'I spoke' in (22g) ends with a superheavy syllable. The paradigm of verb [ħki:t] plus dative clitics in (23) is similar to the paradigm of verb [radd] 'he returned sth' in (4) in that it also involves doubling of the dative marker in order for stress to be uniform among all members of the paradigm.

(23) Defective verb *ħki:.t* 'I spoke' + dative clitics

Optimal Output	'I spoke'
a. ħki:.ˈt-il.li †	'~ for me'
b. ħki:.ˈt-il.na	'~ to us'
c. ħki:.ˈt-il.lak	'~ to you (M)'
d. ħki:.ˈt-il.lik	'~ to you (F)'
e. ħki:.ˈt-il.kun	'~ to you all'
f. ħki:.ˈt-il.lo	'~ to him'
g. ħki:.ˈt-il.la	'~ to her'
h. ħki:.ˈt-il.lun	'~ to them'

† This is an instance of an ethical dative.

Compare the paradigm in (23) with the paradigm of verb [ħikit] 'she spoke' plus dative clitics in (24). The verb [ħikit] does not end with a superheavy syllable. The paradigm receives uniform stress without resorting to the doubling of the dative marker.

(24) Defective verb *ħiki.t* 'she spoke' + dative clitics

Optimal Output	'she spoke'
a. ħi.'kit.li	'~ to me'
b. ħi.'ki.t -il.na	'~ to us'
c. ħi.'kit.lak	'~ to you (M)'
d. ħi.'kit.lik	'~ to you (F)'
e. ħi.'ki.til.kun	'~ to you all'
f. ħi.'kit.lo	'~ to him'
g. ħi.'kit.la	'~ for her'
h. ħi.'kit.lun	'~ to them'

The analysis is not limited to pattern I verbs; consider pattern IX for example. This is the only Arabic pattern whose template has a double consonant at the right edge (i.e., C1C2aC3C3). This pattern is used for colors and bodily defects. For instance, [ʔazraʔ] 'blue' is based on the root Z-R-ʔ. The perfective, pattern IX, 3rd Sg Mas verb meaning 'he turned blue' is [zraʔʔ]. When combined with ethical dative clitics, it behaves like the paradigm of [radd], as (25) illustrates.

(25) Pattern IX verb *zraʔʔ* 'he turned blue' + dative clitics

Optimal Output	'he turned blue'
a. zraʔ.'ʔ-al.li	'~ for me'
b. zraʔ.'ʔ-al.na	'~ for us'
c. zraʔ.'ʔ-al.lak	'~ for you (M)'
d. zraʔ.'ʔ-al.lik	'~ for you (F)'
e. zraʔ.'ʔ-al.kun	'~ for you all'
f. zraʔ.'ʔ-al.lo	'~ for him'
g. zraʔ.'ʔ-al.la	'~ for her'
h. zraʔ.'ʔ-al.lun	'~ for them'

Now we turn to the paradigm of hollow verbs plus dative clitics. Again, we focus on 3rd Sg Mas perfective verbs.

Hollow verbs with dative clitics

Hollow verbs, that is, verbs based on roots with a semi-vowel as C2, satisfy the constraint OP-IDEN-STRESS indirectly by satisfying other constraints. As the paradigm in (26) shows, all the members of the paradigm receive stress on the initial syllable. This is expected in all but two members: (26b) *ʒabilna* 'he brought (sth) to us' and (26e) *ʒabilkun* 'he brought (sth) to you all'. These are realized as ['ʒa.b-il.na] and [ˈʒa.b-il.kun], with stress on the antepenultimate light syllable, rather than as *[ʒa.ˈb-il.na] and *[ʒa.ˈb-il.kun] with stress on the penultimate heavy syllables. This unexpected assignment of stress, however, follows from another constraint; as we mentioned in relation to the paradigms in (14) and (18), stress avoids syllables whose nuclei are epenthetic if these syllables are preceded by a light syllable CV, according to Kager (1999, pp. 240–243; see also Kenstowicz & Abdul-Karim, 1980). The limitation that stress on epenthetic vowels is possible only when the preceding syllable is heavy provides cause to rule out the unacceptable forms *[ʒa.ˈb-il.na] and *[ʒa.ˈb-il.kun], as the [i] in both forms is epenthetic.

(26) Hollow verb *ʒe:b* 'he brought' + dative clitics

Input	Optimal Output	'brought'	Sub-optimal Output(s)
a. /ʒe:b-li/	ˈʒab.li	'~ to me'	
b. /ʒe:b-lna/	ˈʒa.bil.na	'~ to us'	*ʒa.ˈbil.na, *ˈʒe:.bil.na
c. /ʒe:b-lak/	ˈʒab.lak	'~ to you (M)'	
d. /ʒe:b-lik/	ˈʒab.lik	'~ to you (F)'	
e. /ʒe:b -lkun/	ˈʒa.bil.kun	'~ to you all'	*ʒa.ˈbil.kun, *ˈʒe:.bil.kun
f. /ʒe:b-lo/	ˈʒab.lo	'~ to him'	
g. /ʒe:b-la/	ˈʒab.la	'~ to her'	
h. /ʒe:b-lun/	ˈʒab.lun	'~ to them'	

However, note that the paradigm in (26) does not completely conform to the structure of the base. Whereas the base is of the form CV:C, the members of the paradigm surface as CVC. This phenomenon is called Closed Syllable Shortening (Watson, 2002, pp. 66–70) and is attested in several Arabic dialects, such as Cairene Arabic, usually due to a constraint on superheavy syllables in non-final positions. LA, however, allows superheavy syllables of the type CV:C word internally; for example, [ħi.ke:l.kun] 'he said to you all' from Table 5 and [be:b.kun] 'your (Pl) door'. Thus, Closed Syllable Shortening in the paradigm in (26) is unexpected in terms of phonotactic constraints in LA. However, OPIdenStress may offer an answer. A form such as /ʒe:b-lkun/, if realized without shortening, would be forced to have stress on the epenthetic vowel as the initial syllable is heavy. The result [ʒe:ˈbilkun] now does not match the other members of the paradigm in terms of stress location; the optimal output with shortening allows for stress on the first syllable in [ˈʒabilkun], parallel to the location in the outputs of forms such as /ʒe:b -li/ or /ʒe:b -lik/ in the same paradigm. However, these forms also do not surface with a long vowel, though that is phonotactically acceptable; they shorten in order to match their co-members in the paradigm, such as ['ʒabilkun], which resorted to shortening in order to match them in stress location. Thus, the shortening violates the constraint in (27) while satisfying the OP constraint in (28):

(27) MaxIO (μV):
Vocalic moras present in the input are present in the output (no vowel shortening).[10]

(28) OPIdenQuant:
Each member of a paradigm has the same weight (initial) syllable as each other member of that paradigm.

As the tableau in (29) illustrates, the epenthetic vowel is needed for the resolution of a phonotactically unacceptable consonant cluster in the two forms with a consonant initial affix, leading to a potential violation of OPIdenStress as in (29c). Here the usual stress constraints place stress on the initial syllable in six cases, and on the penultimate in two cases, resulting in two OPIdenStress violation marks. The shortening that resolves the stress issues leads to potential OPIdentQuant violations in (29b); two marks are assessed for the failure of the short vowel in two forms to match the long vowels in the other six forms. Placing stress uniformly on the initial syllable without shortening would violate *'σHσHσ in two forms in (29d). Shortening all vowels, as in the winner (29a), allows for initial syllable stress, with all paradigm members matching in quantity as well.

All eight forms of the paradigm are compromised in some (non-phonotactically-driven) way in order to look similar to other members of the paradigm.

(29) Optimal Paradigm of hollow verb *ʒe:b* + dative clitic

Input /ʒe:b+ l/+ {i, na, ak, ik, kun, o, a, un} base prosodic head=[ʒe:b]	OPIden Quant	OPIden Stress	$*'\sigma_H\sigma_H\sigma$	*Stress EpenV	Stress Loc	Head MaxBO	DepIO (V)	Max (μ)
a. ☞ {'ʒab.li, 'ʒa.bil.na, 'ʒab.lak, 'ʒab.lik, 'ʒa.bil.kun, 'ʒab.lo, 'ʒab.la, 'ʒab.lun}					**	*** *** **	**	*** *** **
b. {'ʒe:b.li, 'ʒa.bil.na, 'ʒe:b.lak, 'ʒe:b.lik, 'ʒa.bil.kun, 'ʒe:b.lo, 'ʒe:b.la, 'ʒe:b.lun}	*!*				**	**	**	**

10. We specify vocalic mora here, to keep vowel shortening distinct from consonant degemination. While it is unclear whether the distinction matters in LA, the two are distinct in other dialects, as a reviewer pointed out.

c. {'ʒe:b.li, ʒe:.'bil.na, 'ʒe:b.lak, 'ʒe:b.lik, ʒe:.'bil.kun, 'ʒe:b.lo, 'ʒe:b.la, 'ʒe:b.lun}		**!		*!*			**	
d. {'ʒe:b.li, 'ʒe:.bil.na, 'ʒe:b.lak, 'ʒe:b.lik, 'ʒe:.bil.kun, 'ʒe:b.lo, 'ʒe:b.la, 'ʒe:b.lun}			*!*					

Before we turn to verbs with accusative clitics, and now that we have introduced new stress constraints, it is time to revisit the Tableau of Geminate Verbs with Dative Clitics in (12), with the verb [radd] 'to return sth' as an example. While the tableau was complete in terms of paradigms, the fully complete tableau including full paradigms and all constraints appears below.

(30) Optimal Paradigm of geminate verb *radd* + dative clitic (complete constraints)

Input /radd+ l/ + {i, na, ak, ik, kun, o, a, un} base prosodic head=[radd]	Realze Morph	OPIden Stress	$*'\sigma_H\sigma_H\sigma$	*Stress EpenV	Stress Loc	Head MaxBO	DepIO (V)	DepIO (C)
a. ☞ {rad.'dal.li, rad.'dal.na, rad.'dal.lak, rad.'dal.lik, rad.'dal.kun, rad.'dal.lo, rad.'dal.la, rad.'dal.lun}				*** *** **			*** *** **	*** ***
b. {'rad.da.li, rad.'dal.na, 'rad.da.lak, 'rad.da.lik, rad.'dal.kun, 'rad.da.lo, 'rad.da.la, 'rad.da.lun}		**!		**			*** *** **	
c. {'rad.da.li, 'rad.dal.na, 'rad.da.lak, 'rad.da.lik, 'rad.dal.kun, 'rad.da.lo, 'rad.da.la, 'rad.da.lun}			*!*		**		*** *** **	
d. {'rad.da.li, 'rad.da.na, 'rad.da.lak, 'rad.da.lik, 'rad.da.kun, 'rad.da.lo, 'rad.da.la, 'rad.da.lun}	*!*							

The paradigm in (30a) wins despite violating a constraint against stressing epenthetic vowels. This, however, is possible due to a constraint against stressing a heavy antepenultimate syllable when the penultimate syllable is also closed or heavy. Otherwise, the paradigm in (30c) would win. However, ['rad.dal.na] and ['rad.dal.kun], with initial stress, violate the high ranking constraint *'σHσHσ, resulting in a preference for (30a).

3. Verbs plus accusative clitics

The paradigms discussed in the previous section all involve dative clitics as level-two morphemes. Another type of level-two morpheme that verbs may take is accusative clitics. Three of four paradigms discussed in the previous section satisfy Head-Max (B/O) when the dative clitics are replaced with their accusative counterparts. Three of the four paradigms satisfy OP-Iden-Stress without much ado. These are geminate verbs (31), defective verbs (32), and hollow verbs (33). In all three paradigms, all members receive stress on the same syllable, and it is always a heavy syllable.[11]

(31) Geminate verb *radd* 'he returned' + accusative clitic

Input	Optimal Output	'returned'
a. /radd-ni/	'rad.-ni	'~ me'
b. /radd-na/	'rad.-na	'~ us'
c. /radd-ak/	'rad.d-ak	'~ you (Mas)'
d. /radd-ik/	'rad.d-ik	'~ you (Fem)'
e. /radd-kun/	'rad.-kun	'~ you all'
f. /radd-o/	'rad.d-o	'~ him'
g. /radd-a/	'rad.d-a	'~ her'
h. /radd-un/	'rad.d-un	'~ them'

(32) Defective verb *ħike:* 'he talked' + accusative clitics

Input	Optimal Output	'talked to'
a. /ħi'ke:-ni/	ħi.'ke:.-ni	'~ me'
b. /ħi'ke:-na/	ħi.'ke:.- na	'~ us'
c. /ħi'ke:-k/	ħi.'ke:-k	'~ you (Mas)'
d. /ħi'ke:-ki/	ħi.'ke:.-ki	'~ you (Fem)'
e. /ħi'ke:-kun/	ħi.'ke:.-kun	'~ you all'
f. /ħi'ke:-o/	ħi.'ke:(-h)	'~ him'
g. /ħi'ke:-a/	ħi.'ke:.-ha	'~ her'
h. /ħi'ke:-un/	ħi.'ke:.-hun	'~ them'

11. The 2nd Sg Fem and Mas accusative clitics have two allomorphs: [–ik] and [–ak] with verbs that end with a consonant – e.g., (31) and (33) – and [–ki] and [–k] with verbs that end with a vowel – e.g., (32).

(33) Hollow verb ʒe:b 'he brought' + accusative clitics

Input	Optimal Output	'brought'
a. /ʒe:b-ni/	'ʒe:b.-ni	'~ me'
b. /ʒe:b-na/	'ʒe:b.- na	'~ us'
c. /ʒe:b-ak/	'ʒe:.b-ak	'~ you (Mas)'
d. /ʒe:b-ik/	'ʒe:.b-ik	'~ you (Fem)'
e. /ʒe:b-kun/	'ʒe:b.-kun	'~ you all'
f. /ʒe:b-o/	'ʒe:.b-o	'~ him'
g. /ʒe:b-a/	'ʒe:.b-a	'~ her'
h. /ʒe:b-un/	'ʒe:.b-un	'~ them'

Note that some members of the geminate paradigm in (31) undergo degemination; for example, (31e) [radkun] 'he returned you all' instead of *[raddkun]. As the tableau in (34) shows, this violates the Base-output constraint HeadMax (B/O) in order to satisfy higher ranking constraints. The paradigm in (34c) shows that if the verb remains geminate, the paradigm would be unevenly split between two stress patterns.

(34) Optimal Paradigm of geminate verb *radd* + accusative clitic

Input /radd/ +{ni, na, ak, ik, kun, o, a, un} base prosodic head=[radd]	*COMP	OPIDEN STRESS	HEAD-MAX BO	DEP (V)	MAX (μ_C)
a. ☞ { 'rad.ni, 'rad.na, 'rad.dak, 'rad.dik, 'rad.kun, 'rad.do, 'rad.da, 'rad.dun }			***		***
b. { 'radd.ni, 'radd.na, 'rad.dak, 'rad.dik, 'radd.kun, 'rad.do, 'rad.da, 'rad.dun }	*!**			**	
c. { rad.'da.ni, rad.'da.na, 'rad.dak, 'rad.dik, rad.'da.kun, 'rad.do, 'rad.da, 'rad.dun }		***!		***	

Note that the first syllable in the hollow paradigm in (33) does not undergo Closed Syllable Shortening even when the clitic begins with a consonant. In this sense, this paradigm contrasts with the minimally different paradigm of hollow verbs *plus* dative clitics in (26). In the case of (33), there is no phonotactic/syllable motivation for the epenthetic vowel to be added, so there is no issue of stress location, as in (26). The lack of an epenthetic vowel in the forms in (33) means that OP-IdenStress can be satisfied without alteration in any of the forms.

This leaves us with one paradigm: sound verbs. As (35) and (36) show, sound verbs 'to hear' and 'to push' satisfy HEAD-MAX (B/O). However, stress assignment is not uniform across all the members of the paradigm, and thus the constraint OP-IDEN-STRESS seems to be violated.

(35) Sound verb *simiʕ* 'he heard' $[C_1iC_2iC_3]$+ accusative clitics

Input	Optimal Output	'heard'	Sub-optimal Output
a. /simiʕ-ni/	si.'miʕ.-ni	'~ me'	
b. /simiʕ-na/	si.'miʕ.-na	'~ us'	*s<i>.'miʕ.-na
c. /simiʕ-ak/	'sim<i>.ʕ-ak	'~ you (Mas)'	
d. /simiʕ-ik/	'sim<i>.ʕ-ik	'~ you (Fem)'	
e. /simiʕ-kun/	si.'miʕ.-kun	'~ you all'	*s<i>.'miʕ.-kun
f. /simiʕ-o/	'sim<i>.ʕ-o	'~ him'	
g. /simiʕ-a/	si.'mi.ʕ-a, 'sim.ʕ-a	'~ her'	
h. /simiʕ-un/	si.'mi.ʕ-un, 'sim.ʕ-un	'~ them'	

(36) Sound verb *dafaʃ* 'he pushed' $[C_1aC_2aC_3]$ + accusative clitics

Input	Optimal Output	'pushed'	Sub-optimal Output
a. /dafaʃ-ni/	da.'faʃ.-ni	'~ me'	
b. /dafaʃ-na/	da.'faʃ.-na	'~ us'	*d<a>.'faʃ.-na
c. /dafaʃ-ak/	'da.fa.ʃ-ak	'~ you (Mas)'	
d. /dafaʃ-ik/	'da.fa.ʃ-ik	'~ you (Fem)'	
e. /dafaʃ-kun/	da.'faʃ.-kun	'~ you all'	*d<a>.'faʃ.-kun
f. /dafaʃ-o/	'da.fa.ʃ -o	'~ him'	
g. /dafaʃ-a/	da.'fa.ʃ-a, 'da.fa.ʃ-a	'~ her'	
h. /dafaʃ-un/	da.'fa.ʃ-un, 'da.fa.ʃ-un	'~ them'	

There is no winner stress pattern in these paradigms, and in two forms, there are two options for the location of output stress. The two verbs with the accusative clitics for 'her' and 'them' [(35g) *simiʕa* 'he heard her'/(36g) *dafaʃa* 'he pushed her' and (35h) *simʕun* 'he heard them'/ (36h) *dafaʃun* 'he pushed them'] may be pronounced with stressed assigned either to the first syllable – such as (35c) *simʕak* 'he heard you (Mas)' and (36c) *dafaʃak* 'he pushed you (Mas)' – or with stress assigned to the second syllable – such as (35a) *simiʕni* 'he heard me' and (36a) *dafaʃni* 'he pushed me'. This may be the case because the accusative clitics for 'her' and 'them' have two allomorphs each: [–a] and [–ha] for 'her' and [–un] and [–hun] for 'them'. Note that [–ha] and [–hun] surface when the accusative pronominals cliticize to defective verbs that end with a vowel; e.g., (32g) [ħike:-ha] 'he talked to her' and (32h) [ħike:-hun] 'he talked to them' above. Apparently, even when the [h] is not realized in verbs like *simiʕa* 'he heard her', *simʕun* 'he heard them', speakers optionally treat the verbs in terms of stress as if the [h] is there; thus *simiʕ(h)a* 'he heard her', *simiʕ(h)un* 'he heard them'. This makes the penultimate syllable optionally heavy and accordingly it receives stress.

As a result of the aforementioned alternation/free variation in stress assignment, the paradigms consist of two equal sets of stress patterns. Five members are assigned stress on the left-most syllables (c, d, f, g, h), and five members are assigned stress on the second left-most syllables (a, b, e, g, h). Perhaps in terms of Optimal Paradigms, this means that each member of the paradigm induces an equal number of violations of OP-Stress when compared to the other members. If OP-IDEN-STRESS stops being a decisive factor in stress assignment, as shown in the tableau in (37) by the use of ? for candidate (37a), the pressures for antepenultimate vs. penultimate stress are evenly matched while other constraints weigh in favor of maintaining the split paradigm.

(37) Sound verb *dafaʃ* 'he pushed'[$C_1aC_2aC_3$] + acc the full paradigm

Input /dafaʃ/ +{ni, na, ak, ik, kun, o, a, un} Base prosodic head=[dafaʃ]		STRESS LOC	OPIDEN STRESS	HEAD-MAX BO	DEPIO (V)	MAX (μ_V)
a.	☞ {da.'faʃ.ni, da.'faʃ.na, 'da.fa.ʃak, 'da.fa.ʃik, da.'faʃ.kun, 'da.fa.ʃo, 'da.fa.ʃa~da.'fa.ʃa, 'da.fa.ʃun ~ da.'fa.ʃun }		?????			
b.	{ da.'faʃ.ni, da.'faʃ.na, d.'fa.ʃak, d.'fa.ʃik, da.'faʃ.kun, d.'fa.ʃo, d.'fa.ʃa, d.'fa.ʃun }			***** 		*****
c.	{ 'daf.ʃa.ni, 'daf.ʃa.na, 'da.fa.ʃak, 'da.fa.ʃik, 'daf.ʃa.kun, 'da.fa.ʃo, 'da.fa.ʃa, 'da.fa.ʃun }			***	***	***

Candidate (37b) attempts to satisfy OPIDENTSTRESS by deleting vowels, resulting in uniform stress on the penultimate syllable, while candidate (37c) includes inserted vowels to result in uniform stress on the initial syllable. Doing neither, candidate (37a) has stress in two locations in the paradigm, but each location has an equal number of members.

Note that the tableau in (37) raises two counterfactuals possibilities:

(i) If only *da.'fa.ʃa, da.'fa.ʃun* were available, then the paradigm could tip towards penultimate stress, violating HEADMAX-BO and MAX(μ) to accomplish it in the forms with antepenultimate stress.
(ii) If only *'da.fa.ʃa, 'da.fa.ʃun* were available, then the paradigm could tip towards antepenultimate stress, violating HEADMAX-BO, MAX(μ), and DEPIO (V) in the forms with penultimate stress.

However, leaving both options available results in a paradigm balanced between penultimate and antepenultimate stress; one solution is to propose that in such cases, OPIDEN constraints are neutralized. However, we leave this case for future research.

4. Conclusion

In this paper, we showed that whereas base-output constraints play a role in determining the phonological structure of the members of derivational paradigms, Optimal Paradigm constraints that are normally considered only operative in inflectional paradigms may also play a role in derivational morphology, where derivational means 'derived from a base'. We used evidence from LA verbs. We looked at four different types of verbs (sound, geminate, hollow, and defective) and two types of clitics (dative and accusative), giving rise to eight paradigms.

All paradigms satisfy an OP constraint that we called OP-Iden-Stress. The constraint requires all the members of the paradigm to be stressed on the same syllable. Seven of the eight paradigms we discussed clearly satisfy this OP constraint, many vacuously. That is, no Input-Output constraint or any other constraint is violated *solely* to produce a paradigm that satisfies OP-Iden-Stress. Three cases stand out, however. These are:

i. Paradigms of verbs with dative clitics in which the base ends with a superheavy syllable (e.g., [radd]): In this case, Input-Output faithful constraints are violated in an overkill (non-phonotactically movitated) fashion.
ii. Paradigms of hollow verbs with dative clitics: In this case, HeadMaxBO and MaxμV are violated extensively so that the shortened base vowel allows uniform stress in the paradigm as well as uniform syllable size throughout the paradigm.
iii. Paradigms of sound verbs with accusative clitics. We tentatively suggest that this final case appears to be possible only because each member of the paradigm violates OP-Iden-Stress an equal number of times, leaving the decision to other constraints such as HeadMax BO and DepIO (V).

Acknowledgement

We thank the editors, an anonymous reviewer, and the audience at the 26th Arabic Linguistics Symposium for helpful feedback and discussions. Any remaining errors remain ours.

References

Alhawary, M. T. (2011). *Modern standard Arabic grammar: A learner's guide*. United Kingdom: Wiley-Blackwell.

Benua, L. (1997). *Transderivational identity: Phonological relations between words*. Ph.D. dissertation, University of Massachusetts, Amherst.

Kager, R. (1999). Surface opacity of metrical structure in optimality theory. In B. Hermans, & M. van Oostendorp (Eds.), *The derivational residue in phonological optimality theory* (pp. 207–245). Amsterdam: John Benjamins.

Kiparsky, P. (2002). Syllables and moras in Arabic. In C. Fery, & R. Vijver (Eds.), *The syllable in optimality theory* (pp. 147–182). United Kingdom: Cambridge University Press.

Kenstowicz, M. (1996). Base-identity and uniform exponence: Alternatives to cyclicity. In J. Durand, & B. Laks (Eds.), *Current trends in phonology: Models and methods* (pp. 363–393). Salford: European Studies Research Institute, University of Salford.

Kenstowicz, M., & Abdul-Karim, K. (1980). Cyclic stress in Levantine Arabic. *Studies in the Linguistic Sciences*, 10, 55–76.

McCarthy, J. (2005). Optimal paradigms. In L. Downing, T. A. Hall, & R. Raffelsiefen (Eds.), *Paradigms in phonological theory* (pp. 295–371). Oxford: Oxford University Press.

Prince, A. & Smolensky, P. (1993). *Optimality theory: Constraint interaction in generative grammar*. Unpublished manuscript, Rutgers University.

Prince, A., & Smolensky, P. (2004). *Optimality theory: Constraint interaction in generative grammar*. Malden, MA & Oxford: Blackwell.

Ryding, K. C. (2008). *A reference grammar of modern standard Arabic*. United Kingdom: Cambridge University Press.

Watson, J. (2002). *The phonology and morphology of Arabic*. Oxford: Oxford University Press.

Sociolinguistics

The Arabic of Bukhara

A Principal parts analysis of the effects of contact influence on morphological typology

Keri Miller
The University of Arizona

The Arabic spoken in the Bukhara region of Uzbekistan displays an unusual degree of integration between three distinct language families: Semitic, Indo-Aryan, and Turkic. This paper proposes an analysis of Bukhara Arabic based upon data collected between 1935 and 1943. The analysis involves classification of the verbal system drawing on the principal parts theory of Finkel and Stump (2007a), the purpose of which is to determine the essential parts needed to predict the remaining forms in a lexeme's paradigm. The analysis includes an examination of the productivity of the verbal nonconcatenative derivative systems to reveal the changes involved in the partial adoption of concatenative systems. The results shed further light on the morphological processes of Semitic languages under strong contact conditions with other language families.

Keywords: Principal Parts Analysis, Arabic of Bukhara, verbal nonconcatenative derivatives

1. Introduction

This paper seeks to address the question: How can a principal parts analysis illuminate the relationships between the Semitic, Indo-Aryan, and Turkic morphological elements that Bukhara Arabic incorporates into an essentially consistent grammatical system? In order to engage with this question, background familiarity with both the development of Bukhara Arabic and with principal parts theory is necessary.

Brief historical overview of Bukhara Arabic

Language contact between Arabic and Central Asian languages dates at least as far back as the year 676 CE, when the Islamic Khalifate first reached the lands of Maverannahr (Arabic *maœ wara: ʔ in-nahr)* meaning 'that which is beyond the Oxus – or Amu Darya – River', i.e. Transoxiana). Commander Sayyid bin Othman entered the land of Sogdia, consisting of the most fertile land of present day Uzbekistan and part of Tajikistan, and transferred 30,000 captives from Bukhara to Medina to work on irrigation systems (Chikovani, 2004). Later in the same century, at the time that al-Rabi' ibn Ziyad al-Ḥarithi was ruler of Khorasan in Iran, 50,000 Arab families from al-Baṣra and al-Kufa in Iraq moved to the Khorasan region (Mohamed, 1992). Al-Rabi' later sent a large number of families from Bukhara to live in Baṣra in a special neighborhood which he had built for them, known as the Bukharia neighborhood. Between 709 and 712 CE, the Arabs had taken Balkh, Afghanistan, as well as the Bukhara, Samarqand and Qashqadarya regions of present day Uzbekistan, under the leadership of Qutaiba ibn Muslim of the Bahila Arab tribe.

Qutaiba was determined that the Arabs should settle and live in Transoxiana once they had established Islamic rule, and he wanted them to live in the big cities such as Bukhara and Samarqand, where it would be easier to spread Islam. Mohamed (1992) elucidates that from the beginning, the lands taken by the Arabs were divided among the Qayisite and Yemeni tribes who had participated in the conquest: a portion to Rabi'; a portion to the Mudhari (a North Arabian tribe); and a portion to the Yemenites.

At the time of the conquests, the primary language that was spoken in the regions of Transoxiana was Sogdian, a Middle Iranian language written in an Aramaic script. During the centuries following the settlement of the Arabs in Central Asia, the region experienced many battles for power. The following groups (and branches of these groups) held temporary control of the region: Iranian Samanids and Ghaznavids, the Turkic Seljuks, the Mongols, and the Russians (McChesney, 1998). Throughout the power changes, Samarqand and Bukhara in particular managed to maintain strong intellectual Islamic relations with the Arab heartland, which from the 14th to the 17th century centered particularly on Sufi thought. Though very little appears to be written about the Arab communities that Russian Arabists Izmaylova and Burykina discovered in 1930 in Uzbekistan living primarily as sheep nomads (Ingham, 1994), some religious intellectual communication was maintained at least until the nineteenth century, if not until 1917 with the separation due to the Bolshevik revolution (McChesney, 1998).

In November 2000, Georgian linguist Guram Chikovani visited Central Asia with a team of researchers who confirmed the continued presence of an

Arab speech community in the Bukhara region, much as Tsereteli had described it in the previous century (Chikovani, 2003). They also gathered additional data from the Qashqadarya region, where the Arabic speakers speak a distinctly different dialect. During this expedition, Chikovani discovered an oral tradition by which "the ancestors of the Bukhara Arabs were settled in Central Asia by Tamerlan" (Chikovani, 2003, p. 2; 2004b, p. 259) otherwise known as Timur the Lame, who was recorded to have taken captives from Damascus to Samarqand in the year 1401 for the purpose of conscripting them to work on his architectural projects.

Research carried out in the twentieth century on the Arabic dialects of Central Asia indicates interchange and movement throughout the centuries between shepherds living in Uzbekistan, Tajikistan, Afghanistan, and Turkmenistan (Ingham, 1994; Barfield, 1981). It is inevitable that across the centuries at least some of the Arab groups had contact with Assyrians and Kurds, and on a rare occasion the Arabs of Bukhara mention specifically in their stories intermarriage between Arab men and Kazakh and Turkmen women. By all reports the Arabs of Bukhara are fluent in Uzbek and/or Tajik.[1]

Sources of data

The data examined in this paper come from the stories collected by Tsereteli (1956) and Vinnikov (1969) between 1935 and 1943, comprising 288 pages and representing twenty-one speakers. Of these speakers, the youngest was 20 years old and the oldest was 84 at the time of narration. One 30-year-old speaker was a woman; the rest were men. Occasionally both Vinnikov and Tsereteli transcribed the same story, although clearly at different times due to the differences in the narrative sequence and occasionally word choice. The two contrasting descriptions prove useful for determining the level of obligatoriness for certain grammatical structures. The stories are taken from two villages within the Bukhara province: Arab-Khana and Jogari.

It should be noted that there are phonological differences between these two dialects (see Vinnikov, 1969) as well as a range of pronunciations as with any language. However, for the purposes of this paper I am simplifying the phonology of each word to the most basic representations with the highest frequencies. No significant differences appear in the morphological and syntactic elements of the

1. In fact it is not certain at the time of this writing whether or not the Arab communities who do still reside in Bukhara continue to speak Arabic as a mother tongue, but in a personal visit to the Qashqadarya village Jeynov in August of 2012, I was told by the leader of the Arab community that approximately 40 speakers in Jeynov, all over the age of fifty, still speak Arabic to some degree among one another.

dialects of these villages insofar as they relate to this study if at all. In contrast, the stories Vinnikov collected from the Arabs in the village of Jeynov in the Qashqadarya province exhibit a variety that is different enough to require a separate analysis outside the scope of this paper.

This study made use of Vinnikov's (1962) thorough glossary of the Bukhara dialect, which lists each verb alphabetically by consonantal root followed by each of the templatic patterns in which the verb is found throughout all of the stories. This resource has enabled the charting of the relationships between the particular root and the productivity across the ten templatic patterns, in order to answer the question: To what extent is the nonconcatenative Semitic templatic verbal morphology alive and well in the speech of the Arabs documented by Tsereteli and Vinnikov?

Principal parts theory

The reader is likely to be familiar with the notion of principal parts from the study of other languages such as Latin or German. Principal parts are the forms of the verb that a person must know in order to form conjugations across the full system of a given language. In Latin the majority of verbs have four principal parts: present, infinitive, active perfect, and passive perfect (supine). Some languages are lacking any overt expression of one or more of these parts. A student of Latin can memorize these four forms of a verb (see Table 1) in order to place the particular verb into its matching conjugational paradigm, sparing the effort of learning each verb individually.

A similar process is used for the German 'strong verb' otherwise known as 'irregular verb.' The learner consciously memorizes the irregular forms of each verb until the process has been internalized (see Table 2). The so-called 'regular verbs' essentially fit into one pattern and their internal vowels don't shift so classes need not be memorized – hence the term 'regular': for example, 'to ask' *frag***en** – *frag***t** – *frag***te** – **hast ge***frag***t**, 'to search' *such***en** – *such***t** – *such***te** – **hast ge***such***t**. The so-called mixed verbs fall in between these paradigms.

Table 1. Principal Parts of Latin Verbs.

Conjugation	**Present**	**Infinitive**	**Perfect**	**Supine**
I *love*	*amō*	*amāre*	*amāvī*	*amātum*
II *be silent*	*taceō*	*tacēre*	*tacuī*	*tacitum*
III *read*	*legō*	*legere*	*legiī*	*lēctum*
IV *call*	*acciō*	*accīre*	*accīvī*	*accītum*

Table 2. Principal Parts of Four German Strong Verbs.

Lexeme	Present	Past	Participle
to swear	*schwören*	*schwor*	*geschworen*
to take	*nehmen*	*nahm*	*genommen*
to stand	*stehen*	*stand*	*gestanden*
to read	*lesen*	*las*	*gelesen*

Mailhammer (2007) has identified 42 classes of strong verbs in Modern Standard German. Because there are some verbs in which all three stem forms must be memorized, the German verbal system is considered to have three principal parts: the present stem, the past stem, and the past participle. If speakers of German (or computer programs) know this much about any given verb, they know enough to conjugate the verb for any person, number, and gender.

Based on their work in the field of computational linguistics, Finkel and Stump (2007a) have observed deeper implications for the relationship of principal parts to specific language systems. First, from a morphological point of view, they began to critically examine how many lexemes can be stored in a given language, in a range from complete storage to the absolute “minimum of forms needed to deduce the remaining, unstored forms” (Finkel and Stump, 2007a, p. 40). Their quest was for the optimal amount of storage, with a definition of optimal as “both adequate and minimum” (p. 41), in which all of the conjugations can be deduced based on the knowledge of these parts with no unnecessary storage.

Finkel and Stump have divided principal parts analyses into three categories. The previously described common pedagogical analysis they label a “static principal parts analysis” (Finkel and Stump, 2007a, p. 42). At a deeper level, they have developed an “adaptive principal parts analysis” (p. 42), in which principal part *n* governs which morphosyntactic property set (MPS) identifies the principal part *n + 1*. This type of analysis relies on a series of if-then relationships: If MPS has variant X, then it falls into inflection class Y. In this kind of analysis, there is more than one possible set of optimal morphosyntactic properties.

To exemplify the benefits of utilizing the adaptive analysis, consider the analysis of German verbs. The static analysis of German would simply elicit three principal parts to memorize for conjugating the verbs. However, if we adopt the adaptive analysis, the calculation of the average principal parts across the whole strong verb system totals 2.14 because some verbs in fact only require knowledge of one or two principal parts. For example, *schwören* ‘to swear’ has /ø/ as a present tense stem vowel, and that knowledge is enough to classify the conjugation since no other verb shares that present stem vowel (Mailhammer, 2007a). Table 3 is taken from Finkel and Stump (2007a, p. 42), illustrating a hypothetical language in which

Table 3. Static principal parts (in shaded cells) for a hypothetical system.

	W	X	Y	Z
I	a	e	i	m
II	b	e	i	m
III	c	f	j	n
IV	c	g	j	n
V	d	h	k	o
VI	d	h	l	o

there are six conjugations, four morphosyntactic property sets (W-Z) and three principal parts (W, X, and Y) by which all conjugations can be determined. The shaded cells indicate which of the principal parts remain static in such a system.

If the same language is analyzed from a perspective of adaptive principal parts, new relationships are revealed (Finkel and Stump, 2007a). As illustrated in Table IV, in an adaptive system, conjugations I and II only require one identified principal part in order to determine that a given lexeme belongs to them. Conjugations III – VI, however, would each require two principal parts, though the determining MPSs vary between III and IV and V and VI.

The third and most complex type of analysis that Finkel and Stump have identified is the "dynamic principal parts analysis" (2007a, p. 44). This type of analysis is the most flexible and allows for the deepest understanding of the relationships between the parts at a structural level. It allows for the smallest number of morphosyntactic property sets to be considered principal parts. Since optimal MPSs in the dynamic model do not have to share any property set (contrary to the other two models), the dynamic principal parts analysis enables the perception of more direct relationships between "the pairings of word forms with the morphosyntactic property sets they realize" (Finkel and Stump, 2007a, p. 45).

Table 4. Adaptive principal parts (in shaded cells) for the hypothetical system.

	W	X	Y	Z
I	a	e	i	m
II	b	e	i	m
III	c	f	j	n
IV	c	g	j	n
V	d	h	k	o
VI	d	h	l	o

Table 5. Dynamic principal parts (in shaded cells) for the hypothetical system.

	W	X	Y	Z
I	a	e	i	m
II	b	e	i	m
III	c	f	j	n
IV	c	g	j	n
V	d	h	k	o
VI	d	h	l	o

Table V illustrates the dynamic principal parts analysis, whereby a lexeme which has exponent 'f' in MPS X is conjugation III, and 'g' is IV and so on (Finkel and Stump, 2007a, p. 44). This analysis may not be transparent to a language learner because it requires knowledge about which MPS specifically determines the conjugation to which a lexeme belongs.

Finkel and Stump (2007a) hypothesize that principal parts can reveal "a distillation of the implicative relations that exist among the members of a lexeme's paradigm; they reveal an important domain of typological variation in morphology" (as cited in Mailhammer, 2007, p. 81). In order to classify languages by morphological type according to the principal parts analyses, researchers (e.g., Finkel and Stump, 2007a, pp. 49–64; Mailhammer, 2007, pp. 84–86) have delineated five basic categories that can be examined for comparative purposes, particularly using the dynamic analysis. The first of the categories is referred to as Criterion A, addressing the number of principal parts needed to determine a lexeme's paradigm (Finkel and Stump, 2007a, p. 66). Criterion B refers to whether or not the principal parts differ for various inflection classes according to a dynamic analysis. If they are in fact the same, the language can be classified as a *parallel system*, and if different, then a *skewed system*. All static analyses present parallel systems, but under adaptive and dynamic analyses, skewed systems are possible.

Criterion C differentiates between *segregated* and *integrated systems*, depending on whether a lexeme is predicted by more than one principal part (integrated) or whether separate principal parts distinctly derive non-principal parts (segregated). This is determined by the number of "dynamically defined principal parts needed to determine a given word in a lexeme's paradigm" (Finkel and Stump 2007a, p. 66). In the case of the German strong verbs, for example, by a static analysis, each of the three principal parts is used to determine each lexeme, so it is segregated. However, by an adaptive or a dynamic analysis, there is overlap between some principal parts in determining the nonprincipal parts. For example, as stated previously, only one principal part is needed to determine the conjugation with *schwören* (see Table 2) because in the whole paradigm of German strong verbs it is the only one with /ø/ in the present stem (Mailhammer, 2007, pp. 98–99). However, in the case of *nehmen*, two principal parts – both the present stem and the past participle – are needed, because there are several other conjugations with /e/ in the present stem. Since there are no other conjugations with /e/ in the present stem and /ø/ in the past participle, no other principal parts are required. In the case of *stehen* and *lesen*, three principal parts are required since other conjugations share each of their parts. There is overlap in the number of principal parts needed to determine the category of one lexeme, hence by an adaptive analysis, the German strong verbal system is *integrated* (Mailhammer, 2007, p. 100).

Criterion D addresses the morphological relation between a dynamically defined principal part and the nonprincipal parts assumed to be deduced from it (Mailhammer, 2007, p. 66). The answer distinguishes between a *morphologically coherent system* from an *incoherent system*. This question deals with the relationship between the morphological form of the principal parts versus that of the nonprincipal parts. Because the consonantal part of the stem does not change with the principal part, the strong German verbal system is coherent. The question posed in Criterion E is whether corresponding word forms in distinct paradigms are determined by the same dynamically defined principal parts. The answer leads to the classification as an *isomorphic* or *non-isomorphic system*. In this case, *isomorphic* means that the same dynamic parts in a system determine corresponding word forms (Finkel and Stump, 2006 p. 20). The static analysis of the strong German verbs is isomorphic because all of the principal parts are used to determine all of the forms across the various classes. However in the adaptive and dynamic analyses, the system becomes *non-isomorphic* since different principal parts can determine forms across the class, sometimes present stem, sometimes preterit, sometimes a combination, and so forth. In fact a skewed system can never be isomorphic.

These documented advantages of principal parts analysis led to its selection for the examination of the morphological structure of Bukhara Arabic in this study. The study attempts to answer if such an analysis can reveal how deeply the structures of concatenative Turkic and Iranian languages have affected the nonconcatenative Semitic structure of Bukhara Arabic, or can shed light on the process of language maintenance and change under strong multilingual contact conditions.

2. Methodology

In the quest to understand the relationships between various language systems and the principal parts they employ, Finkel and Stump (2007b) developed an online tool that is available to the public.[2] This online tool requires the user to form a *plat* – a chart of the salient morphosyntactic property sets of a given language system – by compiling the paradigms present in the given system which is later converted to plain text form (Unicode Tf8) for online analysis. The online tool also presents the option to enter a chart of paradigms in order to receive a plat that may need some adjusting.[3]

In this study, a collection of all verbs, infinitives and participles from the data were compiled onto a master list of morphosyntactic properties with categories for tense, voice, mood, and aspect combined with categories for person, number and gender. A verbal paradigm was created based on the compiled list. Cells that were not found in the available corpus were filled based on the analysis of the verbal paradigms by Fischer (1961) and Vinnikov (1965).[4] For transcriptions IPA was used except for the emphatic /sˤ/, /zˤ/, /tˤ/, and /dˤ/, which were denoted using capital letters.[5]

Table 6 was first constructed based on a total of nine distinct inflection classes. Three of these classes were omitted from the table due to their irregular nature.

Table 6. Six productive inflection classes in Bukhara Arabic.

1.	Sound verb with three consonantal roots	*qasam*	'divide'
2.	First consonantal root is semi-vowel '*waw*'	*wazan*	'weigh'
3.	First vowel is underlying '*waw*'*	*kala:*	'eat'
4.	Hollow verb with a semi-vowel in the middle	*qo:l*	'say'
5.	Final consonantal root is a semi-vowel or *alif*	*baka:*	'cry'
6.	Final consonantal root is a geminate	*shadd*	'tie'

*The 'waw' appears in forms II and VII of the plat.

2. This tool is located on the web at: http://www.cs.uky.edu/~raphael/linguistics/analyze.html.

3. I am considerably indebted to Raphael Finkel for his assistance in developing the plat for Bukhara Arabic, as well as to Gregory Stump for his insights in analyzing the results

4. In his 1965 article, Vinnikov has created elaborate inflection charts for the Qashqadarya Arabic dialect, and I was able to make some inferences based on what I observed in the Bukhara Arabic data and what is happening in the QA data. The information in Fischer's (1961) article was also of enormous value for filling in gaps and double checking my comprehension of the verb forms.

5. I have also opted to simplify Vinnikov and Tsereteli's transcriptions, aiming to choose the most frequently occurring variants of any alternations. Again I have leaned on the descriptions of Fischer (1961) and Vinnikov (1965) to guide my choices. For a more detailed description of the Bukhara Arabic phonological system see Tsereteli's Foreword (1956:XX-XXIV).

The three irregular types of verbs that were not included in the analysis due to their exceptional nature (see also Fischer, 1961, p. 257), are some of the highest frequency verbs in the language. One is *sawa* 'to do', a type of verb called in Standard Arabic *lafi:f maqru:n* 'clustered together' because there are a vowel and a semi-vowel next to each other in the root. No other similar lexemes were detected in the BA corpus. Nevertheless, it should be noted that, even were these lexemes added to the plat, the fundamental results would remain unchanged. The second is *anTa* 'to give', a Form IV (causative) verb that does not appear in Form I of Bukhara Arabic verbs. Fischer elaborates on the phonological variations and the probable connection with Aramaic (1961, p. 257, note 3).

The other irregular verb forms are *ɣada:* and *ʤa:* 'to go' and 'to come'. In preterit form they usually take a 'k' at the end that has several variations and sometimes extends to '*kən*' (*ɣadak*, *ɣadakən*). Johnstone (1968) suggests that this feature may be a remnant of a deictic particle found in some Northern and Eastern Arab tribes. Again, including these on the plat will not change the number of principal parts, and since it is a category completely restricted to two verbs with no distinct semantic or grammatical meaning attached to the suffix it was excluded from analysis in this study.[6]

In order to fully consider the morphosyntactic properties of BA, the Semitic verbal derivative patterns also need to be taken into account. In the data examined, all of the ten common patterns in Standard Arabic were exemplified. Form I is the basic form. Form II, which has a distinct transitive/causative/intensification connotation, is the second most frequent form, and there is a very high number of roots that appear both in Form I and Form II (see Table VIII, p.17). This can be considered evidence that Form II has a productive derivational property and it was included in the plat. The other causative, Form IV, only appears with the lexeme *anTa* 'to give', as mentioned above. Form VII, which has a passive meaning, is found throughout the stories frequently and consistently in combination with words that were used in Form I, though unlike Form II, it is seldom found with verbs that have no Form I counterparts in the stories. The occurrence of roots used in more than three pattern templates was observed for two roots only: *qaʕad* 'he sat' – *qaʕʕad* 'he set (s.thing) down' – *itqaʕʕad* 'he sat' – *inqaʕad* 'he sat'[7] with a root meaning 'to sit' (I, II, V, VII), and *Tab* 'to be nice,

6. For a quick overview of these irregular verbs, see Fischer 1961 p.257 under *Unregelmäßig gebildete Verbalformen*

7. Vinnikov (1962: 172) provides two examples of uses of form V and VII that appear to be interchangeable:

(1) *min Harr fi bejt ma: yitqaÀÀad* (2) *min Harr arun bejt ma: yinqaÀad*
Both carry the impersonal meaning: "Due to the heat one doesn't sit at home"

to like' – *Tayyab* 'to clean up, repair' – *iTTayyab*' to be garnished, cleared out, brushed' – *inTayyab* (same as Form V) with a root meaning 'to be good' (also I, II, V, VII). Notably in neither of these cases does Form V differ in meaning from Form VII.

The rarity of the instances of roots used in multiple patterns and the fact that most often the roots in the higher forms did not also show up in Form I leads me to the conclusion that perhaps many of the patterns are no longer productive in BA but rather the lexemes have been memorized without referral to the root and pattern combination. For this reason they were excluded from the morphosyntactic property sets to be analyzed, which included only forms I, II and VII.

Table 7 shows the breakdown of patterns as they are found in Vinnikov's glossary in terms of how many occurrences of verbs in each form are found throughout the whole glossary, in comparison with the number of instances of verbs in one form not appearing in any other form. For this study, word frequency was not measured, though the results of such a test would surely be pertinent, begging future treatment. Table 8 illustrates the number of times that each verb form is found in combination with other forms. If a verb is listed in the category 'Single Verb Form' then no other form is listed for it in the glossary.

Table 9 illustrates the plat for preterit Form I across all six verb classes and all conjugations,[8] singular and plural. The full plat is listed in the Appendix. The conjugations for Bukhara Arabic person/number/gender are to a large extent the same

Table 7. Verb patterns found in Vinnikov's Glossary (1962).

Form	Total occurrences	Single verb form
I	231	87
II	132	48
III	4	0
IV	1	0
V	3	3
VI	2	1
VII	79	3
VIII	15	6
IX	6	3
X	5	4

8. Here the KATR output is included below each line of the plat in the singular for ease of reading. The plural can be inferred in the same manner. KATR is the name of a set-based computational language used for representation of morphological systems.

Table 8. Occurrences of Verb Pattern Combinations.

Combination of forms	Total occurrences in Glossary
I, VII	58
I, II	48
II, V	12
I, II, VII	10
I, II, V	6
I, VIII	5
I, V, VII	4
I, V, VIII	3
II, IX	2
I, II, V, VII	2
I, II, VI	1
I, VII	58

as those of the majority of Arabic dialects. TEMP stands for template, for which '1A' equals the tri-consonantal root.[9] If the verb only exhibits two consonants, such as *baka*: 'cry', a zero-morpheme is placed in the slot exhibiting the third consonant (each dash line represents a consonant). The morphosyntactic property sets (MPSs) run across the top; they are the combination of the semantic, lexical, and grammatical morphemes that conjugate variously in the language system, such as preterit Form I first person singular. Of the 10 MPSs in the preterit pattern I, nine are unique, because 1sg and 2msg are the same across all conjugations. Every box in which the conjugation of a lexeme corresponds with a given MPS is called an *exponence*, so the complete plat for the preterit pattern I includes a total of 66 exponences. Form II and Form VII have fewer exponences, because some of the lexemes are not used in those patterns; for example, there is no form II of *qo:l* 'say' or form VII of *baka:*. Table 10 follows up with a complete list of the MPSs following the singular forms of *qasam* 'divide'.

From the MPS chart, I omitted three properties because they do not change across the conjugations: negative (formed with *la* or *ma* interchangeably; see Fischer, 1961, p. 258), optative (always formed with a /ta-/ prefix attached to the subjunctive form), and past continuous (always formed with a variation of *kən* preceding the imperfective or participle and sometimes assimilated to the first phoneme of the verb itself). According to Finkel (personal communication, December 7, 2011), even if included in the plat, those MPSs "will not show up as separate distillations (a distillation is an MPS whose pattern differs from all other MPSs)".

9. According to Finkel, another way to set up the plat would be to have separate templates; one for tri-consonantal roots (sound verbs) and one for the other types. However the single template is a more straight-forward option. (Personal communication, 12/7/2011).

Table 9. Preterit Form I.

CONJ	**pretI1Sg**	**pretI2mS**	**pretI2fS**	**pretI3mSg**	**pretI3f**
TEMP	1A	1A	1A	1A	1A
qasam(divide)	-a-a-t	-a-a-t	-a-a-ti	-a-a-	-a-a-et
	qasamt	qasamt	qasamti	qasam	qasamet
wazan(weigh)	-a-a-t	-a-a-t	-a-a-ti	-a-a-	-a-a-et
	wazant	wazant	wazanti	wazan	wazanet
kala (eat)	-a-ej-t	-a-ej-t	-a-ej-ti	-a-a-	-a-∅-et
	kalejt	kalejt	kalejti	kala	kalet
qol (say)	-u-∅-t	-u-∅-t	-u-∅-ti	-o-∅-	-o-∅-et
	qult	qult	qulti	qol	qolet
baka (cry)	-a-ej-t	-a-ej-t	-a-ej-ti	-a-a-	-a-∅-et
	bakejt	bakejt	bakejti	baka	baket
shadd (tie)	-a-ej-t	-a-ej-t	-a-ej-ti	-a-a-	-a-∅-et
	shaddejt	shaddejt	shaddejti	shadda	shaddet
CONJ	**pretI1Pl**	**pretI2mPl**	**pretI2fP**	**pretI3mPl**	**pretI3fP**
TEMP	1A	1A	1A	1A	1A
qasam(divide)	-a-a-na	-a-a-tu	-a-a-tin	-a-∅-u	-a-∅-in
wazan(weigh)	-a-a-na	-a-a-tu	-a-a-tin	-a-∅-u	-a-∅-in
kala (eat)	-a-ej-na	-a-ej-tu	-a-ej-tin	-a-∅-u	-a-∅-in
qol (say)	-u-∅-na	-u-∅-tu	-u-∅-tin	-o-∅-u	-o-∅-in
baka (cry)	-a-ej-na	-a-ej-tu	-a-ej-tin	-a-∅-u	-a-∅-in
shadd (tie)	-a-ej-na	-a-ej-tu	-a-ej-tin	-a-∅-u	-a-∅-in

Table 10. Bukhara Arabic Morphosyntactic Property Sets: *qasam* 'divide'.

MPS	**1 singular**	**2 feminine singular**	**3 masculine singular**
preterit I	*qasamt*	*qasamti*	*qasam*
preterit II	*qasamt*	*qasamti*	*qasam*
preterit VII	*inqasamt*	*inqasamti*	*inqasam*
imperfect I	*maqsum*	*tuqsumi:n*	*muqsum*
imperfect II	*maqussum*	*mitqussumi:n*	*muqussum*
imperfect VII	*minqasam*	*tinqasami:n*	*minqasam*
imperative I	________	*uqsumi:*	*uqsum*
imperative II	________	*qussumi:*	*qussum*
subjunctive	*aqsim*	*tiqsimi:n*	*jiqsim*
inflected infinitive	*qosmaha:ni*	*qosmaha:nki*	*qosmaha:nak*
inflected participle	*qo:sminni*	*qo:smetinni*	*qo:sminnak*
		3masculine singular	**3 feminine singular**
active participle I	________	*qo:sim*	*qo:sma*
active participle II	________	*miqussum*	*miqussuma*
active participle VII	________	*minqusum*	*minqusma*

3. Results

The synopsis of the principal parts analysis online tool demonstrates that the combination of MPSs and inflectional classes in BA yields 385 distinct exponences, 105 MPSs, 100 unique MPSs, and 14 distillations. The distillations, the distinct patterns that are different from all other MPSs on the plat and therefore crucial to deduction of the paradigm, are listed in Table 11.

Understanding distillations is the key to understanding the predictive relationships analyzed by Finkel and Stump's (2007b) online tool. The program analyzes the patterns of each conjugation in order to determine which patterns can be predicted based on each morphosyntactic property set, marking each new series of predictions as a new distillation. For example, the first morphosyntatic property set in this plat, preterit Form I first person singular, is by default a distillation because it is the first pattern of predictability that the computer program encounters. This becomes visible on the output by selecting 'predictiveness' from the menu of possible computations provided with the online tool. The level of predictiveness is marked with a binary code: Either the pattern can be distinguished in X conjugation for Y set of morphosyntactic properties (1) or it cannot (0). So MPS 1 shows the following pattern, as compared to the input given in the plat. As the comparison in Table 12 indicates, the input for the conjugation of *qoːl* 'say' is the only input that can be consistently predicted in this set. Comparing this information with Table 9, one can see that *qasam* 'divide' and *wazan* 'weigh' share the same conjugational pattern, and *kalaː 'eat'*, *bakaː 'cry'*, and *shadd* 'tie' share another one which is reflected in the binary codes of the distillation pattern. The next distillation is the

Table 11. MPS distillations for Bukhara Arabic (the full plat is given in the Appendix).

1.	preterit	1 singular
2.	preterit I	3 masculine singular
3.	preterit I	1 singular
4.	preterit II	3 masculine singular
5.	preterit VII	1 singular
6.	preterit VII	3 masculine singular
7.	imperfect I	1 singular
8.	imperfect II	1 singular
9.	imperfect II	1 plural
10.	imperative II	2 masculine singular
11.	subjunctive	3 masculine plural
12.	active participle I	3 feminine singular
13.	inflected active participle I	3 masculine singular
14.	inflected infinitive	1 plural

third masculine singular of preterit Form I, which Table 9 displays with no distinctions between any of the conjugations except *qo:l 'say'*.

As is visible in Table 14, Distillation 7, the Form I imperfect first person singular is the only distillation that is completely filled with (1) and no (0); therefore, every lexeme is predictable and is recognized as the principal part in this language system. Table 15 shows the full layout of the Form I imperfect singular MPSs with the KATR output beneath each exponent for clarity. Distillation 7 is highlighted in bold for contrast. Therefore, in this distillation, all of the classes show their unique characteristics that enable a learner of the language to place them in the correct inflection class and thereby deduce the rest of the paradigm.

Table 12. Predictiveness of distillation 1 (preterit I 1Sg). Overall predictiveness of (pretI1Sg): 0.423.

Lexeme	Predictiveness across all |average 14 distillations
qasam	X 1 1 1 1 1 0 1 0 0 0 1 0 0 | 0.538
wazan	X 1 1 1 1 1 0 1 0 0 0 1 0 0 | 0.538
kala	X 1 0 0 0 0 0 0 0 0 0 0 0 1 | 0.154
qol	X 1 1 1 1 1 1 1 1 1 1 1 1 1 | 1.000
baka	X 1 0 0 0 0 0 0 0 0 0 0 0 1 | 0.15
shadd	X 1 0 0 0 0 0 0 0 0 0 0 0 1 | 0.154

Table 13. Predictiveness of distillation 1 (preterit I 3mSg). Overall predictiveness of (pretI3mSg): 0.167.

Lexeme	Predictiveness across all |average 14 distillations
qasam	0 X 0 0 0 0 0 0 0 0 0 0 0 0 | 0.000
wazan	0 X 0 0 0 0 0 0 0 0 0 0 0 0 | 0.000
kala	0 X 0 0 0 0 0 0 0 0 0 0 0 0 | 0.000
qol	1 X 1 1 1 1 1 1 1 1 1 1 1 1 | 1.000
baka	0 X 0 0 0 0 0 0 0 0 0 0 0 0 | 0.000
shadd	0 X 0 0 0 0 0 0 0 0 0 0 0 0 | 0.000

Table 14. Predictiveness of Distillation 7 (imperfI1cSg). Overall predictiveness of (imperfI1cSg): 1.000.

Lexeme	Predictiveness across all |average 14 distillations
qasam	1 1 1 1 1 1 X 1 1 1 1 1 1 1 | 1.000
wazan	1 1 1 1 1 1 X 1 1 1 1 1 1 1 | 1.000
kala	1 1 1 1 1 1 X 1 1 1 1 1 1 1 | 1.000
qol	1 1 1 1 1 1 X 1 1 1 1 1 1 1 | 1.000
baka	1 1 1 1 1 1 X 1 1 1 1 1 1 1 | 1.000
shadd	1 1 1 1 1 1 X 1 1 1 1 1 1 1 | 1.000

Table 15. Plat View of Distillation 7 in bold print.

CONJ	impfI1Sg	impf2mSg	impfI2fSg	impfI3mSg	impfI3fSg
TEMP	**1A**	1A	1A	1A	1A
qasam	**ma**-∅-**u**-	tu-∅-u-	tu-u-∅-i:n	mu-∅-u-	tu-∅-u-
divide	**maqsum**	tuqsum	tuqsumi:n	muqsum	tuqsum
wazan	**ma**-∅-**a**-	tu-∅-i-	tu-∅-i-i:n	mu-∅-i-	tu-∅-i-
weigh	**mawzin**	tuzin	tuzini:n	muzin	tuzin
kala	**mo**-**e**-∅-	to-e-∅-	to-∅-∅-i:n	mijo-e-∅-	to-e-∅-
Eat	**mokel**	tokel	tokli:n	mijokel	tokel
qol	**ma**-**u**-∅-	mit-u-∅-	mit-u-∅-i:n	me-u-∅	ti-u-∅-
Say	**maqul**	mitqul	mitquli:n	mequl	tiqul
baka	**ma**-∅-∅-**I**	ti-∅-∅-I	ti-∅-∅-i:n	mi-∅-∅-i	ti-∅-∅-i
cry	**mabki***	tibki	tibki:n	mibki	tibki
shadd	**ma**-**i**-∅-	mit-i-∅-	mit-i-∅-i:n	me-i-∅-	mit-i-∅-
tie	**mashidd**	mitshidd	mitshiddi:n	meshidd	mitshidd

*The full form of *mabki:* is mab ∅ k ∅ ∅ i:, in which the three root consonants are represented by b, k, ∅.

The program enables the computer to recognize all of the classes in this distillation, and it does not need any other information to place verbs into classes. This means that for this plat, Distillation 7 is the optimal amount of information storage necessary for completing all of the conjugations for the given linguistic system.

For the adaptive principal parts analysis, again only one principal part is required, which the program translates into an if-then statement. In other words, if Distillation 7 takes variation 1, the inflection class is *qasam*. If it takes variation 2, the inflection class is *wazan*, and so on. The computer program reads as follows:

We need 1 adaptive principal parts:
... if distillation 7 (imperfI1Sg) has variant e31_3 (mo1Se2S3S)
the inflection class is kala
...if distillation 7 (imperfI1Sg) has variant e31_5 (ma1S2S3Si)
the inflection class is baka
...if distillation 7 (imperfI1Sg) has variant e31_1 (ma1S2Su3S)
the inflection class is qasam
... if distillation 7 (imperfI1Sg) has variant e31_6 (ma1Si2S3S)
the inflection class is shadd
... if distillation 7 (imperfI1Sg) has variant e31_2 (maw1S2Si3S)
the inflection class is wazan
... if distillation 7 (imperfI1Sg) has variant e31_4 (ma1Su2S3S)
the inflection class is qol

As for the dynamic analysis, the computer program calculated all of the distillations in which each class appears then calculated the average principal parts needed to deduce the exponence of each class. The average for each class and for each distillation is one; therefore the total average is one, which holds to the rule that the dynamic analysis cannot have more principal parts than the static analysis.

5. Language classification implications

Knowing that the system has exactly one principal part puts it in one of the simpler morphological classes in terms of the amount of information needed to deduce the whole verbal paradigm. Looking at the other four criteria by which Finkel and Stump classify typologies, Bukhara Arabic also falls consistently into the most regularly predictable types. The fact that there is only one principal part means that by default the part has to be the same for all inflection classes – there is no alternative, so it is a parallel system according to Finkel and Stump's Criterion B. Since it only has one principal part, Criterion C (segregated versus integrated) does not apply. As for Criterion D, the BA system is very much a coherent system because all of the forms of the verbs are based on the same tri-consonantal system (Template 1A in the plat). In answer to Criterion E, logically BA must be isomorphic because there is only one principal part to determine the corresponding word forms. This analysis would determine that BA is a language with one principal part that is parallel, coherent and isomorphic according to static, adaptive, and dynamically defined principal parts analyses.

During the process of gathering the data and classifying words into inflectional classes for the purpose of the principal parts analysis, it has become evident that the BA verbal system is organized into three morphological levels. Though syntactically BA takes Subject Object Verb word order, which is rare for Arabic varieties, morphologically at its base level it is a verbal system that closely resembles that of the majority of Arabic varieties. The second level is specific to the active participles in the third person singular and plural forms. In the event that they do not take an object there is nothing unusual in the world of Arabic varieties about their taking either progressive or past perfect verbal force. For example the verb *qo:l* 'he said' becomes *qo:jil* as the active participle 'he who is (or has been) saying'. However, if an object is added to this participle, an /-in(n)-/ buffer morpheme is inserted between the participle and the clitic object suffix, so *qo:jlinnu* means 'he said it (masculine) '. Likewise, *qo:jli:n* 'they said' becomes *qo:li:nnu* 'they said it (masculine)'.

Owens (2006) has traced this buffer /-in(n)-/ extensively in Arabic varieties that essentially form a path from Nigeria to Uzbekistan (for details on the debate regarding the origins of this morpheme see also Retsö (1988), Blau (1981), Brusted (2000), Eksell (1984), and Fischer (1961)). However, apart from some dialects of Arabic found in Central Asia such as BA, such an /-in(n)-/ morpheme is always followed by an object suffix, across the inflectional paradigm. This differs from the Arabic in the dialects of Uzbekistan and Afghanistan (see Barfield (1981) and Ingham (2003) for more information on the connections between these speech communities). The Arabs in these regions share the unusual innovation of inflecting the participle with the /-in(n)-/ morpheme with a subject pronoun, but only in the first and second person singular and plural paradigms. Additionally, the Arabs of Uzbekistan inflect a copula in a similar manner (see Windfuhr 2004, Versteegh 1984 and Ratcliffe 2004).

Both the inflected participle and another unusual feature specific to Central Asian Arabic, the inflected infinitive, show up very distinctly in the plat for the principal parts online tool. It is logical that they are concatenatively attached to the end of the Arabic participial forms, and so the innovation does not interfere with the base Semitic root and template morphology. Tables 16 and 17 illustrate this process. See Appendix for the full plat.

Table 16. Singular distillations of inflected participles.

CONJ	inflAP1mSg	inflAP1fSg	inflAP2mSg	inflectedAP2fSg
TEMP	1A	1A	1A	1A
qasam	-o:-∅-inni	-o:-∅-etinni	-o:-∅-innak	-o:-∅-inki
wazan	-o-∅-inni	-o-∅-etinni	-o-∅-innak	-o-∅-inki
kala	-a:-∅-inni	-a:-∅-tinni	-a:-∅-innak	-a:-∅-inki
qol	-oj-∅-inni	-oj-∅-etinni	-oj-∅-innak	-oj-∅-inki
baka	-a:-∅-inni	-a:-∅-etinni	-a:-∅-innak	-a:-∅-inki

Table 17. Plural distillations of inflected infinitives.

CONJ	inflInfin1Pl	inflInfin2mPl	inflInfin2fPl
TEMP	1A	1A	1A
qasam	-a-∅-aha:na	-a-∅-aha:nkum	-a-∅-aha:nkin
wazan	-o-∅-aha:na	-o-∅-aha:nkum	-o-∅-aha:nkin
kala	-a-∅-aha:na	-a-∅-aha:nkum	-a-∅-aha:nkin
qol	-oj-∅-aha:na	-oj-∅-aha:nkum	-oj-∅-aha:nkin
baka	-a-∅-aha:na	-a-∅-aha:nkum	-a-∅-aha:nkin
shadd	-a-∅-aha:na	-a-∅-aha:nkum	-a-∅-aha:nkin

6. Conclusions

The principal parts analyses place the Bukhara Arabic verbal system in essentially the simplest combination of the criteria which Finkel and Stump use to categorize languages by morphological typology. Mailhammer (2007) suggests a hierarchy of complexity based on correlations for principal parts systems. A system with fewer principal parts is a simpler system. A parallel system is simpler than a skewed one. A morphologically incoherent system is more complex than one that is coherent. Isomorphic systems are not as complex as those that are non-isomorphic. Moreover Mailhammer suggests that a complex system may be "more difficult to learn than a less complex one" (p. 85), noting that such a claim would require further examination.

Finkel and Stump (2007a, p. 65) also consider that, "Although language learners may not rely directly on principal parts as a basis for inferring unencountered members of a paradigm, there is an important sense in which principal parts provide a measure of a paradigm's learnability." The higher the predictability according to a language's dynamically defined principal parts, the more easily learnable they expect the language to be.

The study revealed that in regards to the question of the productivity of the root and pattern systems, it appears that only the most salient elements (Form II as transitive/causative/intensifying and Form VII as passive) in the Arabic verb pattern system have been productively retained. Clearly the root and pattern derivation processes are less salient in the verbal system than the intrinsic stem conjugational morphology of this 'peripheral' form of Arabic.

It appears plausible to suggest that the nomadic peoples who developed and maintained over the centuries the Arabic variety which became BA distilled the language to a very basic form – more easily transferable through generations faced with multilingualism and severely restricted contact with the Arabic heartland. The fact that none of the more 'innovative' inflections specific to BA that show up in the infinitives or the inflected active participles even appear as a distillation illustrates that they are very predictable in their use and they bear no relation to the stems, in a similar manner to the other forms I omitted like the optative and the past continuous. Apparently in order to preserve the nature of their highly consistent system, the Bukhara Arabic speakers found ways to work the most complex changes in their language into the morphosyntax external to the core of the stem, treating the differences in a fundamentally similar manner to clitic suffixes.

The results of this study provide an examination of the elements inherent to the BA verbal system and their relation to each other in terms of complexity and conjugational predictability. A comparison of principal parts analyses of other varieties of Arabic would generate useful further study on one hand, as would further historical analysis of Bukhara Arabic on the other, in order to establish more distinctly the morphological aspects of the diachronic process of language change under intense multilingual contact conditions.

References

Barfield, T. J. (1981). *The central asian Arabs of Afghanistan*. Austin: University of Texas Press.

Blau, J. (1981). *The emergence and linguistic background of Judaeo-Arabic: A study of the origins of Middle Arabic* (2nd edn). Jerusalem: Ben Zvi Institute.

Brusted, K. (2000). *The syntax of spoken Arabic: A comprehensive study of Moroccan, Egyptian, Syrian, and Kuwaiti dialects. Washington, D*.C.: Georgetown University Press.

Chikovani, G. (2003). Central Asian Arabic dialects. In I. Ferrando, & J. J. S. Sandoval (Eds.), *Proceedings of the 5th Conference of the Association Internationale de Dialectologie Arabe (AIDA)* (pp. 1–12). Cádiz: Universidad de Cádiz Publicationes.

Chikovani, G. (2004). Some evidence on the history and language of the Arabs of Central Asia. In G. Rasuly-Paleczek, & J. Katschnig (eds.), *Central Asia on display: Proceedings of the VII Conference of the European Society for Central Asian Studies* (pp. 255–260). Berlin: LIT Verlag.

Eksell, K. (1984). On participle constructions with an n-element in some Arabic dialects. *Studia Orientalia*, 55(20), 395–409.

Finkle, R., & Stump, G. (2006). *Principal parts and morphological typology*. Department of Computer Science, University of Kentucky, Lexington, KY, Available from ftp://ftp.cs.uky.edu/cs/techreports/459-06.pdf. (Ms., Technical Report No. 459-06).

Finkle, R., & Stump, G. (2007a). Principal parts and morphological typology. *Morphology*, 17, 39–75. DOI: 10.1007/s11525-007-9115-9

Finkle, R., & Stump, G. (2007b). Principal parts analysis online tool. http://www.cs.uky.edu/~raphael/linguistics/analyze.html

Fischer, W. (1961). Die Sprache der arabischen Sprachinsel in Uzbekistan [The language of the Arabic Linguistic Enclave in Uzbekistan]. *Der Islam*, 35, 232–263.

Ingham, B. (1994). *The effect of language contact on the Arabic dialect of Afghanistan*. In J. Aguadé, F. Corriente, & M. Marugán (Eds.), Actas del Congreso Internacional sobre Interferencias Lingüisticas Arabo-Romances y Paralelos Extra-Iberos [Proceedings of the International Conference on Arabic-Romance linguistic interference and Parallel Extra-Iberians] (pp. 105–117). Zaragoza: Navarro & Navarro.

Ingham, B. (2003). Language survival in isolation: The Arabic Dialect of Afghanistan. In I. Ferrando, & J. J. S. Sandoval (Eds.), *Proceedings of the 5th Conference of the Association Internationale de Dialectologie Arabe (AIDA)* (pp. 21–37). Cádiz: Universidad de Cádiz Publicationes.

Johnstone, T.M. (1968). The verbal affix -k in spoken Arabic. *Journal of Semitic Studies*, 13(2), 249–252. DOI: 10.1093/jss/13.2.249

Mailhammer, R. (2007). Islands of resilience: The history of the German strong verbs from a systemic point of view. *Morphology*, 17, 77–108. DOI: 10.1007/s11525-007-9113-y

McChesney, R. D. (1998). Central Asia's place in the Middle East: Some historical considerations. In D. Menashri (Ed.), *Central Asia Meets the Middle East* (pp. 25–51). London: Frank Cass Publishers.

Mohamed, A. M. (1992). *Bukhara fii ṣadr al-Islām* [Bukhara in the heart of Islam] Cairo: Dar al-fikr al-ʕarbii.

Owens, J. (2006). *A linguistic history of Arabic*. Oxford, New York: Oxford University Press. DOI: 10.1093/acprof:oso/9780199290826.001.0001

Ratcliffe, R. R. (2004). Bukhara Arabic: A metatypized dialect of Arabic in Central Asia. In E. Csató, B. Isaksson, & C. Jahani (eds.), *Linguistic convergence and areal diffusion. Case studies from Iranian, Semitic and Turkic* (pp. 141–159). London: Routledge.

Retsö, J. (1988). Pronominal suffixes with -n(n)- in Arabic dialects and other Semitic languages. *Zeitschrift für arabische Linguistik [Journal of Arabic Linguistics]*, 18, 77–94.

Tsereteli, G. (1956). *The Arabic dialects of Central Asia* (Vol. 1: Bukhara Arabic Dialect). Tbilisi: Publishing House of the Academy of Sciences of the Georgian SSR.

Versteegh, K. (1984). Word order in Uzbekistan Arabic and universal grammar. *Orientalia Suecana*, 33(5), 443–453.

Vinnikov, I. N. (1962). *Slovar' dialekta bukharskikh arabov* [A dictionary of the dialect of the Bukhara Arabs]. Moscow, Leningrad: Palestinskij Sbornik 73.

Vinnikov, I. N. (1965). Materialien zur Grammatik der Kashka-Darjiner Araber: Pardigmen der Verbalformen [Materials of the Qashqadarya Arabs: verbal paradigms]. In S. Segert (Ed.), *Studia Semitica philologica necnon philosophica Ioanni Bakoš dicata* [Semitic philological and philosophical studies dedicated to Jan Bakos] (pp. 261–276). Bratislava/Leningrad: Vydalo Vydavatelstvo Slovenskej Akadémie.

Vinnikov, I. N. (1969). *Yazik ii folklor Buxarskix Arabov* [The language and folklore of the Bukhara Arabs]. Moscow: Nauka Press.

Windfuhr, G. L. (2004). Central Asian Arabic: The Irano-Arabic dynamics of a new perfect. In E. Csató, B. Isaksson, & C. Jahani (Eds.), *Linguistic convergence and areal diffusion. Case studies from Iranian, Semitic and Turkic* (pp. 111–123). London: Routledge.

Appendix: Bukhara Arabic Plat

% Bukhara Arabic Paradigm

% Keri Miller

% qasam wazan kala qol baka shadd
% ‘he divided’ ‘he weighed’ ‘he ate’ ‘he said’ ‘he cried’ ‘he tied’

ABBR 1 1C1S2C2S3C3S4C % use 1A for forms with all three stem consonants

CONJ	**pretI1Sg**	**pretI2mSg**	**pretI2fSg**	**pretI3mSg**	**pretI3fSg**
TEMPLATE	1A	1A	1A	1A	1A
qasam	-a-a-t	-a-a-t	-a-a-ti	-a-a-	-a-a-et
wazan	-a-a-t	-a-a-t	-a-a-ti	-a-a-	-a-a-et
kala	-a-ej-t	-a-ej-t	-a-ej-ti	-a-a-	-a-∅-et
qol	-u-∅-t	-u-∅-t	-u-∅-ti	-o-∅-	-o-∅-et
baka	-a-ej-t	-a-ej-t	-a-ej-ti	-a-a-	-a-∅-et
shadd	-a-ej-t	-a-ej-t	-a-ej-ti	-a-a-	-a-∅-et

CONJ	preteritI1Pl	preteritI2mPl	preteritI2fPL	preteritI3mPl	preteritI3fPl
TEMPLATE	1A	1A	1A	1A	1A
qasam	-a-a-na	-a-a-tu	-a-a-tin	-a-a-u	-a-a-in
wazan	-a-a-na	-a-a-tu	-a-a-tin	-a-a-u	-a-a-in
kala	-a-ej-na	-a-ej-tu	-a-ej-tin	-a-∅-u	-a-∅-in
qol	-u-∅-na	-u-∅-tu	-u-∅-tin	-o-∅-u	-o-∅-in
baka	-a-ej-na	-a-ej-tu	-a-ej-tin	-a-∅-u	-a-∅-in
shadd	-a-ej-na	-a-ej-tu	-a-ej-tin	-a-∅-u	-a-∅-in

CONJ	pretII1Sg	pretII2mSg	pretII2fSg	pretII3mSg	pretII3fSg
TEMPLATE	1A	1A	1A	1A	1A
qasam	-aD-a-t	-aD-a-t	-aD-a-ti	-aD-a-	-aD-a-et
wazan	-aD-a-t	-aD-a-t	-aD-a-ti	-aD-a-	-aD-a-et
kala	waD-a-∅-t	waD-a-∅-t	waD-a-∅-ti	waD-a-∅-	waD-a-∅-et
qol	!none	!none	!none	!none	!none
baka	-aD-ej-t	-aD-ej-t	-aD-ej-ti	-aD-a-	-aD-et-
shadd	-aD-a-dit	-aD-a-dit	-aD-a-dti	-aD-a-d	-aD-a-det

CONJ	preterII1Pl	preterII2mPl	preterII2fPL	preterII3mPl	preterII3fPl
TEMPLATE	1A	1A	1A	1A	1A
qasam	-aD-a-na	-aD-a-tu	-aD-a-tin	-aD-a-u	-aD-a-in
wazan	-aD-a-na	-aD-a-tu	-aD-a-tin	-aD-a-u	-aD-a-in
kala	waD-a-∅-na	waD-a-∅-tu	waD-a-∅-tin	waD-a-∅-u	waD-a-∅-in
qol	!none	!none	!none	!none	!none
baka	-aD-ej-na	-aD-ej-tu	-aD-ej-tin	-aD-u-	-aD-in-
shadd	-aD-a-dna	-aD-a-dtu	-aD-a-dtin	-aD-a-du	-aD-a-din

CONJ	pretVII1Sg	pretVII2mSg	pretVII2fSg	pretVII3mSg	pretVII3fSg
TEMPLATE	1A	1A	1A	1A	1A
qasam	in-a-a-t	in-a-a-t	in-a-a-ti	in-a-a-	in-a-a-et
wazan	in-a-a-t	in-a-a-t	in-a-a-ti	in-a-a-	in-a-a-et
kala	inwa-a-∅-t	inwa-a-∅-t	inwa-a-∅-ti	inwa-a-∅-	inwa-a-∅-et
qol	!none	!none	!none	in-o-∅-	in-o-∅-et
baka	!none	!none	!none	!none	!none
shadd	in-a-ej-t	in-a-ej-t	in-a-∅-ti	in-a-a-	in-a-∅-et

CONJ	preterVII1Pl	preterVII2mPl	preterVII2fPL	preterVII3mPl	preterVII3fPl
TEMPLATE	1A	1A	1A	1A	1A
qasam	in-a-a-na	in-a-a-tu	in-a-a-tin	in-a-a-u	in-a-a-in
wazan	in-a-a-na	in-a-a-tu	in-a-a-tin	in-a-a-u	in-a-a-in
kala	inwa-a-∅-na	inwa-a-∅-tu	inwa-a-∅-tin	inwa-a-∅-u	inwa-a-∅-in
qol	!none	!none	!none	in-o-∅-u	in-o-∅-in
baka	!none	!none	!none	!none	!none
shadd	in-a-∅-na	in-a-∅-tu	in-a-∅-tin	in-a-∅-u	in-a-∅-in

CONJ	imperfI1Sg	imperfI2mSg	imperfI2fSg	imperfI3mSg	imperfI3fSg
TEMPLATE	1A	1A	1A	1A	1A
qasam	ma-∅-u-	tu-∅-u-	tu-∅-u-i:n	mu-∅-u-	tu-∅-u-
wazan	maw-∅-i-	tu-∅-i-	tu-∅-i-i:n	mu-∅-i-	tu-∅-i-
kala	mo-e-∅-	to-e-∅-	to-∅-∅-i:n	mijo-e-∅-	to-e-∅-
qol	ma-u-∅-	mit-u-∅-	mit-u:-∅-i:n	me-u-∅-	ti-u-∅-
baka	ma-∅-∅-i	ti-∅-∅-i	ti-∅-∅-i:n	mi-∅-∅-i	ti-∅-∅-i
shadd	ma-i-∅-	mit-i-∅-	mit-i-∅-i:n	mi-i-∅-	mit-i-∅-

CONJ	imperfI1Pl	imperfI2mPl	imperfI2fPl	imperfI3mPl	imperfI3fPl
TEMPLATE	1A	1A	1A	1A	1A
qasam	nu-∅-u-	tu-∅-u-u:n	tu-∅-u-in	mu-∅-u-u:n	mu-∅-u-in
wazan	new-∅-i-	tu-∅-i-u:n	tu-∅-i-in	mu-∅-i-u:n	mu-∅-i-in
kala	no-e-∅-	to-∅-∅-u:n	to-∅-∅-in	mijo-∅-∅-u:n	mijo-∅-∅-in
qol	min-u-∅-	mit-u-∅-u:n	mit-u-∅-in	me-u-∅-u:n	me-u-∅-in
baka	ni-∅-∅-i	ti-∅-∅-u:n	ti-∅-∅-in	mi-∅-∅-u:n	mi-∅-∅-in
shadd	min-i-∅-	mit-i-∅-u:n	mit-i-∅-in	mi-i-∅-u:n	mi-i-∅-in

CONJ	imperfII1Sg	imperfII2mSg	imperfII2fSg	imperfII3mSg	imperfII3fSg
TEMPLATE	1A	1A	1A	1A	1A
qasam	ma-uD-u-	mit-uD-u-	mit-uD-u-i:n	mu-uD-u-	mit-uD-u-
wazan	ma-uD-u-	mit-uD-u-	mit-uD-u-i:n	mu-uD-u-	mit-uD-u-
kala	ma-uD-u-	mit-uD-u-	mit-uD-u-i:n	mu-uD-u-	mit-uD-u-
qol	!none	!none	!none	!none	!none
baka	ma-iD-∅-i	mit-iD-∅-i	mit-iD-∅-i:n	mi-iD-∅-i	mit-iD-∅-i
shadd	ma-iD-∅-id	mit-iD-∅-id	mit-iD-∅-idi:n	mi-iD-∅-id	mit-iD-∅-id

CONJ	imperfII1Pl	imperfII2mPl	imperfII2fPl	imperfII3mPl	imperfII3fPl
TEMPLATE	1A	1A	1A	1A	1A
qasam	min-uD-u-	mit-uD-u-u:n	mit-uD-u-in	mu-uD-u-u:n	mu-uD-u-in
wazan	min-aD-a-	mit-aD-a-u:n	mit-aD-a-in	mi-aD-a-u:n	mi-aD-a-in
kala	min-uD-u-	mit-uD-u-u:n	mit-uD-u-in	mi-uD-u-u:n	mi-uD-u-in
qol	!none	!none	!none	!none	!none
baka	min-iD-∅-i	mit-iD-∅-u:n	mit-iD-∅-in	mi-iD-∅-u:n	mi-iD-∅-in
shadd	min-iD-∅-id	mit-iD-∅-idu:n	mit-iD-∅-idin	mi-iD-∅-idu:n	mi-iD-∅-idin

CONJ	imprfVII1Sg	imprfVII2mSg	imprfVII2fSg	imprfVII3mSg	imprfVII3fSg
TEMPLATE	1A	1A	1A	1A	1A
qasam	min-a-a-	tin-a-a-	tin-a-a-i:n	min-a-a-	tin-a-a-
wazan	min-a-a-	tin-a-a-	tin-a-a-i:n	min-a-a-	tin-a-a-
kala	minwa-a-∅-	tinwa-a-∅-	tinwa-a-∅-i:n	minwa-a-∅-	tinwa-a-∅-
qol	!none	!none	!none	min-o-∅-	tin-o-∅-
baka	!none	!none	!none	!none	!none
shadd	min-a-∅-	tin-a-∅-	tin-a-∅-i:n	min-a-∅-	tin-a-∅-

CONJ	imprfVII1Pl	imprfVII2mPl	imprfVII2fPl	imprfVII3mPl	imprfVII3fPl
TEMPLATE	1A	1A	1A	1A	1A
qasam	nin-a-a-	tin-a-a-u:n	tin-a-a-in	min-a-a-u:n	min-a-a-in
wazan	nin-a-a-	tin-a-a-u:n	tin-a-a-in	min-a-a-u:n	min-a-a-in
kala	ninwa-a-∅-	tinwa-a-∅-u:n	tinwa-a-∅-in	minwa-a-∅-u:n	minwa-a-∅-in
qol	!none	!none	!none	min-o-∅-u:n	min-o-∅-in
baka	!none	!none	!none	!none	!none
shadd	nin-a-∅-	tin-a-∅-u:n	tin-a-∅-in	min-a-∅-u:n	min-a-∅-in

CONJ	imperativeI2mSg	imperativeI2fSg	imperativeI2mPl	imperativeI2fPl
TEMPLATE	1A	1A	1A	1A
qasam	u-∅-u-	u-∅-u-i	u-∅-u-u	u-∅-u-in
wazan	i-∅-i-	i-∅-i-i	i-∅-i-u	i-∅-i-in
kala	u-u-∅-	u-∅-∅-i	u-∅-∅-u	u-∅-∅-in
qol	-u-∅-	-u-∅-i	-u-∅-u	-u-∅-in
baka	i-∅-∅-i	i-∅-∅-i	i-∅-∅-u	i-∅-∅-in
shadd	-i-∅-	-i-∅-i	-i-∅-u	-i-∅-in

CONJ	imperativeII2mSg	imperativeII2fSg	imperativeII2mPl	imperativeII2fPl
TEMPLATE	1A	1A	1A	1A
qasam	-uD-u-	-uD-u-i	-uD-u-u	-uD-u-in
wazan	!none	!none	!none	!none
kala	!none	!none	!none	!none
qol	!none	!none	!none	!none
baka	-iD-∅-i	-iD-∅-i	-iD-∅-u	-iD-∅-in
shadd	-i-∅-id	-i-∅-idi	-i-∅-idu	-i-∅-idin

CONJ	subCnd1Sg	subCnd2mSg	subCnd2fSg	subCnd3mSg	subCnd3fSg
TEMPLATE	1A	1A	1A	1A	1A
qasam	a-∅-i-	ti-∅-i-	ti-∅-i-i:n	ji-∅-i-	ti-∅-i-
wazan	aw-∅-a-	tiw-∅-a-	tiw-∅-a-i:n	jiw-∅-a-	tiw-∅-a-
kala	o-e-∅-	to-e-∅-	to-e-∅-i:n	jo-e-∅-	to-e-∅-
qol	a-u-∅-	it-u-∅-	it-u-∅-i:n	ji-u-∅-	it-u-∅-
baka	a-∅-∅-i	ti-∅-∅-i	ti-∅-∅-i:n	ji-∅-∅-i	ti-∅-∅-i
shadd	a-i-∅-	it-i-∅-	it-i-∅-i:n	ji-i-∅-	it-i-∅-

CONJ	subCnd1Pl	subCnd2mPl	subCnd2fPl	subCnd3mPl	subCnd3fPl
TEMPLATE	1A	1A	1A	1A	1A
qasam	ne-∅-i-	ti-∅-i-u:n	ti-∅-i-in	ji-∅-i-u:n	ji-∅-i-in
wazan	ne-∅-i-	ti-i-∅-u:n	ti-i-∅-in	ji-i-∅-u:n	ji-i-∅-in
kala	no-e-∅-	to-e-∅-u:n	to-e-∅-in	jo-e-∅-u:n	jo-e-∅-in
qol	ne-u-∅-	it-u-∅-u:n	it-u-∅-in	ji-u-∅-u:n	ji-u-∅-in
baka	ni-∅-∅-i	ti-∅-∅-u:n	ti-∅-∅-in	ji-∅-∅-u:n	ji-∅-∅-in
shadd	ni-i-∅-	it-i-∅-u:n	it-i-∅-in	ji-i-∅-u:n	ji-i-∅-in

CONJ	actPrtImSg	actPrtIfSg	actPrtImPl	actPrtIfPl
TEMPLATE	1A	1A	1A	1A
qasam	-o:-i-	-o:-∅-a	-o:-∅-i:n	-o:-∅-a:t
wazan	-o:-i-	-o:-∅-a	-o:-∅-i:n	-o:-∅-a:t
kala	-a:-i-	-a:-∅-a	-a-∅-i:n	-a-∅-a:t
qol	-o:ji-∅-	-oj-∅-a	-oj-∅-i:n	-oj-∅-a:t
baka	-a:-i:-	-a:-∅-a	-a:-∅-i:n	-a:-∅-a:t
shadd	-o-∅-	-o-∅-a	-o-∅-i:n	-o-∅-a:t

CONJ	actPrtIImSg	actPrtIIfSg	actPrtIImPl	actPrtIIfPl
TEMPLATE	1A	1A	1A	1A
qasam	-uD-u-	-uD-u-a	-uD-u-i:n	-uD-u-a:t
wazan	-uD-u-	-uD-u-a	-uD-u-i:n	-uD-u-a:t
kala	wuD-u-∅-	wuD-u-∅-a	wuD-u-∅-i:n	wuD-u-∅-a:t
qol	!none	!none	!none	!none
baka	-iD-∅-i:	-iD-∅-a	-iD-∅-i:n	-iD-∅-a:t
shadd	-iD-i-d	-iD-i-d	-iD-i-di:n	-iD-i-da:t

CONJ	actPrtVIImSg	actPrtVIIfSg	actPrtVIImPl	actPrtVIIfPl
TEMPLATE	1A	1A	1A	1A
qasam	mun-u-u-	mun-u-∅-a	mun-u-∅-i:n	mun-u-∅-a:t
wazan	mun-u-u-	mun-u-∅-a	mun-u-∅-i:n	mun-u-∅-a:t
kala	munwu-u-∅-	munwu-∅-∅-a	munwu-∅-∅-i:n	munwu-∅-∅-a:t
qol	min-u:ji-∅-	min-u:j-∅-a	min-u:j-∅-i:n	min-u:j-∅-a:t
baka	!none	!none	!none	!none
shadd	min-i-∅-	min-i-∅-a	min-i-∅-i:n	min-i-∅-a:t

CONJ	inflectedAP1mSg	inflectedAP1fSg	inflectedAP2mSg	inflectedAP2fSg
TEMPLATE	1A	1A	1A	1A
qasam	-o:-∅-inni	-o:-∅-etinni	-o:-∅-innak	-o:-∅-inki
wazan	-o-∅-inni	-o-∅-etinni	-o-∅-innak	-o-∅-inki
kala	-a:-∅-inni	-a:-∅-tinni	-a:-∅-innak	-a:-∅-inki
qol	-oj-∅-inni	-oj-∅-etinni	-oj-∅-innak	-oj-∅-inki
baka	-a:-∅-inni	-a:-∅-etinni	-a:-∅-innak	-a:-∅-inki
shadd	-o-∅-inni	-o-∅-etinni	-o-∅-inna	-o-∅-inki

CONJ	inflectedAP1Pl	inflectedAP2mPl	inflectedAP2fPl
TEMPLATE	1A	1A	1A
qasam	-o:-∅-inna	-o:-∅-inkum	-o:-∅-inkin
wazan	-o-∅-inna	-o-∅-inkum	-o-∅-inkin
kala	-a:-∅-inna	-a:-∅-inkum	-a:-∅-inkin
qol	-oj-∅-inna	-oj-∅-inkum	-oj-∅-inkin
baka	-a:-∅-inna	-a:-∅-inkum	-a:-∅-inkin
shadd	-o-∅-inna	-o-∅-inkum	-o-∅-inkin

CONF	inflInfin1Sg	inflInfin2mSg	inflInfin2fSg	inflInfin3mSg	inflInfin3fSg
TEMP	1A	1A	1A	1A	1A
qasam	-o-∅-aha:ni	-o-∅-aha:nak	-o-∅-aha:nki	-o-∅-aha:nu	-o-∅-aha:na
wazan	-o-∅-aha:ni	-o-∅-aha:nak	-o-∅-aha:nki	-o-∅-aha:nu	-o-∅-aha:na
kala	-a-∅-aha:ni	-a-∅-aha:nak	-a-∅-aha:nki	-a-∅-aha:nu	-a-∅-aha:na
qol	-oj-∅-aha:ni	-oj-∅-aha:nak	-oj-∅-aha:nki	-oj-∅-aha:nu	-oj-∅-aha:na
baka	-a-∅-aha:ni	-a-∅-aha:nak	-a-∅-aha:nki	-a-∅-aha:nu	-a-∅-aha:na
shadd	-a-∅-aha:ni	-a-∅-aha:nak	-a-∅-aha:nki	-a-∅-aha:nu	-a-∅-aha:na

CONJ	inflInfin1Pl	inflInfin2mPl	inflInfin2fPl
TEMPLATE	1A	1A	1A
qasam	-a-∅-aha:na	-a-∅-aha:nkum	-a-∅-aha:nkin
wazan	-o-∅-aha:na	-o-∅-aha:nkum	-o-∅-aha:nkin
kala	-a-∅-aha:na	-a-∅-aha:nkum	-a-∅-aha:nkin
qol	-oj-∅-aha:na	-oj-∅-aha:nkum	-oj-∅-aha:nkin
baka	-a-∅-aha:na	-a-∅-aha:nkum	-a-∅-aha:nkin
shadd	-a-∅-aha:na	-a-∅-aha:nkum	-a-∅-aha:nkin

CLASS consonant b c d f g j k l m n p q r s t v z

SANDHI w [:consonant:] => $1

SANDHI D [:consonant:] => $1 $1 % doubling

LEXEME divide	qasam	1:q	2:s	3:m
LEXEME weigh	wazan	1:w	2:z	3:n
LEXEME eat	kala	1:k	2:l	3:∅
LEXEME say	qol	1:q	2:l	3:∅
LEXEME cry	baka	1:b	2:k	3:∅
LEXEME tie	shadd	1:sh	2:dd	3:∅

Semantic/Pragmatics

Terms of endearment and anger in Levantine Arabic

Praying for and against someone*

Mohammad A. Mohammad
University of Texas at Austin

This article provides a detailed description of a specific structure in Arabic referred to as "God-wishes" by Fergusson (1997). In spite of their singular third person agreement features, the article argues that they are actually imperatives in disguise. Evidence supporting this conclusion is cited across the paper. The data described is derived from hundreds of hours of television series in Levantine Arabic broadcast in Ramadan over the last fifteen or so years.

Keywords: God-wishes, Palestinian Arabic, Levantine Arabic, imperative, negation.

Introduction

In a phenomenon probably unique to Arabic, two interlocutors in a discourse use a kinship title that establishes a hierarchy between the two. To clarify, consider (1) and (2) in which a grandson gives his grandmother something:

(1) a. Child:
ya sitti xuði
ya grandma take...
'Grandma, here it is.'

* I am grateful to the audience at the Arabic Linguistics Symposium XXVI, especially Elabbas Benmamoun, Reem Khamis-Dakwar and Hamid Ouali for their insightful comments. I am particularly grateful for Enam Al-Wer for her patient and careful reading of the manuscript. Her suggestions and corrections were numerous and valuable. Remaining errors are mine. My biggest debt, however, goes to the editors of this volume and to the anonymous reviewers.

b. Grandmother:
ya sitti ʔallah yeʕṭīk
ya grandma Allah give.you
'Sweetie, may Allah give you (whatever you wish').'

In (1) the grandmother is happy with the grandson. Her response shows this in two ways: (a) by repeating the exact form of address that the grandson uses; and (b) by praying for the grandson. Consider (2):

(2) a. Child:
ya sitti xuði
ya grandma take...
'Grandma, here it is.'

b. Grandmother:
hāt ʔallah yōxðak
give Allah take.you
'Hand it over, may Allah take you.'

In (2) the grandmother is angry with the grandson. Her response shows her displeasure with him in two ways: (a) by not repeating the exact form of address that the grandson uses; and (b) by praying against the grandson.

The kinship forms of address typically used are *yamma/yamu/mama/yāmu* 'mother', etc.; *yaba/baba/yabi* 'father', etc.; *ʕam* 'paternal uncle'; *ʕammeh* 'paternal aunt'; *xāl* 'maternal uncle'; *xāleh*, 'maternal aunt'; *sitt/ʤiddah* 'grandmother; and *sīd/ʤidd* 'grandfather.' These forms clearly exhibit a hierarchy between the two interlocutors. The more interesting part, however, is the form of prayer used in the (b) examples.

Ferguson (1997) refers to the forms exhibited in (1) and (2) above as "God-wishes." To the best of my knowledge, Ferguson's paper is the first and only paper to discuss these expressions. Noting the complexity of these forms, he stated that it would take years to write their grammar. After nearly four years of 'fieldwork', the present paper provides a grammar of these expressions.

The paper is organized into three parts. In part one, I provide some preliminary remarks on the grammar needed for the discussion to follow. In part two, the bulk of the paper, I provide a detailed description of God-wishes and propose their grammar. In the third and last part, I suggest an explanation for the unique grammar of God-wishes.

A note on the data

All the data reported in this paper, except for the ungrammatical examples, are authentic and attested. The majority of the examples cited come from Palestinian

Arabic, taken from personal observation in which I was a witness to the exchange. The phenomena are consistent in a large number of Arabic varieties. The facts were confirmed from viewing approximately two thousand episodes of TV series broadcast during Ramadan over the last fifteen years. The Arabic varieties used in these videos insofar as I was able to establish are: Modern Standard Arabic, rural and urban Palestinian Arabic, Damascene Arabic, Aleppo Arabic, Bedouin Jordanian Arabic, Jordanian Arabic (mixture of Ammani and regional varieties), Syrian Bedouin Arabic, Emirati Bedouin Arabic, Emirati urban Arabic, Saudi Arabic, Cairene Arabic, Saʕīdi Arabic (Upper Egyptian), Syrian Coast Arabic (presumably around Latakia and Tartous), Syrian Horani Arabic, Syrian Druze Arabic, Lebanese Arabic and Iraqi Arabic. I have been unsuccessful in securing TV series from Libya, Tunisia and Morocco. I did, however, consult with speakers of these dialects and I was assured that the data reported in this paper are identical in these Arabic varieties.

2. Verb types in Arabic

In order to set the stage for this paper, I classify the Arabic verb into the following classes, leaving aside any issues of tense, aspect and mood (for a detailed discussion and description see Brustad, 2000 and Holes, 2004):

a. verbs that show morphological changes by suffixes only (referred to in grammar books of Arabic as past tense verb);
b. verbs with both prefixes and suffixes (referred to in Arabic grammar books as *almuDāreʔ almarfūʔ)*;
c. verbs that occur in environments in which a lexical item requires a specific form of the verb (referred to in Arabic grammar books as *almuDāreʔ almansūb)*;
d. verbs in which a prefix is deleted accompanied by some other changes depending on a form of the verb after the prefix is deleted (referred to as the imperative *alʔamr)*.

Past tense verbs

There are small differences in this form across Arabic dialects. The differences among these dialects bear no effect on the issues raised by this paper. What is needed here is data on how these verbs are negated. Of all the dialects that I am familiar with and in all the published work, I am not aware of a single Arabic dialect that negates this form with a simple *lā*. They all use *ma* in the negative here. Some dialects use the suffix *-ʃ* together with *ma*. The rarest among the dialects are those that negate with *-ʃ* without the *ma*. That is, we have the following paradigm (see Brustad, 2000 and Benmamoun, 1992, 1997, 2000 for details):

(3) a. ʔadʒa
came.3SM
'He came.'

b. ma-dʒā-ʃ
NEG-came-NEG
'He did not come.'

c. ma-ʔidʒa
NEG-came
'He did not come.'

d. ʔadʒā-ʃ
came,NEG
'He did not come.'

e. *la-ʔidʒa
NEG.come
'He did not come.'

Example (e) is ungrammatical in all the dialects as a simple negative sentence. However, it is grammatical in extremely rare cases, when a God-wish is expressed in Bedouin Arabic, as we will see below.

AlmuDāreʔ almarfūʔ

Modern Arabic dialects can be divided into two types: those with a prefix such as *b-* or *ka-* and those with no such prefix. For ease of reference, I refer to the first group as the *baka* (which means in some Palestinian varieties 'still', 'remains') dialects and the second group will be referred to as the *bala* (which means in some Palestinian varieties 'without') dialects. To illustrate, consider:

(4) bedrus
studies.3SM
'He is studying'

(5) kayudrus
studies.3SM
'He is studying'

(6) yadrus
studies.3SM
'He is studying'

(4) and (5) are of the *baka* type while (6) is of the *bala* type. The crucial point here is that the forms in (4)–(6) are the forms that require no lexical trigger.

The negation patterns with present tense verb follow the same pattern briefly outlined above. As far as I know, all Arabic dialects have the same negation forms both in past tense and present tense verbs. Some may exhibit a different pattern in the imperative in that *lā* is permitted or required.

Word order

Unlike Classical Arabic or Modern Standard Arabic, modern Arabic dialects tolerate less word order variation. Not counting *Wh*-questions where OVS is permitted or required in some Arabic dialects, they all have SVO, some permit VSO, and fewer others also permit VOS.

3. God-wishes

As noted earlier, I follow Ferguson (1997) in naming the phenomena under discussion in this paper 'God-wishes'. As Ferguson noted, these expressions come in pairs. I will term the first member of the pair *trigger* and the second *response*. The trigger may be verbal or an event not verbally or overtly articulated.

While Ferguson focused his attention on pairs that are usually found in social exchanges that may give the impression that the speakers are almost on automatic pilot. This paper goes beyond social niceties and provides a detailed description of a larger array of God-wishes. Consider first some typical expressions that one might encounter on a daily basis:

(7) a. maʕ essalāmeh
with safety
'(Proceed) with safety (goodbye).'

b. ʔallāh ysallmik
Allah grant safety.you.fs
'May Allah grant you (too) safety.'

(8) a. mabrūk
blessed (lit.)
'Congratulations!'

b. ʔallāh ybārek fīk
Allah bless.you.sm
'May Allah bless you (too). (Thanks)'

(9) a. ʔallāh yeʕtīk lʕāfyeh
Allah give the.health
'May Allah give you good health.'

b. ʔallāh yʕāfīk
Allah give good health.you.sm
'May Allah give you (too) good health.'

In the above examples the trigger is verbally expressed in (a) with response using the root used in the trigger (*slm* in (7), *brk* in (8) and *ʕfw* in (9)). Consider now the following where the trigger is an event given in (a) and the verbal response given in (b):

(10) a. Someone sympathetically responds to a person who loses a valuable item:

b. ʔallāh yʕawweðˤ ʕaleik
Allah compensate on.you
'May Allah compensate you.'

(11) a. A beggar asks for money:

b. ʔallāh yyasser lak
Allah provide for.you.sm
'May Allah provide for you.'

(12) a. A person is about to ask someone for a favor:
b. ʔallāh yerðˤa ʕaleik +request
Allah be pleased on.you.sm
'May Allah be pleased with you. (would you do ...)'

While the above forms are taken from Palestinian Arabic, they are universal across the Arabic dialects I checked. The verbs and/or their pronunciation obviously may vary, but the forms are the same. The crucial point here is that the form of the verb is identical in both the *baka* and the *bala* dialects.

Consider now the following where the verb is in the past tense form:

(13) a. taqabbal ʔallāh
accepted Allah
'May Allah accept (your prayers).'

b. minna waminkum sˤāleħ ilʔaʕmāl
from.us and.from.you.p good deeds
'May Allah accept good deeds from both of us.'

(14) a. la tiʕibt
not tired
'May you not be tired.'

b. la nidimt
not regretted
'May you never regret anything!'

The above examples, unlike the ones preceding them, are rare, not productive and tend to have a Classical Arabic flavor. Example (14), however, is a greeting expression found in Jordanian Bedouin Arabic. I found this example in two separate Jordanian Bedouin Ramadan series: *Biyareg Alʕarba* 'The Biyaregs of Alʕarba' and *Qabāʔel Aʃ-ʃarq* 'Tribes of the East.'

Beyond social niceties

The examples described in the previous section are used on specific social situations. In this section, I describe the phenomena outside the confines of social etiquette. I begin with kinship terms. We look at exchanges in which (a), the junior member of the pair, provides the trigger and the senior member provides the response. Consider first the following:

(15) Junior: yāba …
Dad

Senior: naʕam yāba
Yes Dad

(16) Junior: yamma
Mom

Senior: xeir yamma
good Mom

(17) Junior: xāli hatlak nusˁ leira
uncle give half dinar
'Uncle, give me half a dinar.'

Senior: wallah ya xāli mfalles
by.allah ya uncle.my penniless
'I swear, I am penniless.'

The interesting thing to note about the above exchanges is that both parties use the same form of address belonging to the senior member. In these examples the senior member is happy with the junior member. Consider now the following when the senior member is unhappy with the junior member:

(18) Junior: yāba … yāba … yāba …
Dad Dad Dad

Senior: walak baw lli yexlaʕ nīʕak. ʃu biddak
walak baw which break jaw.your what want.you
'walak, may a baw break your jaw. What do you want?'[1]

(19) a. marħaba yaxālti
hello ya aunt.my
'Hi, aunt'

b. ya mi:t marħaba ya xālti
ya hundred hello ya aunt.my
'One hundred hellos to you!'

c. xill yxill ʕðˁāmak
xill loosen /break bones.your.MS
'May a xill loosen your bones!'

d. *xill beyxill ʕðˁāmak
xill loosen /break bones.your.MS
'May a xill loosen your bones!'

The above examples I heard from Palestinians living in Jordan. Note five interesting things here regarding (c). First, unlike (b), the adult did not repeat the form of address, thus, indicating her displeasure. Second, she took the form of address *xālti* and derived from it a non-existing and nonsensical verbal noun, *masˁdar*, *xill* and a verb '*yxill*' that also does not exist in the dialect. Third, she uses an indefinite noun as a *mubtadaʔ* 'subject of a verbless sentence' where a definite noun is typically expected. Fourth, she uses the *bala* form of the verb by dropping the b-prefix, in a *baka* dialect. Fifth, if the *baka* form is used, the outcome is unacceptable as a wish (see (d) above).

Consider the following example from Upper Egyptian 'sʕaʕi:di' Arabic taken from the TV series *Sheikh Alʕarab Hammam* 'Hammam, the Sheik of the Arabs.' Example (20) is said by Hammam's first and older wife to his second and younger wife when the latter addressed the former using the form *xālti* 'my aunt':

(20) xulxāl yxalxel ʕdˁāmik ʕan laħmik
anklet (lit.) loosen bones.your.FS from flesh.your.FS
'May an anklet separate your bones from your flesh!'

1. The word 'baw' is non-existent and nonsensical. The context provides the interpretation that it refers to something that can break a jaw. The expression 'walak' can be used as a form of insult or as a form of affection, depending on the context.

Unlike (19), (20) contains lexical items that exist in the dialect. However, their literal meaning is somehow suspended or expanded. An anklet is not known to separate bones from the flesh. With this proviso in mind, it seems that all five remarks on the Palestinian Arabic example in (19) equally apply to (20).

Fully fledged God-wishes

In this section I provide a description of God-wishes in which the first member of the pair provides the trigger and the lexical base for the response.

Proper names as sources

Most Arabic proper names are derived from meaningful roots with the word used as a proper name maintaining its dictionary meaning. Consider the following example where the source is the proper name *tawfīq* 'success':

(21) ʔallah ywafgak ya tufīg
Allah grant.you with success ya Tawfiq
'Tawfiq, may Allah grant you success.'

In (21), the response in the pair responds to something that *Tawfiq*, the trigger, did. Here, the response is derived from the name of the trigger.

Consider now the following example taken from the Syrian TV series *AlKhirbeh* 'The Village' in which the character also named *Tawfiq* does something deemed objectionable by the speaker. The speaker takes the root *wfq* as his base:

(22) ʔallah lā ywafqak ya tufīq
Allah not grant.you with success ya Tawfiq
'Tawfiq, may Allah never grant you success.' (*alxirbeh*, Syrian Druze Arabic in a village presumably in the Suwaida area)

In the same series, a character commenting on a woman whose name is *raħmeh* 'mercy' says:

(23) hadi raħmeh ʔallah lā yirħamha
this Rahmeh Allah not have mercy.her
'This is Rahmeh (mercy), may Allah not grant her mercy.'

Example (23) is an instance where the speaker's tone of voice expresses extreme sarcasm which is also encoded in the God-wish. The speaker takes the root *rħm* as his derivational base.

The following example from Palestinian Arabic uses the *sʕd* root from which the proper name is derived as the base:

(24) wein rāħ saʕīd ʔallah lā yisiʕdo
Where went Said Allah not make happy.him
'Where did Said (lit. happy) go? May Allah never grant him happiness.'

The last example of this category is the comment about a prostitute named *Radiyyeh* (lit. content):

(25) radˤiyyeh ʔallah lā yirdˤa ʕaleiki mnein
Raddiye Allah not be pleased on.you how
biddu yerdˤa ʕaleiki wʕendek kul hazzabayen
want be pleased with.you and.you have all these.customers
'Radiyyeh, may Allah never be pleased with you. How can he be pleased with you when you have all these customers?' (from the TV series *Asad Alwarraq*, Damascene Arabic)

The above examples show us how the response is based on taking a proper name, the trigger, and derives from its root a verbal God-wish response.

Nouns and verbs sources

Consider the following example from '*Tribes of the East*, (Jordanian Bedouin Arabic)' in which the younger sister is reprimanding her sister over her socially unacceptable behavior:

(26) a. wallāhi inneč msˤībah ya bint abūy
By Allah you disaster ya daughter father.my
'I swear you are a catastrophe, sister.'

b. msˤībah tsˤībeč
Catastrophe strikes.you
'May a catastrophe strike you.'

In (26) (b) the response repeats the noun provided by the trigger and uses the corresponding verb as a comeback.

In the following example, a husband and wife are discussing a problem their tribe is facing. The problem is that there is a lone warrior raiding the tribe in search of vengeance. The husband, referring to the warrior as 'the night wolf', states that the wolf is about to be captured by one of the many ambushes put in place. The wife responds:

(27) a. ʕugub mā daraʕ bilɣanam
after when burped.3SM with.the.sheep
'Now that he is burping (after he devoured the sheep) (i.e. isn't it too late?)'

b. widʒaʕ yedraʕ iθmeč
pain burps.3SM mouth.your
'May pain break your mouth.' (*Biyāreg Alʕarba 'The Biyreqs of Alarba'*, Jordanian Bedouin Arabic)

The husband shows his displeasure with his wife's sarcasm in (b). Both husband and wife use the same verb: The wife uses it in its dictionary meaning, and the husband adds a twist to the verb meaning.

The event as the source

Consider now the following examples where the trigger is an event with the God-wish acting as a commentary on the event:

(28) ʔallah yzillhon
Allah humiliate.them
'May Allah humiliate them.' (Syrian TV series *Xān Alharīr 'The Silk Market'*, Aleppo Arabic)

(29) ʔallah yiblāk bʕaglak
Allah afflict.you brain.your
'May Allah afflict you with your brain.' (Jordanian Bedouin TV series *Rās ɣleis 'Ghleis' Head'*, Part 1)

(30) ʃofto ʔallah yugsˤof ʕomro
saw.1s.him Allah cut short life.his
'I saw him. May Allah cut his life short.' (From the TV series *Finjān Addam 'The Cup of Blood'*, Jordanian Maani Arabic).

(31) xiðha ʔallah lā yruddak inta wayyāha
take.her Allah not bring.you you and.her
'Take her and to Hell with both of you!' (from the TV series *Toug*, Syrian Bedouin Arabic).

(32) ya rabb ykūn māt
ya Lord be dead
'I pray to the Lord that he is dead.' (from the TV series *Asad alwarraq*, Damascene Arabic).

By looking at the above examples and hundreds like them, the following generalizations can be made:

a. word order used is always SVO;
b. The tense of the verb in the response is always in the *bala* form, where the prefix b- is dropped.

Word order in God-wishes

As remarked above, word order in God-wishes is always SVO in all the dialects investigated, regardless of whether the dialect permits VSO or not. Consider the following paradigm from Palestinian Arabic:

(33) a. ʔallah yifridʒhā
Allah provide a way out
'May Allah provide a way out.'

b. *yifridʒhā ʔallah
provide a way out Allah
'May Allah provide a way out.'

c. ʔallah bifridʒhā
Allah provide a way out
'Allah will provide a way out.'

b. bifridʒhā ʔallah
provide a way out Allah
'Allah will provide a way out.'

The above paradigm demonstrates two things: Word order is rigid in God-wishes and using the *bala* form in a VSO order results in an ungrammatical output. By contrast, the *baka* form allows both word orders. Consider now the same paradigm in Jordanian Bedouin Arabic where the examples occur in the TV series *Biyareg alarba*:

(34) a. ʔallah yafridʒhā
Allah provide a way out
'May Allah provide a way out.'

b. yafridʒhā ʔallah
provide a way out Allah
'May Allah provide a way out.'

The dialect in question is a *bala* dialect and freely permits VSO. In the TV series, SVO order always occurs in a God-wish with VSO allowing the reading where the speaker is making a statement, not a God-wish. Thus, it seems each dialect has its own strategy: rely on word order and/or the form of the verb to avoid ambiguity. Palestinian Arabic and the majority of Levantine dialects rely on both word order and the *bala* form to encode the God-wishes whereas Bedouin Arabic and presumably other *bala* dialects rely on word order to encode them.

On the nature of bala verbs

The *bala* form in the *baka* dialects is not reserved for these constructions. It is abundant in Arabic dialects (for an extensive discussion of these forms, see Brustad 2000). The *bala* form in Levantine is required when preceded by one of the following:

(35) *biddi* 'I want;' *lāzem* 'must;' *Darūri* 'necessary;' *mumken* 'possible;' *raħ* 'will;' *rāyeħ* 'on his way;' *masmūħ* 'permitted;' *mamnūʕ* 'prohibited;' *ħābeb* 'I feel like;' and many verbs expressing wishing, telling, permitting, etc. (for details, see Brustad, 2000).

Brustad (2000) notes one further use of the *bala* form. Consider the following from Palestinian Arabic:

(36) a. teʃrab gahweh
drink coffee
'Would you like some coffee?'

b. bteʃrab gahweh
drink coffee
'Are you drinking coffee?'

Note that if the *baka* form is used as in (b) above, the question force shifts to asking a person already drinking something with the speaker asking to find out whether the person is drinking coffee or something else.

The obvious question, then, is whether the *bala* form used in God-wishes and the those used in contexts mentioned in (35) is the same or different (for details see Brustad, 2000). The answer is that they are different. The God-wishes, as seen in the negative examples cited so far, use *lā*. If *ma* is used instead, the sentences become ungrammatical. If, on the other hand, *lā* is used with *bala* forms triggered by any member of the list in (35), the output is ungrammatical or marginal at best. Consider what happens if (36) is negated:

(37) a. lā tiʃrab gahweh
not drink coffee
'Do not drink coffee.'

b. ma tiʃrab gahweh
not drink coffee
'Why don't you drink coffee?'

c. ma tiʃrabeʃ gahweh
not drink.NEG coffee
'Do not drink coffee.'

Negation with *lā* renders the sentences imperative. Negating it with *ma* without the suffix *-ʃ* changes the sentence into some kind of a suggestion. Negating it with both *ma* and the *-ʃ* suffix provides an imperative sentence synonymous with the (a) sentence. This puts the God-wishes in a class of their own.

Are they idioms?

Now that I have established that God-wishes are a class by themselves, one may wonder if they are actually idioms. Typically idioms have the following properties: (a) they are not productive; (b) the meaning of the idiom cannot be deduced from its component parts; and (c) the idiomatic portion consists of the verb and its object (for a detailed discussion see Mohammad, 2000, and Aoun, Benmamoun and Chouerie, 2010). God-wishes do not exhibit any of these properties; they are very productive, their meaning can be determined from the meaning of their component parts and they never involve the verb and object as a single unit. Therefore, it can be safely concluded that God-wishes are not idioms.

Toward identifying God-wishes

In the preceding sections, I have established that God-wishes are different from what Brustad (2000) called subjunctive verbs. While the latter do share the physical form with God-wishes, they differ in two crucial properties: (a) subjunctive verbs are a consequence forced by the presence of Brustad's semi verbs; and (b) subjunctive verbs cannot be negated with *lā*. By contrast, God-wishes have no lexical trigger present and can only be negated with *lā*.

In order to identify God-wishes, let us take a closer look at their structure. Obviously on the surface they seem as if they are expressing a wish of the speaker. This impression is created due to the fact that the verb is in the third person singular with Allah as the subject, whether the subject is lexically present or not. To show that this is actually a false impression, consider the following examples:

(38) ʔallah yyesserlak bʤāh errasūl
Allah provide.you for.sake the.prophet
'May Allah provide you (a living, with success), for the Prophet's sake.'

(39) ʔallah yfikk ʔasirna karāmeh lamuħammad ya rabb
Allah undo captivity. our for.honor Muhammad ya Lord
'May Allah free us from captivity, in honor of Muhammad, oh dear Lord'
Rjāl lʕez, Damascene Arabic).

(40) ʔallah yensˁurkun bʤāh elʕadra
Allah grant victory.you for the sake the-Virgin
'May Allah grant you victory, for Virgin Mary's sake.' *Rjāl Lʕez 'Men of Glory'*, Damascene Arabic).

(41) ʔallah yrayyeħ balak bʤāh ʔallah min fōg
Allah put at ease mind for the sake Allah from above
'May Allah grant you peace of mind, for Allah' sake (who is) above.'

The sentences above consist of two parts: The first part contains the prayer/wish, and the second part contains a name for whose sake the speaker is soliciting Allah, the subject of the first part, to grant the wish. The most interesting and telling example is (41). At first glance, it sounds and looks like blasphemy. The speaker seems to imply that there are two Gods. Given the culture of the speaker (a Sunni Muslim), that cannot possibly be the case. The only explanation consistent with above assumptions is to assume that the above examples, and all forms of God-wishes, are, in reality, imperatives in disguise. Making this assumption explains the above facts and explains the appearance of the negative particle *lā*.

In support of the above conclusion, consider the following (probably rare) examples that I recorded from the Syrian TV series *Xān Elħarīr* (Aleppo Arabic), Part Two and from the Jordanian Bedouin series *Biyāreg Alʕarba*, respectively.

In the first example, the local imam is talking about someone whose only fault is that he drinks. In all other respects, he is not bad.

(42) ya rabb ya ʔimma tesˁelħu ya imma tuʔsˁuf ʕumru
ya God ya either repair.him.2SM ya or chop.2SM life.his
ya allah yesʕelħu ya imma yuʔsˁuf ʕumru
ya God repair.him.3SM ya or chop life.his
'Oh Lord, either lead him to the right way or cut his life short. May Allah either lead him to the right way or cut his life short.'

The rarity of the above example stems from the use of the vocative form *ya* and the use of the third person form of the verb. The expected form in all the verbal forms in the example above is second person singular. Thus, in spite of the example's rarity, it provides direct evidence in support of our conclusion that we are dealing with an imperative verbal form, albeit in the unexpected third person.

The second example is from the Jordanian Bedouin TV series *Biyāreg Alarba*. The characters are standing by the grave of a brutal sheikh's father. The sheikh asks them to recite *AlFāteHa 'The Opening Chapter of the Holy Qur'an'* for the benefit of his father's soul. One of the people present, an enemy of the sheikh, says (with the actor using voice over) the following prayer:

(43) allahumma lā terħam waldeihum madām
Lord NEG have mercy.2SM parents.their for as long
lana arðˤ ʕindahum lamma yruddūnha ʔallah
for.us land with.them until return.3MP.it Allah
lā ywaffighum allahumma unsˤurna
NEG grant success.3SM.them Lord grant victory.2SM.us
ʕaleihum beddunya wbelʔāxrah
over.them in.this world and.in.the.hereafter
'Dear Lord, never have mercy on their parents until they return our land they are holding. May Allah never grant them success. Lord grant us the upper hand over them both in this world and in the hereafter.'

Example (43) is a fascinating example in that it contains two instances of direct imperative: one negative and the other positive. In between these imperatives a God-wish is intercalated. It makes a great deal of sense to interpret the God-wish as an indirect imperative as previously suggested.

In order to explain the surprising conclusion above, we need to leave syntax and seek an explanation elsewhere. It is my claim here that the explanation can probably be stated best in functional terms. But before offering my explanation, it is perhaps useful to take a closer look at the wishes (in this case, a better word is prayer) when an imperative form is used. Consider:

(44) a. ya ʔallah ʔerħmna braħemtak
ya Allah have mercy.us with.mercy.your
'Dear Allah, have mercy on us.'

b. ʔerħmna braħemtak ya ʔallah
have mercy.us with.mercy.your ya Allah
'Dear Allah, have mercy on us.'

In (44) a direct imperative form of the verb is used. Note further that the name of Allah can occur either in sentence initial position or in sentence final position. Consider now a different form of imperative:

(45) a. ya ʔallah terħmna braħemtak
ya Allah have mercy.us with.mercy.your
'Dear Allah, have mercy on us.'

b. *terħamna braħemtak ya ʔallah
have mercy.us with.mercy.your ya Allah
'Dear Allah, have mercy on us.'

The form of the verb in (45) is the *bala* form in second person singular. The presence of the vocative *ya* and Allah's name must occur sentence initially to license this form of the verb. Note further that, as far as I know, no one but Allah may occupy the subject position of this type of imperative. I believe it is used as a softened imperative, more of a request than a command. I submit that the use of this softened imperative is dictated by politeness and respect for Allah.

We are now in a position to suggest an explanation for the use of third person verbs in God-wishes. This form of the verb is dictated by the fact that Allah is addressed indirectly since the direct addressee is someone else. Thus, this form is motivated by the two factors of politeness toward Allah and that the addressee is someone else. Put differently, there are two addressees in God-wishes: Allah and the listener.

Conclusion

In this paper, I have met Ferguson's challenge and provided a grammar of God-wishes. Ferguson was right in that it indeed took years to come up with a description. I have shown that God-wishes have the following properties:

i. Word order used is always SVO;
ii. The verb is always in the *bala* form;
iii. Unlike Standard Arabic, the past tense form of the verb is nearly totally missing;
iv. Only *lā* can be used to negate them;
v. Unlike idioms, they are productive; and, finally,
vi. They are imperatives in disguise.

References

Aoun, J. E., Benmamoun, E., & Choueiri, L. (2010). *The syntax of Arabic*. New York: Cambridge University Press.

Benmamoun, E. (1992). Inflectional and functional morphology: Problems of projection, representation and derivation. Unpublished Ph.D. dissertation, USC, Los Angeles.

Benmamoun, E. (1997). Licensing of negative polarity in Moroccan Arabic. *Natural Language and Linguistic Theory*, 15, 263–287. DOI: 10.1023/A:1005727101758

Benmamoun, E. (2000). *The feature structure of functional categories: A comparative study of Arabic dialects*. New York: Oxford University Press.

Brustad, K. (2000). *The syntax of spoken Arabic*. Washington, DC: Georgetown University Press.

Ferguson, C. (1997). 'God-wishes in Syrian Arabic'. In R. K. Belnap, & N. Haeri (Eds.), *Structuralist studies in Arabic Linguistics: Charles Ferguson papers, 1954-1994* (pp. 212–228). La Haye/New York/Cologne: Brill.

Holes, C. (2004). *Modern Arabic structures, functions, and varieties. Washington D.C.*: Georgetown University Press.

Mohammad, M. (2000). *Word order, agreement, and pronominalization in standard and Palestinian Arabic*. Amsterdam: John Benjamins. DOI: 10.1075/cilt.181

Language acquisition

On the L1 development of final consonant clusters in Cairene Arabic*

Marwa Ragheb and Stuart Davis
Indiana University

This study presents data on the first language development of final consonant cluster acquisition in Cairene Arabic. We compare the production of final consonant clusters of two siblings (an older brother and a younger sister) acquiring Cairene Arabic in a monolingual setting when both were 2 years, 8 months (2;8). Since one child had more target-appropriate clusters than the other at that age, we get a glimpse of the developmental path of final consonant cluster acquisition in Cairene Arabic. Our findings include that pharyngeal-initial final clusters are acquired early and that gemination is the common "repair" strategy for clusters not yet acquired. We conclude by relating our findings to theories regarding the nature of first language phonological acquisition.

Keywords: Phonology, child language acquisition, consonant clusters, Cairene Arabic, geminates.

1. Introduction

While there have been an increasing number of studies on Arabic phonological acquisition, certain areas of such research have not been addressed. Many of the

* This research has been partially supported by an Indiana University Grant-in-Aid for Research and was done under Indiana University IRB study number 08-13285. We would like to thank MG, RG, and their family for their time and cooperation. Various aspects of this paper have been discussed with Jon Anderson, Jessica Barlow, Outi Bat-El, Kenneth de Jong, Daniel Dinnsen, Heather Goad, Reem Khamis-Dakwar, Ghada Khattab, Michele Morrisette, Traci Nagle, Mitsuhiko Ota, Yvan Rose, Natsuko Tsujimura, Marilyn Vihman, and Islam Youssef. Preliminary versions of this paper have been presented at the 24th (2010) and 26th (2012) meetings of the Arabic Linguistics Society, the Georgetown University Roundtable in March 2010, the Old World Conference on Phonology in Marrakech Morocco in January 2011 and the International Child Phonology Conference in June 2011 and 2013. We thank the participants at these meetings for their feedback. We also thank the anonymous reviewers and the editors for their comments.

available studies focus on the normative ages of acquisition of consonants (e.g., Amayreh, 1994, 2003; Amayreh and Dyson, 1998; Omar, 1973; Saleh, Shoeib, Hegazi and Ali, 2007). Few studies, though, have examined the acquisition of consonant clusters. The only documented work on the acquisition of final consonant clusters in any Arabic dialect that we are aware of is our own preliminary (unpublished) work on one child acquiring Cairene Arabic (Ragheb, 2010; Ragheb and Davis, 2010). The aim of the current paper is to present and describe data collected from two children who are siblings (an older brother and a younger sister) acquiring Cairene Arabic (CA) in a monolingual setting when both were 2 years, 8 months (2;8) with a focus on their final consonant cluster acquisition. Because these two children were at somewhat different stages of cluster development, the comparison of the two children allows us to begin to understand the developmental path of cluster acquisition in Arabic. An additional aim of our paper is to relate our findings to the larger debate in acquisitional phonology as to the very nature of the acquisition process: One view holds that phonological acquisition is essentially reducible to articulatory development reflecting performance factors and not the phonological grammar (Hale and Reiss, 2008; Blevins, 2009), while a competing view maintains that phonological development reflects language competence that entails knowledge of the phonological structure of the language being acquired (e.g., Fikkert, 1994; Demuth, 1996; see also Rose and Inkelas, 2011 for a recent overview).

This paper proceeds as follows: in Section 2, we present background briefly mentioning some previous work on Arabic phonological acquisition and relevant information on Cairene Arabic phonology. In Section 3, we summarize and discuss our preliminary findings of the production of final clusters in CA of one child at the age of 2;8. In Section 4, we present data on final cluster acquisition from a second child, the younger sibling of the first, acquiring CA when she was also at the age of 2;8. In Section 5, we outline a predicted trajectory or developmental path for consonant cluster acquisition in CA, based on the described and analyzed production of the two children. We are able to get a glimpse at the trajectory since one child had more target-appropriate clusters than the other, even though both were at the same age when data were collected. In Section 6, we discuss the major findings of our work as it relates to the ongoing debate on the nature of L1 phonological acquisition.

2. Background

In this Section, we briefly discuss previous work on Arabic acquisitional phonology and then present some aspects of Cairene Arabic phonology that are relevant for our current study.

2.1 Arabic acquisitional phonology

As both Dyson and Amayreh (2007) and Khattab (2007) specifically note, there is a lack of studies on the acquisition of consonant clusters in Arabic. Nonetheless, there have been observations such as that of Dyson and Amayreh (2000) that there are low percentages of coda deletion or consonant cluster simplification in normally developing 2–4 year olds (based on Jordanian Arabic). These observations are consistent with a finding of Khattab and Al-Tamimi (2011) who note that the Lebanese Arabic children of their study did not go through a CV stage or even a CVC stage. Their first content words were already minimally bimoraic (e.g., CVCC or CVVC). What emerges from these reports on Jordanian and Lebanese phonological acquisition is the absence of the simple deletion of a consonant to avoid complex syllables. By simple deletion, we mean deletion that has no noticeable compensatory effects such as gemination or vowel lengthening. This differs from English phonological acquisition where the simple deletion of a consonant to avoid complexity in syllable structure is commonly observed (McLeod, van Doorn and Reed, 2002).

Most previous work on the acquisitional phonology of Arabic dialects has largely focused on examining the age of the mastery of production for each consonant, distinguishing between consonants that are acquired early from those that are acquired late (cf. for Jordanian Arabic: Amayreh, 1994, 2003; Amayreh and Dyson, 1998; for Egyptian Arabic: Omar, 1973 and Saleh et al., 2007). For example, with respect to Jordanian Arabic, Amayreh (1994, 2003) and Amayreh and Dyson (1998) studied the ages of acquisition of consonants concentrating on which individual consonants were acquired early and which were acquired later. In a similar study of Egyptian Arabic by Saleh et al. (2007), 30 Cairene-speaking children ranging in age between 12 to 30 months were examined. While Saleh et al. did not provide a detailed description on consonant cluster development, they report the occurrence of glottal stop replacement (that is, the use of a glottal stop as replacement for other consonants), which has been similarly observed in our study, too.

A final report that we will mention, by Ayyad and Bernhardt (2009), presents data from a normally developing (bilingual) child aged two years, four months (2;4) who was acquiring Kuwaiti Arabic in an English-speaking environment (North America). Based on the production of 38 words, the child had already acquired labials, dorsals, the uvular and pharyngeal fricatives and the emphatic /$t^ʕ$/ with 100% accuracy. The child had not yet acquired the other emphatics, the interdentals, /ʃ/, and /r/. Moreover, the data analysis shows that the child's productions included medial geminates and some consonant clusters in all positions (initially, medially, and finally). While this study is in some ways relevant for our own work, since we also observe the late acquisition of /r/ and the early acquisition of geminates and pharyngeal fricatives, the data did not report on certain structures such as words with final geminates. Also, Kuwaiti Arabic does not have the range of final consonant cluster types that are witnessed in Cairene Arabic.

2.2 Consonant clusters in Cairene Arabic

One of the main characteristics of Cairene Arabic in comparison to other Arabic dialects is that it allows for final clusters consisting of any two consonant phonemes. While a word-final syllable in CA can end in (maximally) two consonants, a non-final syllable can end in at most one consonant. On the other hand, CA lacks word-initial and syllable-initial consonant clusters generally, although [kw] occurs marginally in words like [kwajjis] 'good'. Examples of final clusters are given in (1). Cairene Arabic allows word-final clusters of any sonority profile (falling, level, rising). In addition, words with final geminates are common, such as [nusˤsˤ] 'half' and [sitt] 'woman'. (Note, transcriptions are essentially phonemic: low-level allophonic changes are not indicated.)

(1) Final consonant clusters in Cairene Arabic

Cairene	**Gloss**
a. bint	daughter
b. ʃams	sun
c. ʔism	name
d. ʔult	I said
e. ʔatl	killing
f. ʕaks	opposite
g. misk	musk
h. ʕabd	slave
i. kidb	lies
j. masˤr	Egypt
k. ʔursˤ	tablet, pill

The data in (1) a., b., d., and k. exemplify words where the final consonant cluster has falling sonority going from a sonorant consonant into an obstruent. Such final clusters are fairly common in Arabic dialects. The final clusters in (1) c., e., and j. show rising sonority with an obstruent followed by a sonorant. In many dialects, an epenthetic vowel would occur at least optionally to break up such clusters so that these words would be pronounced as two syllables. In CA, the words in (1) c., e., and j. are pronounced as monosyllables with some degree of phonetic devoicing of the word-final sonorant. They cannot be pronounced with an epenthetic vowel, not even optionally. The words in (1) f.–i. end in two obstruents and can be considered as displaying level sonority of the final cluster. Again, such words are pronounced as single syllables.

Our study focuses on consonant cluster development in children acquiring CA. Given the range of final consonant clusters in CA as exemplified in (1) along with the difficulty that children acquiring English have in mastering final clusters (McLeod et al., 2002), it could be hypothesized that final clusters would be difficult for children acquiring CA to master. Moreover, if they are difficult to master, a sonority effect can be assumed in which clusters of falling sonority should be more easily mastered (i.e., occur earlier in acquisition) than clusters of rising sonority, given that they are much more common across languages. Furthermore, it could be hypothesized that typically developing CA-speaking children might delete one consonant of the final cluster or insert a vowel into the cluster, similar to typically developing English-speaking children. Before turning to our acquisition data that bear on these matters, we briefly discuss the prosodic nature of CA word-final syllables since this will be of importance in understanding the acquisition data.

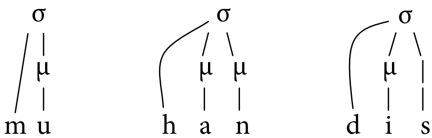
The prosodic nature of word-final syllables in CA is reflected by the stress pattern of the language. Any word in which the final syllable ends in two consonants will have stress on the final syllable as exemplified in (2) a. Similarly, any word that ends in a geminate consonant will have final stress as indicated in (2) b. On the other hand, a word that ends in a CVC syllable with a single word-final consonant does not have stress on the final syllable. It will normally have stress on the penultimate syllable (2) c. or the antepenultimate syllable in longer words (2) d. (specifically when the antepenultimate syllable is CV or light), unless the penultimate syllable is heavy (e.g., CVC), in which case the penultimate syllable attracts stress as in (2) e.

(2) Representative stress patterns of Cairene Arabic (period indicates syllable boundary, the stress syllable is in bold)

a.	ka.**tabt**	'I wrote'
b.	ʔa.**xaff**	'lightest'
c.	**ka**.tab	'he wrote'
d.	**ku**.tu.bak	'your (masc.) books'
e.	mu.**han**.dis	'engineer'

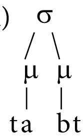
The CA stress pattern, especially as it relates to final syllables, can be understood through the notions of moraic weight and final consonant extraprosodicity (see Hayes, 1995 and Watson, 2002 concerning CA). Essentially, a coda consonant (but not a word-final consonant) adds weight to the syllable. That is, a coda consonant is moraic except if it is at the end of the word. Given that short vowels also add a mora to the syllable, we see that in comparing (2) a. with (2) c., a final syllable receives stress if it is bimoraic. If the final syllable is not bimoraic, then the penultimate syllable receives stress if it is bimoraic, as illustrated by a comparison between (2) e. and d. Words ending in a geminate always receive stress on the final syllable ((2) b.) thus suggesting that a geminate consonant always adds a mora to the syllable (see Watson, 2002 on this point). In (3), the mora structure of the word [muhandis] 'engineer' is illustrated, and in (4) and (5) we show the mora structure of the final syllable of [katabt] 'I wrote' and [ʔa.xaff] 'lightest', respectively.

(3) Moraic representation of [mu.**han**.dis] 'engineer' (σ indicates syllable; μ indicates mora)



(4) Moraic representation[1] of the final syllable of [ka.**tabt**] 'I wrote'

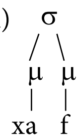
(Final)



(bimoraic, receives stress)

(5) Moraic structure of the final syllable of [ʔa.**xaff**] 'lightest'

(Final)



(bimoraic, receives stress)

The examples in (2) and the illustrations in (3)–(5) show that a final syllable receives stress if it is bimoraic. What will be important for our study is the parallel prosodic structure shown in (4) and (5) between words ending in two consonants and a final geminate. Both types of words end in a bimoraic syllable that attracts stress. We now turn to our acquisition data.

3. Child 1: Word list MG

In this section we report on the data and observations from our previous (unpublished) study (Ragheb, 2010; Ragheb and Davis, 2010). For that study, the first author elicited data in July 2008 from one male child, MG, aged two years, 8 months (2;8), who was typically developing and was acquiring Cairene Arabic in a monolingual environment in Cairo. Data were gathered using pictures that elicited word responses from a pre-designed word-list (focused on clusters) and spontaneous speech, which was recorded and later transcribed. The pictures used to gather

1. For (4) and (5), only the moraic elements are shown.

data were specifically chosen to elicit target words ending in consonant clusters.[2] The first author used a book that had different pictures or scenes in it and which was generally aimed at increasing a child's vocabulary. The researcher would ask MG to identify or search for certain objects, animals, and actions in the book in order to elicit the target words. Another task that relied more on spontaneous speech involved the researcher and MG engaging in telling stories or recounting certain events. Data were collected at multiple times over a one-month period, resulting in 10 sessions of about one hour in length. In (6), we see a representative sample of MG's production of word-final consonant clusters. The first column in (6) shows the target adult pronunciation of the CA word, and the second column demonstrates MG's pronunciation of the target CA word.

(6) A representative sample of MG's pronunciation of CA target words with final consonant clusters

Target pronunciation	**MG's pronunciation**	**Gloss**
a. nusˤsˤ	ʔusˤsˤ	half
b. ward	wadd	flowers
c. bint	bitt	girl
d. kalb	kabb	dog
e. ʃiribt	ʔitt	I drank
f. miʃtˤ	ʔitt	comb
g. naml	ʔall	ants
h. ħabl	ʔall	rope
i. ʔism/ʔismu/ʔismi	ʔimm/ʔimmu/ʔimmi	name/his name/ my name
j. malħ	ʔall	salt
k. ʔamħ	ʔamm	wheat
l. taħt	taħt	under
m. baħr	baħl	sea
n. ʃaʕr/ʃaʕru	saʕl/saʕlu	hair/his hair

2. In the process of elicitation, multiple tokens of each target words were collected over the different recording sessions. The spontaneous speech recorded also included multiple tokens of different target words over different points in time. The researchers only included for analysis those tokens that the children produced without any help from adults. Any tokens resulting from repetition after an adult modeling were disregarded, as imitation is not a reliable method of getting at a child's phonological knowledge. The data presented here for both children constitute a representative sample of the gathered data. The initial probe (word list) both changed and expanded according to the individual knowledge of the child. For example, the word [ʔatˤr], 'train' was initially on the list, but one of the children called it by another name, thus resulting in replacing this word by another target word with a final CC cluster that the child knew.

The present paper is focused on the production of final clusters, so we will not discuss MG's use of onset glottal replacement, especially common in words where the final clusters are more marked. (6) a. shows that MG had target-appropriate final geminates. These seem to be acquired very early in the acquisition process. (6) b.–d. shows that in a target word where the final cluster has falling sonority, MG deleted the first consonant of the cluster and geminated the second (i.e., the word-final consonant geminates). Similarly, in words where the target final cluster has level sonority (as in the obstruent clusters of (6) e.–f.) or rising sonority (as in (6) g.–i.), MG deleted the first consonant of the cluster and geminated the second or final consonant. Thus, (6) b.–i. shows no effect of the level of sonority on target cluster production. However, in producing pharyngeal final target clusters, a different pattern emerged, as illustrated in (6) j.–k. Here, MG deleted the final pharyngeal consonant and geminated the first consonant of the cluster. Importantly, from our more comprehensive set of data on MG at 2 years 8 months, we know that he did not have geminate pharyngeals in his system (though he clearly had pharyngeal consonants as singletons). Consequently, final gemination for the target words in (6) j.–k. was not a possible outcome. He nonetheless still geminated, but it was the initial consonant of the cluster that geminated rather than the final. Thus, the data in (6) b.–k. show a pattern of gemination for final target clusters. It is interesting to note that such a gemination pattern has not been reported as a manifestation of target final clusters in English L1 acquisition.

We now consider the data in (6) l.–n. These CA words contain final consonant clusters where the first consonant of the cluster is a pharyngeal fricative and the second is either an obstruent or a sonorant. Essentially, MG produced these clusters as target-appropriate.[3] This is most clearly seen with the data item in (6) l., /taħt/, which MG pronounced correctly. (6) m.–n. are almost pronounced target appropriately, with the word-final /r/ being pronounced as [l]. However, a more complete examination of MG's data shows that he had not acquired /r/ at this stage and normally substituted [l] for target-appropriate /r/. Consequently, we conclude that MG acquired a final cluster in (6) m.–n. where the first consonant is a pharyngeal and the second is a liquid.

MG's final consonant cluster data in (6) raises two issues. First, assuming that MG's language acquisition was normal, why do the first types of final clusters acquired include a pharyngeal-initial cluster? Is this related to sonority? Second, why is final gemination the common "repair" strategy for clusters that are not yet

3. It is crucial to point out that, with both children, the consonant clusters produced target-appropriately were done so every time they were elicited or when they occurred in spontaneous speech. Thus, even with the most stringent of measures of L1 acquisition, they are considered to have been acquired.

acquired appropriately, especially in light of the fact that this strategy has not been noted generally in the literature on L1 final cluster acquisition? In considering the first issue, it is important to mention that phonetic work on Egyptian Arabic by Elgendy (2001) has shown that pharyngeal fricatives have phonetic characteristics of glides. This characterization of pharyngeals has also been espoused by McCarthy (1994) and Halle (1995). Given this, we maintain that the final clusters acquired earliest by MG are those in which the first consonant of the cluster has the highest sonority, namely a glide. It should be noted that while CA has both the palatal glide /j/ and the labiovelar glide /w/, they do not appear in consonant clusters for independent reasons: CA underlying sequences with final glide clusters such as /bajt/ 'house' and /lawn/ 'color' surface as [beet] and [loon], respectively, because of the independent process of monophthongization. (See Youssef, 2010, for specific argumentation justifying that [ee] and [oo] derive from /aj/ and /aw/, respectively, in a synchronic analysis of the phonology of CA). Thus, we maintain that MG's final cluster acquisition is constrained by sonority preferences in that the first member of a final cluster should be of highest sonority so that there will be a sonority fall in the first clusters acquired. In other words, while CA final clusters often violate preferred sonority sequencing (i.e., clusters can be of rising sonority), the L1 acquisition of such clusters may reflect markedness considerations whereby preferred final clusters with falling sonority are acquired earlier.

With respect to the second issue as to why MG always manifested final gemination for the clusters that were not yet acquired target appropriately as seen by the data shown in (6) b.–k., we have argued in our previous work (Ragheb, 2010; Ragheb and Davis, 2010) that gemination reflected MG's knowledge of the prosodic structure of Cairene Arabic, since by gemination MG was able to preserve the prosodic structure (i.e., mora structure) of the target word without having to pronounce two adjacent consonants that have two different articulations. Recall from the discussion in Section 2 that, as seen in (4), a final syllable that ends in a consonant cluster (or a monosyllabic word of the shape CVCC) is bimoraic. Further, as seen in (5), a final syllable that ends in a geminate is also bimoraic. This means, as we show in (7) below, that both the target pronunciation of words like [bint] in (6) c. and MG's pronunciation of it as [bitt] have the same prosodic structure.

(7) Moraic structure of target syllable [bint] and MG's pronunciation of it as [bitt]

```
Target [bint] 'girl':  a. Syllable          b. MG:  Syllable
                            /  \                     /  \
                            μ   μ                    μ   μ
                            |   |                    |   |
                            bi  nt (bimoraic)        b i   t  [bitt] (bimoraic)
```

Note that simply deleting one of the final consonants without geminating is problematic because the final syllable would be prosodically different from the target as shown in (8).

(8) Mora structure of target [bint] being pronounced as [bit]
[bint] 'girl' (Final)Syllable

```
        (Final)Syllable
              |
              μ
              |
            b i t
```

(monomoraic, no preservation of prosodic structure)

Moreover, it should be noted that CA lacks content words that are CVC, so the potential pronunciation of [bit] for target [bint] 'girl' would be in violation of the phonotactics of the language. Consequently, MG's gemination reflected his tacit knowledge of the prosodic structure of CA. By gemination he preserved the prosodic structure of the target word. MG's tendency to geminate the word-final consonant (as opposed to the first consonant of the final cluster) may just reflect the saliency of the right edge of the word. We now consider a second child whose cluster data provide more insight into the developmental path of cluster acquisition in CA.

4. Child 2: RG

In this Section, we report on the data and observations of a second child, a female, RG, also 2 years 8 months of age (2;8) at the time of data collection, who is acquiring Cairene Arabic as her L1 in a monolingual setting. She is the sister of MG, three years younger. In the course of 12 sessions during May 2011, the first author elicited target data through several picture-naming tasks, as well as through spontaneous speech. The data collection procedure and instruments were similar to those in MG's study, except for allocation of more time to spontaneous speech production. Thus, the same book was used, and RG was asked to perform the same tasks (e.g., search for, or identify, an object, animal, etc.). Data were recorded and then broadly (i.e., phonemically) transcribed.

In (9), we present a representative sample of RG's production of word-final consonant clusters. The first column shows the target adult pronunciation of the CA word and the second column indicates RG's pronunciation of the target CA word. We can initially observe from RG's data in (9) that she has produced more types of target-appropriate final clusters in comparison to the final clusters produced by MG at the same chronological age as was seen in (6).

(9) A representative sample of RG's pronunciation of CA target words with final consonant clusters

Target Pronunciation	RG's pronunciation	Gloss
a. taħt	taħt	under
b. baħr	baħd	sea
c. ʃibiʕt	ʃibaʕt	I'm full
d. ʃaʕr	ʃaʕd	hair
e. ʕuʔd	ʕuʔd	necklace
f. issubħ	issuħħ	(in) the morning
g. malħ	laħħ	salt
h. jaʕʕ	jaʕʕ	yuck!
i. ʃams	sans	sun
j. kinz	tinz	treasure
k. bint	mint	girl
l. miʃtˤ	miʃtˤ	comb
m. masˤr	masˤd	Egypt
n. ʔirʃ	ʔiʃʃ	shark
o. ʔird	ʔidd	monkey
p. durg	dudd	drawers
q. furn	funn	oven
r. maltˤ	matˤtˤ	naked/scantily dressed
s. kalb	tabb	a dog
t. libs	liss (i) beet, liss	clothes, indoor clothes
u. ʔakl	ʔatt	food
v. naml	namm	ants
w. ħabl	ħabb	rope
x. ʔismu (ʔism)	ʔissu (ʔiss)	his name (name)

Examination of the data in (9) reveals the following observations on RG's final consonant cluster development. First, as a general observation, RG had target appropriate word-final geminates. Unlike MG, this also included geminate pharyngeals, as indicated by the target appropriate form in (9) h. Second, like her brother at this age, RG had target-appropriate pharyngeal-initial consonant clusters in word-final position, as seen in (9) a.–e. This is most clearly seen in data items (9) a., c., and e., where the target cluster is a pharyngeal followed by an obstruent. The target pharyngeal-rhotic final clusters in (9) b. and d. were realized with the rhotic consonant as [d]. It should be noted that at this stage, RG did not have /r/ in any position. While MG frequently substituted [l] for target /r/, we observed that RG did not have [l] in coda position; thus she substituted [d] for target /r/ in (9) b. and (9) d. We consider her pronunciation of the clusters in (9) b. and d. as target-appropriate in the sense that she pronounces two different consonants in a pharyngeal-initial final cluster, with the first consonant being accurately produced as a pharyngeal.

In addition to the pharyngeal-initial clusters, RG had more target-appropriate final clusters than MG at this age. This includes Nasal + Obstruent final clusters as in (9) j.–k., but note instances of non-target-appropriate place assimilation as in (9) i., where the bilabial nasal of /ʃams/ 'sun' was assimilated in RG's pronunciation to the coronality of the word-final /s/, resulting in [sans]. It should be noted that Nasal + Sonorant clusters were not yet acquired target-appropriately by RG at this stage, as demonstrated by her pronunciation of the target form [naml] 'ants' in (9) v. as [namm] with gemination. Also, as indicated in (9) l.–m., Sibilant + Obstruent final clusters were also target-appropriate. This is clearly seen in (9) l., but can also be observed in 9 m., where the word-final rhotic was treated as the obstruent [d], just as in (9) b. and d. This should be compared with (9) x., where the rising sonority Sibilant + Nasal cluster shows gemination in RG's pronunciation rather than the target appropriate sequence of consonants. To summarize, at 2 years 8 months, RG had acquired three types of final clusters target appropriately: pharyngeal-initial final clusters, Nasal + Obstruent clusters, and Sibilant + Obstruent clusters.

With respect to the other clusters shown in (9) that RG has not yet acquired target appropriately, RG displayed a gemination strategy similar to her brother for the target final clusters, though her specific pattern of gemination was somewhat more complicated. In falling and same sonority clusters as in (9) n.–t., which had not yet been acquired target appropriately, RG deleted the first consonant of the final cluster, geminating the second. This should be contrasted with rising sonority final clusters as exemplified in (9) u.–x., where RG deleted the final consonant of the cluster, geminating the first. The specific example of target [ʔakl] 'food' in (9) u., which RG realized as [ʔatt], reflected her independent manifestation of velar stop fronting where the velar stops /k/ and /g/ were realized as coronal stops. This can be seen by her pronunciation of target /kinz/ 'treasure' in (9) j. as [tinz]. The important observation is that, like MG, RG showed gemination as the "repair" strategy for final clusters that had not yet been acquired target-appropriately. We maintain that this reflected her tacit knowledge of the prosodic (i.e., moraic) structure of Cairene Arabic, as was shown in (7) with respect to MG's pattern of gemination. Moreover, RG's choice of which consonant to geminate for the most part reflected the sonority of the consonant, namely, that the consonant with lower sonority in the cluster tends to geminate. This is most clearly seen in (9) n.–s., where gemination is applied to the lower sonority final consonant in the cluster, and in (9) v.–x., where the initial consonants of the clusters have lower sonority and are geminated. The data item (9) t. seems to suggest that in final clusters consisting of two obstruents, it is the second one that geminates. The only exceptions to this pattern of gemination are the clusters in (9) f. and g. that end in pharyngeal consonants. Since these clusters have rising sonority, they were not yet acquired target-appropriately by RG. Yet she geminated the more sonorant pharyngeal consonant, rather than the preceding consonant. This perhaps has to do with the saliency of pharyngeals in CA (or, restated from an optimality-theoretic perspective, the constraint requiring faithfulness to pharyngeal consonants was highly-ranked in her system).

Having presented the final consonant cluster systems of both RG and MG, we now turn to a discussion of a predicted trajectory or developmental path for consonant cluster acquisition in CA.

5. Developmental path

In the previous sections, we have described the pattern of word-final consonant clusters of two siblings at the same age (2;8) acquiring Cairene Arabic in a monolingual setting. While the children were at the same age when the data were collected, RG seemed to be further ahead in her cluster acquisition than MG. By comparing the productions of the two children, we can hypothesize a developmental path for CA cluster acquisition. Recall from Section 2 that CA allows for words to end in any two consonant phonemes regardless of their sonority relation. This is different from languages like English as well as other dialects of Arabic such as Lebanese (Haddad, 1984), which, while allowing for word-final clusters, do not normally permit such clusters with rising sonority. Given the general rarity of rising sonority final clusters in the world's languages and the more frequent occurrence of falling sonority clusters, one might hypothesize that a developmental path for final cluster acquisition in CA would entail that falling sonority clusters (and level sonority clusters) are acquired before rising sonority clusters. Although this hypothesis is generally true for both children in this study, since neither child had accurately acquired rising sonority clusters at 2 years 8 months, the comparison of the productions of these two children suggests that the developmental path of cluster acquisition is more nuanced in that certain falling sonority clusters are acquired before others. More specifically, it is of note that both children had acquired pharyngeal-initial final clusters. Since these are the only clusters that MG produced target-appropriately, it may be that these are the earliest clusters acquired. If we assume that the pharyngeals of Cairene Arabic have the phonetic properties of glides, as has been argued for by Elgendy (2001), then we would maintain that the cluster type that is acquired first is the one in which the first consonant of the cluster is of the highest sonority among consonants.

In addition to the pharyngeal-initial final clusters, RG had acquired a second type of falling sonority cluster by the age of 2 years 8 months: Nasal + Obstruent clusters. Compare for example, RG's pronunciation of /bint/ 'girl' in (9) k., where the cluster was pronounced target-appropriately, with MG's pronunciation of the same word shown in (6) c., in which there is gemination ([bitt]). The comparison of the two children suggests that RG was further along in her acquisition of final clusters and that Nasal + Obstruent clusters are acquired relatively early in the final cluster developmental path. The reason for this should be clear in that as seen in a word like [bint], there is only one place of articulation (coronal) in the cluster; thus, in an articulatory sense, these clusters are "easier" than clusters where the two consonants are not homorganic. That this is at issue can be seen in RG's pronunciation of target [ʃams] 'sun' in (9) i. as [sans], where she showed exceptional assimilation of the bilabial nasal to the coronal fricative so that the final cluster surfaced as homorganic. Importantly, sonority fall is still a factor since a rising sonority target cluster with an initial nasal consonant as in [naml] 'ants' ((9) v.) was pronounced with gemination, [namm], and so is distinct from the falling sonority Nasal + Obstruent clusters which are acquired earlier.

A third type of final consonant clusters that RG had acquired by 2 years 8 months was the Sibilant + Obstruent variety. From a certain perspective, the early acquisition of this type of obstruent-obstruent cluster is somewhat striking since in many languages, including English, Sibilant + Obstruent clusters are special because they can occur as onset clusters (or coda clusters), even though other types of obstruent-obstruent clusters are disallowed or restricted. In English, for example, the clusters [sp], [st], and [sk] are the only obstruent-obstruent clusters permitted in complex onsets. With respect to codas, [sp] and [sk] are the only obstruent-obstruent codas that end in a non-coronal consonant. While the special nature of Sibilant + Obstruent clusters for languages like English is well-known (Goad, 2011), these clusters are typically not treated as special in Arabic. Thus, it is of note that RG treated them as a distinct type in her acquisition of final clusters. Goad (2011) points out that for languages like English and Dutch, the acquisition of s-clusters is often independent of the acquisition of other cluster types with respect to sonority. We suggest that for RG the early acquisition of Sibilant + Obstruent clusters was independent of the acquisition of other clusters based on sonority, and we leave it as just an observation that RG distinguished Sibilant + Obstruent clusters from other cluster types even though these do not seem to have unique properties in Arabic phonology.

Finally, with respect to the final consonant clusters that neither MG nor RG had acquired target appropriately, we can divide them into three types: other falling sonority clusters such as liquid-obstruent and liquid-nasal clusters; level sonority clusters; and rising sonority clusters. These clusters were all realized with gemination of one of the last two consonants for both children, though the gemination pattern applied by RG differed slightly from that applied by MG. While MG typically geminated the final consonant of these clusters as long as it was not a pharyngeal (seen in (6) d.–h.), RG geminated the final consonant in a falling sonority cluster ((9) n.–s.) and the initial consonant in a rising sonority cluster ((9) v.–x.). Given that RG was distinguishing between rising and falling sonority in these cluster types, we speculate that, in the further developmental path for final cluster acquisition, other falling sonority clusters would be acquired before rising sonority clusters (abstracting away from the difficulty that both children have with the phoneme /r/). Although we leave for future research a more detailed examination of longitudinal data, it appears that, in general, falling sonority final clusters are acquired before rising ones even though both cluster types are common in CA.

6. Major findings and conclusions

As far as we are aware, this detailed presentation of final consonant cluster acquisition data of two children (aged 2;8) who are acquiring Cairene Arabic in a monolingual setting, is the first study that has a specific focus on the acquisition of such clusters. As previously mentioned, examining clusters in CA is particularly interesting because of the full range of consonant clusters allowed in word-final position in CA. Thus, we find it significant that cluster acquisition seems to be sensitive to sonority considerations, and, for RG, to the special status of Sibilant + Obstruent clusters even though neither of these is apparent in the adult phonology (which allows for any cluster type). However, what we consider to be our most important finding is the evidence that gemination is the major "repair" strategy attested in L1 acquisition of final consonant clusters for Cairene Arabic for both children at this early stage of development. This strategy has not been documented in languages like English and Dutch, where the simple deletion of a consonant or even of the final cluster can occur (e.g., Fikkert, 1994). In fact, as seen in work like that of McLeod et al. (2002), gemination is not even considered as a possible strategy in the development of final consonant cluster acquisition. McLeod et al., for example, list processes like deletion, epenthesis, and metathesis in cluster development, but not gemination. However, we suspect that gemination is a common strategy for target final clusters in Arabic dialects in general, given that all Arabic dialects seem to have final geminates and that they seem to be learned very early in the acquisition process.

There are other findings in our study, not necessarily related to consonant cluster development, that are worth mentioning. First, both children had acquired pharyngeal consonants early, a finding that is consistent with Ayyad and Bernhardt's (2009) work on Kuwaiti Arabic, though not consistent with the work on Jordanian Arabic (Amayreh, 1994, 2003; Amayreh and Dyson, 1998). An important difference regarding pharyngeals between our two subjects is that MG lacked geminate pharyngeals, but RG had them. Second, /r/ was acquired late for both children. However, the replacement strategy of target /r/ was quite different for each of these two children. MG tended to replace /r/ with [l]. RG often treated target /r/ as [d], and sometimes /r/ underwent consonant harmony. Further, RG also seemed to only have /l/ in onset position; she did not have /l/ in coda position. MG did not demonstrate difficulty with coda /l/. Third, RG had across-the-board velar fronting for /k/ and /g/, as seen by the data items in (9) j. and p., in which these target sounds were realized as [t] and [d], respectively. RG also showed occasional instances of consonant harmony of initial onsets, as in (9) i. and k. MG, on the other hand, showed no signs of velar fronting or consonant harmony. Instead, MG had extensive glottal replacement in word-initial position, which RG did not have. We suspect that these phenomena may be common in Arabic developmental phonology. For example, Saleh et al. (2007) reported glottal replacement in children acquiring Egyptian Arabic, and there is some consonant harmony in the Kuwaiti data reported by Ayyad and Bernhardt (2009). Future research is needed for the investigation of these processes in Arabic developmental phonology.

Finally, we would like to relate our findings to the ongoing debate in the literature on acquisitional phonology as to the very nature of the acquisition process (Rose and Inkelas, 2011). In the relevant literature on acquisition, two contrasting viewpoints can be found. One is the view that phonological acquisition is just articulatory development (Hale & Reiss, 2008; Blevins, 2009). Blevins (2009, p. 328) maintains that, "A wealth of data illustrate that the majority of recurrent features of child phonology (e.g., CV syllable stage, cluster reduction stage, consonant harmony) are reflections of articulatory developmental stages, indicating developmental constraints on performance, not on language competence." This can be interpreted as implying that children across different languages should manifest very similar strategies in phonological acquisition, since all children have essentially the same articulatory apparatus. The other view is that the nature of developmental errors is dependent on the structure of the language being learned. That is, errors in development reveal the linguistic competence that the child has with respect to the language (Fikkert, 1994; Demuth, 1996). This implies that children learning different languages will manifest different strategies in acquisition, reflecting the structures of the languages being acquired. Our claim is that the Cairene Arabic L1 cluster acquisition data presented in this paper support the second position. The structure of the phonological grammar plays an important role in the nature of the child's performance. This is most clearly seen by the predominance of gemination for the target final cluster. As noted earlier, gemination as a strategy in L1 acquisition for the pronunciation of final clusters has not been witnessed in other languages such as English, where children often delete consonants or insert vowels in final clusters. The strategy of word-final consonant gemination seen with both MG and RG can be understood as a means of preserving the prosodic moraic structure of the bimoraic final syllable in words that end in two consonants without the need for making two distinct consonantal gestures. This is seen by the parallel moraic structure in (7) for target /bint/ 'girl' and MG's pronunciation [bitt] with a geminate. Moreover, our acquisition data is consistent with the observation of Khattab and Al-Tamimi (2011), that children acquiring Lebanese Arabic do not seem to go through a CV (or CVC) stage. Such word forms would be monomoraic in Arabic, and many Arabic dialects (including Cairene) require content words to be minimally bimoraic (e.g., CVCC). Neither MG nor RG seem to have gone through a CV stage nor do they really have a process of consonant cluster reduction. This is surprising, given a view like that espoused by Hale and Reiss (2008) and Blevins (2009) who contend that acquisition is largely reducible to articulatory development. We thus conclude that the Cairene Arabic acquisition patterns evident in data from RG and MG provide insights into their linguistic competence, showing tacit knowledge of the moraic structure of the language. We would contend that knowledge of the nature of the grammar, that is, linguistic competence, plays an important role in determining the specific manifestation of the performance. The nature of the child's performance seems to be controlled by the higher-level linguistic structure. It is not reducible to just articulatory development.[4] Finally, although our two case studies are not longitudinal, we hope that

---

4. A couple of reservations about our analysis have been raised by two anonymous reviewers that we address in this footnote. One reviewer has concerns that the two children, MG and RG, are siblings. The implication is that since the children are receiving similar input, it would not be surprising that they both have gemination. This, then, would make it harder to generalize our finding to a larger child Arabic population. To respond to this, we would like to make two points: First, as detailed in the second paragraph of Section 2, MG and RG had quite different phonologies. For example, RG had across-the-board velar fronting of target /k/ and /g/ and instances of consonant harmony, while MG did not show any velar fronting and has widespread glottal replacement of onset consonants. Further, while neither child had target appropriate /r/, MG consistently replaced /r/ with [l] while RG frequently replaced it with [d]. Given that these two children are quite different in their phonological development, we find it even more telling that they both used gemination for target consonant clusters. Second, CA baby talk words often display final geminates. By 'baby talk,' we mean the way that adults imitate the speech of young children. Such words include [kuxx] 'something bad', [mamm] 'food', and [daħħ] 'something good'. This implies that adults perceive that it is common for children to make final geminates.

these preliminary findings will encourage such studies with larger samples in a range of Arabic dialects.

References

Amayreh, M. (1994). *A normative study of the acquisition of consonant sounds in Arabic*. Unpublished Ph.D. dissertation, University of Florida, Gainesville.

Amayreh, M. (2003). Completion of the consonant inventory of Arabic. *Journal of Speech, Language, and Hearing Research*, 46, 517–529. DOI: 10.1044/1092-4388(2003/042)

Amayreh, M. M., & Dyson, A. T. (1998). The acquisition of Arabic consonants. *Journal of Speech, Language, and Hearing Research*, 41, 642–653.

Ayyad, H. & Bernhardt, B. M. (2009). Phonological development of Kuwaiti Arabic: Preliminary data. *Clinical Linguistics & Phonetics*, 23, 794–807. DOI: 10.3109/02699200903236493

Blevins, J. (2009). Phonetic explanations for recurrent sound patterns: Diachronic or synchronic? In E. Raimy, & C. Cairns (Eds.), *Contemporary views on architecture and representations in phonology* (pp. 325–336). Cambridge, MA: MIT Press.

Demuth, K. (1996). The prosodic structure of early words. In J. Morgan, and K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 171–184). Mahwah, NJ: Lawrence Earlbaum Associates.

Dyson, A. T., & Amayreh, M. (2000). Phonological errors and sound changes in Arabic-Speaking children. *Clinical Linguistics & Phonetics*, 14, 79–100. DOI: 10.1080/026992000298850

The same reviewer wonders if the children were hearing a lot of deleted consonants in the input final clusters given that it is well known from English that word-final /t/ and /d/ are often deleted (especially after another consonant) by speakers of all ages. Relatedly, the other reviewer wonders about the functional load of consonants. We are quite certain that children are not hearing deleted consonants as part of the lexical input. For example, it is unlikely that adults are pronouncing a word like /kalb/ 'dog' as [kab] or [kal] or even as [kabb] or [kall] when speaking to children. Keep in mind that final CVCC syllables always carry stress whereas final CVC syllables do not. Consequently, a consonant cannot just be deleted at the end of a word as in English, since it would play havoc with the stress pattern. Moreover, in monosyllabic CVCC words, a consonant cannot just be deleted since there are no real content words in CA that are CVC. Recall from Section 2 that Khattab and Al-Tamimi (2011) found that the Lebanese Arabic children of their study did not witness a CV or CVC stage. This is consistent with our findings and makes sense because CV and CVC content words in CA (and Lebanese Arabic) are rare or nonexistent. Further, given the root and pattern system of Arabic morphology, deleting a consonant would play havoc with the lexical meaning of a root, which is expressed by the consonants. In this sense, we suspect that Arabic consonants carry a high functional load. Consequently, our strong impression is that adult Arabic speakers do not delete a final consonant of a content word in a way that English speakers delete a final /t/ or /d/. As is well known, dialects like CA often use the strategy of epenthesis over a word boundary so that a word-final consonant would not be deleted as exemplified by the typical CA pronunciation of *bint kibiira* as [bin.tik.bii.ra] 'a big girl'. Compare this with English "grant competition" where the final /t/ is likely to delete and there is no regular process of epenthesis to prevent its deletion.

Dyson, A. T., & Amayreh, M. (2007). Jordanian Arabic speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 287–299). Thomson Delmar Learning.

Elgendy, A. (2001). *Aspects of pharyngeal coarticulation*. Netherlands: LOT.

Fikkert, P. (1994). *On the acquisition of prosodic structure*. The Hague: Holland Academic Graphics.

Goad, H. (2011). The representation of sC clusters. In M. van Oostendorp, C. Ewen, E. Hume, and K. Rice (Eds.), *The Blackwell companion to phonology* (Vol. 2, pp. 898–923). Oxford: Wiley Blackwell.

Haddad, G. (1984). Epenthesis and sonority in Lebanese Arabic. *Studies in the Linguistic Sciences*, 14, 57–88.

Hale, M., & Reiss, C. (2008). *The phonological enterprise*. Oxford: Oxford University Press.

Halle, M. (1995). Feature geometry and feature spreading. *Linguistic Inquiry*, 26, 1–46.

Hayes, B. (1995). *Metrical stress theory*. Chicago: University of Chicago Press.

Khattab, G. (2007). Lebanese Arabic speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 300–340). Thomson Delmar Learning.

Khattab, G., & Al-Tamimi, J. (2011). *The role of geminates in shaping early word patterns by Lebanese Arabic speaking children*. International Child Phonology Conference, University of York, York, UK, June 15–18.

McCarthy, J. (1994). The phonetics and phonology of Semitic pharyngeals. In P. Keating (Ed.), *Papers in laboratory phonology III: Phonological structure and phonetic form* (pp. 191–233). Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511659461.012

McLeod, S., van Doorn, J., & Reed, V. A. (2002). Typological description of the normal acquisition of consonant clusters. In F.Windsor, L. Kelly, & Hewlett (Eds.), *Themes in clinical phonetics and linguistics* (pp. 185–200). Hillsdale, NJ: Lawrence Erlbaum.

Omar, M. (1973). *The acquisition of Egyptian Arabic as a native language*. Cambridge: Cambridge University Press.

Ragheb, M. (2010). *The phonological acquisition of word-final consonant clusters in Cairene Arabic*. Unpublished manuscript, Indiana University, Bloomington.

Ragheb, M., & Davis, S. (2010). *The acquisition of word-final clusters in Cairene Arabic: An OT analysis*. 24th Arabic Linguistics Symposium. University of Texas at Austin, April 9–11.

Rose, Y. & Inkelas, S. (2011). The interpretation of phonological patterns in first language acquisition. In M. van Oostendorp, C. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (Vol. 4, pp. 2414–2438). Oxford: Wiley Blackwell.

Saleh, M., Shoeib, R., Hegazi, M., & Ali, P. (2007). Early phonological development in Arabic Egyptian children: 12–30 months. *Journal of Folia Phoniatrica et Logopaedica* 59, 234–240. DOI: 10.1159/000104461

Watson, J. (2002). *The phonology and morphology of Arabic*. Oxford: Oxford University Press.

Youssef, I. (2010). Against underlying mid vowels in Cairene Arabic. *Zeitschrift für Arabische Linguistik* 52, 5–38.

Neurolinguistics

Neurocognitive modeling of the two language varieties in Arabic Diglossia

Reem Khamis-Dakwar and Karen Froud

Diglossia is one distinctive feature of Arabic and other languages, such as Swiss-German (Ferguson 1959). Neurocognitive studies aiming to understand the neural mechanisms of diglossia in general, and Arabic diglossia in particular, are sparse. This paper provides a framework for discussing neurophysiological approaches to questions concerning the representation and processing of languages exhibiting diglossia in the brains of native speakers, as well as understanding the potential contribution of such approaches for applied linguistics and teaching Arabic as a first or second language.
The first section introduces electroencephalography (EEG), and some event-related potentials (ERPs) known to correlate with aspects of language processing. The second section reviews neurophysiological studies of language representation and processing in Arabic diglossia to date. Finally, we discuss some potential contributions of neurophysiological studies to the fields of Arabic linguistics, applied linguistics and teaching Arabic. The review points to separateness of the two lexicons in Arabic diglossia and highlights the need for further neurophysoplogiocal and behavioral studies to revisit the suggested model of mental representation in diglossia.

Keywords: Diglossia, Event related potentials, mental representation.

1. Introduction

Diglossia is a term used to describe a specific sociolinguistic situation in which a single speech community shares two language varieties, one used as the spoken in everyday communication, and the other, usually highly codified, used for formal communication such as literature, education or journalism (Ferguson, 1959). Diglossia, a distinctive feature of Arabic, also pertains to several other linguistic situations; Ferguson mentions demotic and katharevousa Greek, Creole and French in Haiti, and Swiss-German (Schwyzertütsch) and German in Switzerland. Despite the exponential increase in neurocognitive studies of language processing and bilingualism in the past decade, neurocognitive studies aiming to understand the neural mechanisms of diglossia in general, and Arabic diglossia in particular, remain sparse. This paper aims to provide a foundation for discussion and development of neurophysiological experimental approaches to questions concerning the mental representation of diglossic languages. We suggest that neurocognitive approaches are essential for enhancing our understanding of the nature of diglossia and that such understanding in turn provides a valuable underpinning for future theoretical and applied studies of Arabic diglossia.

The first section of this paper provides a short, lay introduction to one neurocognitive methodology, the Event-Related Potential (ERP) method. We will briefly describe the methods that are used to track the neurophysiological responses associated with some aspects of linguistic processing and will provide examples of studies that have utilized ERP methods to shed light on neural underpinnings of language. The second section reviews neurophysiological studies of language representation and processing in Arabic diglossia to date. Finally, we outline one way that the neurophysiological findings thus far could map onto a mental model of the cognitive inter-relationships between two language varieties and discuss the potential contributions of further neurophysiological studies to our knowledge about Arabic diglossia.

1.1 Derivation of event-related potentials

Cognitive neuroscience is the field that is concerned with the neural underpinnings of cognitive functions. In this domain, a primary aim is to "understand how cognitive functions, and their manifestations in behavior and subjective experience, arise from the activity of the brain" (Rugg, 1997: 1). In service of such investigations, various neuroimaging techniques are used to provide information about the location and timing of brain activity that can be experimentally associated with stimulus processing. Metabolic measures, such as functional Magnetic Resonance Imaging (fMRI), track what is referred to as BOLD signal – that is, a signal that is Blood Oxygenation Level Dependent. By establishing where oxygen is taken up in the brain under controlled conditions, it is possible to determine – with millimeter precision – the location of brain activity that is related to a specified experimental task. Such spatial resolution is valuable in answering questions about the specific network of regions that is involved in a task such as the co-relationships between Broca's and Wernicke's areas for language processing. However, due to physical constraints on the equipment and the refractory time of the measure itself, BOLD signal does not permit temporal resolution finer than a few seconds to minutes (Luck, 2005). In the case that we are investigating, processes known to unfold on the order of milliseconds, a more direct tracking of neuronal communication is called for. Electroencephalography (EEG) is a non-invasive means for measuring voltage fluctuations associated with the post-synaptic potentials that are generated as populations of neurons fire together. EEG is typically recorded on a millisecond scale, and the recent advent of high density EEG has also made it more possible to associate the measurement of task-related neural activity with specific brain sources. Therefore, EEG is considered among the best neuroimaging techniques for tracking sequences of rapid neural activations that are associated with complex mental operations, such as language processing (Friedrici, 2002).

Because EEG provides a continuous measure of voltage change across the scalp during the recording epoch, it is necessary to distinguish brain activity that is related to the event of interest from background brain activity. This is usually achieved through processes of data segmentation and averaging, in order to derive the Event-Related Potential (ERP). ERP refers to the average electrical activity elicited as a response to the experimental condition. It is obtained by segmenting EEG recordings into epochs, which are segments of time linked to the event of interest, and conducting several operations to exclude responses that are not related to the examined condition/event, such as eye movements, EKG (the

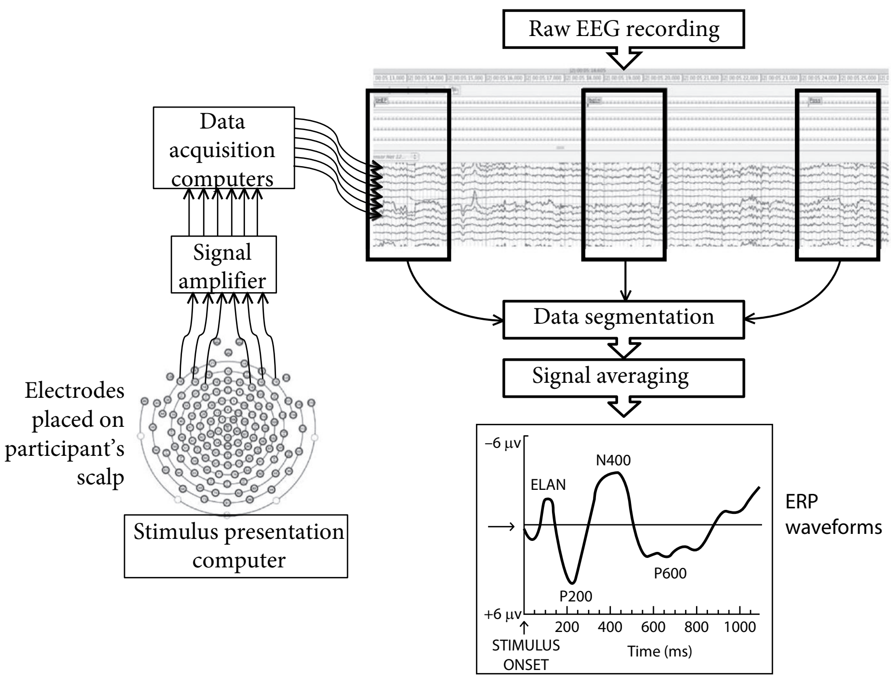


Figure 1. Steps in acquisition and derivation of Event Related Potentials in a cognitive experiment with EEG recording.

electrocardiogram), or skin potentials. The segments of data associated with specific events are then averaged together, which has the effect of "canceling out" information about brain activation that is not associated with experimental manipulations, ideally leaving behind only information about the brain responses associated with the examined task/stimuli – the ERP. Figure 1 above depicts the process of eliciting, acquiring and deriving ERPs during a cognitive experiment.

In contrast to the continuous EEG recording, ERPs index changes in electrophysiological brain responses that are associated with a particular stimulus, process or event – that is, ERP waveforms offer "a depiction of the changes in scalp recorded voltage over time that reflects the sensory, cognitive, affective, and motor processes elicited by a stimulus" (Kappenman & Luck, 2012: 4). In EEG/ERP studies of language, the main focus is typically on the brain's responses to a specific experimental event presentation, such as processing a word or the structure of a sentence. These event-related electrical potentials are referred to and analyzed in terms of their characteristics, especially timing, direction of voltage deflection (negative or positive), and/or scalp topographies (represented by electrode sites).

1.2 Key language-related ERP signatures

Certain ERP "signatures" have been shown, experimentally, to be associated with different aspects of language processing, and here we provide an introductory overview of key language-related ERPs: the Mismatch Negativity (MMN), the N400, and the P600.

The three ERPs that form the basis of this short tutorial segment were selected because they index different levels of language processing and representation: phonology, lexical semantics, and syntactic integration, respectively. They therefore provide a representative sample of the methods that have potential for application in the study of different levels of linguistic representation in Arabic diglossia. Thanks to ERP studies a greater understanding of cognitive processes supporting phonological, lexical semantic and syntactic operational levels of linguistic representation has been obtained, over many years of investigation. The reader should note that there are many more ERPs known to index different stages or processes that are crucially involved in language, and the underlying mechanisms associated with some of these remain poorly understood. (For a more detailed review of cognitive components, see Luck & Kappenman, 2012.)

For each ERP described here, we provide an example of a novel application or interpretation by way of illustration of the utility of such approaches for complex questions in language acquisition, representation, processing or pathology. Examples of applications to Arabic diglossia are discussed in more detail in Section 2 below.

1.2.1 *Mismatch negativity (MMN) ERP*

The Mismatch Negativity is an ERP component known to be associated with sensory-memory updating and detection of change or of rule violation. Generators of the MMN are known to play a central role in attention (Näätänen, Kujala, & Winkler, 2010), and MMN has been shown to be elicited in response to changes in sound duration, frequency, intensity, and timbre as well as language-specific speech-sound changes and grammatical violations (Näätänen & Kreegipuu, 2012). MMN is usually elicited using an *oddball paradigm* – that is, an experimental procedure in which auditory stimuli are presented in a continuous, repetitive stream that is interrupted by occasional (less than one in five) presentations of a deviant sound. The MMN is a negative voltage fluctuation that is measured using one or more electrodes positioned fronto-centrally, with a peak around 100–250 milliseconds post-presentation of a deviant auditory stimulus.

Experimental studies have revealed that the MMN response can be enhanced to deviant stimuli that constitute language-specific phonological contrasts (e.g. Cheour, Ceponiene, Lehtokoski, Luuk, Allik, Alho, & Näätänen, 1998; Dehaene-Lambertz, 1997; Dehaene-Lambertz, Dupoux, & Gout, 2000; Näätänen, Lehtokoski, Lennes, Cheor, Houtilainen, Iivonen, Vainio, Alku, Ilmoneiemi, Luuk, Allik, Sinkkonen, & Alho, 1997; Pulvermüller, Kujala, Shtyrov, Simola & Tiitinen, 2001; Sharma & Dorman, 2000). For example, in a group of native speakers of Finnish, Näätänen et al. (1997) found that a larger MMN response was elicited in response to a contrast between two Finnish vowels than for an Estonian vowel contrast in spite of the fact that the Estonian vowel contrast incorporated a larger acoustical difference. Such findings have suggested that MMN can be used to "probe permanent language-specific speech-sound memory traces" (Näätänen & Kreegipuu, 2012: 149).

MMN has been commonly used in language studies due to its ease of elicitation and the lack of need for conscious task participation (since MMN can be elicited while individuals are watching a silent movie or even sleeping), as well as its robustness as a correlate of language-specific phonological representations (Conboy, Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2008). MMN has been found to develop as an index of language-specific speech-sound memory traces within the first year of life (around the age of 6–12 months; Näätänen & Kreegipuu, 2012), so it is also helpful for experimental paradigms that aim to track the development of language-specific phonological representations in young children or in adult L2 learners. Furthermore, current studies reveal that MMN can be elicited by sensory or cognitive violations at different levels of central auditory processing, which means that MMN approaches have the potential to contribute to studies of abstract feature analysis, parallel and sequential representations, language learning, and therapy effects (inter alia) (Näätänen & Kreegipuu, 2012).

In one example of a novel application of the MMN methodology, Froud & Khamis-Dakwar (2012) used an oddball task to examine the phonological representations of speech sounds in monolingual English-speaking children with childhood apraxia of speech (CAS). MMN responses to native and non-native speech sound contrasts were compared between monolingual English-speaking children with and without CAS. The sounds used for the native contrasts were standard /pa/ and deviant /ba/ (this contrast is phonemically significant in English); for the non-native contrast, standard /pa/ was contrasted with deviant /pha/ (this contrast is allophonic in English). Children without CAS showed, as expected, clear MMN responses to the phonemic condition but not to the allophonic condition. However, children with CAS did not show an MMN in the phonemic condition, and showed some evidence of immature MMN responding to the allophonic contrasts. These findings are suggestive of phonological involvement in CAS, so that it cannot be viewed as a primary disorder of motor speech planning, contrary to the current prevailing view (e.g., ASHA 2007). By applying the standard MMN methodology, then, it was possible to determine new (and potentially clinically significant) information about speech-sound representations in children with disordered speech. Equally, such methods could be applied to shed light on whether speakers of Arabic distinguish between speech sounds within and between diglossic language varieties, when in development children are able to distinguish between phonological representations that are associated with different diglossic varieties, or when adult learners of Arabic are able to internalize speech sound representations in the standard and spoken phonological systems.

1.2.2 *N400 ERP*

The N400 is considered "the best-studied language-related ERP component" (Swaab, Ledoux, Camblin & Boudewyn, 2012: 398). Kutas and Hillyard (1980) first reported the N400 effect, showing that semantically inappropriate words in a sentential context (e.g., *He spread the warm bread with <u>socks</u>*) elicited a negative-going voltage fluctuation over centro-parietal electrodes that peaked around 400 milliseconds after onset of the unexpected word, relative to the ERPs elicited by semantically appropriate words (e.g., *It was his first day at <u>work</u>*). Since then, typical language N400 studies utilize presentations of congruent and incongruent conditions at different levels: the word level, sentence level, and during discourse (Swaab et al., 2012). N400 is accepted as an index of lexical and semantic processing; however, it is still debated whether the N400 is a general index of semantic retrieval of stored conceptual knowledge associated with the presented context or semantic integration into the proceeding context. In linguistic presentations, the N400 was found to be sensitive to meaning manipulations and to the presentation of an unexpected word at both the word and sentence levels, during visual or oral presentations, and with different kinds of anomalies, such as mismatches in phonology, faces, color patches, and pictures (for a review, see Kutas, Van Petten & Kluender, 2006; or Lau, Phillips & Poeppel, 2008).

In a study of Arabic diglossic code-switching (described in more detail below), Khamis-Dakwar & Froud (2007) presented oral presentations of congruent and incongruent sentences in spoken (Palestinian) and Modern Standard Arabic (MSA). Although this study targeted codeswitching rather than semantic representations, N400 components were observed in response to semantic anomalies in MSA that were processed auditorily by native speakers of Palestinian Arabic. These findings parallel those from other studies that have used N400 to index semantic processing in bilingualism, showing that semantic processing can be robust across multiple linguistic systems (e.g. Weber-Fox & Neville, 1996; Ojima, Nakata & Kakigi, 2005). Moreno and Kutas (2005) showed that changes in N400 latency in response to semantic incongruities in a first and second language are sensitive to multiple factors, including language dominance, vocabulary proficiency, and age of exposure. Hence, N400 can be used to identify the effects of various sociolinguistic factors on the relative robustness of semantic processing in different language varieties.

1.2.3 *P600 ERP*

The P600 is one of a complex of ERP components that reflect different stages in syntactic processing (Swaab et al 2012; Hahne & Friederici, 1999). The P600, also known as the Syntactic Positive Shift (SPS), is a slow, late, positive-going voltage shift observed over posterior electrode sites around 600 milliseconds after the onset of a sentence-structure violation (Hahne & Friederici, 1999). The P600 has been experimentally shown to correlate with late-controlled grammatical processing and can be elicited by syntactic anomalies (such as violations of phrase structure or agreement) in either meaningful or meaningless sentences. P600 is also elicited by processing of garden path sentences, in which two plausible structural configurations are available until a specific resolution point (Osterhout, Holcomb, & Swinney, 1994), suggesting that P600 is an ERP index of syntactic reanalysis and repair. Moreover, recent reports showed that P600/SPS can be elicited in response to lexical code-switching between L1 and L2 (Moreno et al., 2002; Jackson, Swainson, Cunnington, & Jackson 2001) or between two language varieties in diglossic situations (e.g. Khamis-Dakwar & Froud, 2007). This suggests that P600/SPS is associated with semantic-syntactic interface violations (Swaab et al., 2012).

1.2.4 *Mental chronometry of language*

In this section we have provided a brief overview of the methods involved in deriving event-related potentials from continuous recordings of electrophysiological activity generated by the brain and have provided some background information about the nature of specific ERPs known to be associated with aspects of language processing. Although partial, this introduction hints at a much broader and more fine-grained chronometry of language (Friederici, 2002) involving multiple cortical generators and complex interactions. Although each ERP has its own timing and topographical characteristics, it should be made clear that language-related ERPs do not necessarily proceed in a stepwise manner, any more than the linguistic processes that are thought to be associated with them. Rather, linguistic processes advance cyclically, with feedback and feedforward interactions, and the chain of ERP indices provides a reflection of the brain processes supporting this complexity.

In what follows, we provide a summary of the neurocognitive investigations of Arabic diglossia reported to date and suggest a possible theoretical framework that may be emerging from this small but growing literature. We propose directions for continuing the rigorous investigation of the mental representation of diglossia that can be catalyzed by the approaches implemented thus far.

2. Neurophysiological studies of language representation and processing in Arabic diglossia

Most of the ERP studies conducted in Arabic have addressed different questions than the relationship between the two language varieties, such as the assessment of the effects of orthographic transparency on event-related potentials (e.g. Simon, Bernard, Lalonda, & Rebaï, 2006) or the neural basis of morphological representation of the Arabic consonantal root (Boudelaa, Pulvermüller, Hauk, Shtyrov, & Marslen-Wilson, 2010). Moreover, these investigations studied Modern Standard Arabic (MSA) stimuli only and did not attempt to address the mental representation of the two language varieties in Arabic diglossia. This paper presents the few neurophysiological studies on the representation of the two language varieties in the brain of native Arabic speakers and discusses how innovative use of ERP methodologies can meaningfully expand our current and future research in this area.

The first study examining cognitive representation in Arabic diglossia, by Khamis-Dakwar & Froud (2007), focused on an examination of brain responses of native speakers of Arabic to lexical switching between the two language varieties. The two ERP components of interest in this study were the N400 and P600/SPS. Previous studies showed that codeswitching *between* languages elicits P600/SPS (Moreno, Federmeier & Kutas, 2002; Jackson, Swainson, Mullin, Cunnington, & Jackson, 2004), whereas incongruent and unexpected word presentations *within* a language were found to be correlated with an N400 response. Hence, Khamis-Dakwar & Froud hypothesized that, if the cognitive interrelationship between the two language varieties of Arabic is underlined by a separation at the lexical level, then switching between the two language varieties would elicit a P600/SPS effect, similar to brain responses of bilingual Spanish-English speakers to switches between the two languages. However, if the underlying neural mechanisms of the two varieties were unified, then an N400 response, typically observed between dialects or registers, would be elicited for codeswitching between the two language varieties.

Khamis-Dakwar & Froud (2007) designed a study in which diglossic codeswitching, language variety, and congruency were manipulated such that the last word of a spoken sentence was either congruent in a specific language variety, incongruent in a specific language variety, or congruent but codeswitched between language varieties. Examples are provided in (1)–(3) below:

Thirty-nine sentences in each condition were recorded by a native speaker of Palestinian Arabic (Northern dialect) and were presented auditorily to participants while continuous high density EEG was recorded. Participants were asked to press a button to indicate whether the last word of each sentence was presented in the same or a different language variety to the rest of the sentence.

(1) Congruent sentences in spoken and Standard Arabic

a. Palestinian Colloquial Arabic
ʔiza bitʃu:f ʔil-ʔahrama:t btinisweħ
if see-2.SG.MASC. the-pyramids astonished-2.SG.MASC
If you see the pyramids you will be astonished.

b. Modern Standard Arabic
ʔal-bu:ma taṣħa: bi-ððulma
the owl wake-2.SG.FEM at-dark
The owl wakes at dark.

(2) Incongruent sentences in spoken and Standard Arabic

a. Palestinian Colloquial Arabic
ʔiza bitʃu:f ʔil-ʔahrama:t btifʔar
if see-2.SG.MASC.the-pyramids get poor-2.SG.MASC
If you see the pyramids you will get poor.

b. Modern Standard Arabic
ʔalbu:ma taṣħa: bilʕiṭr
the owl wake-2.SG.FEM at-perfume
The owl wakes at perfume.

(3) Code-switched sentences
  a. Spoken → Standard
     we:n *na:m-at* *ʔil-: qiṭṭa?*
     where slept-2.SG.FEM the-cat-MSA
     Where did the cat sleep?
  b. Standard → Spoken
     la: *tukθər* *mina ʔal-ħaki*
     NEG. excess-2.SG.MASC. from the-talk-PCA
     Don't speak too much.

In the code-switched condition, semantically equivalent words with no phonemic correspondence (lexically different items) between the language varieties were used in order to minimize word form processing effects.

Five native speakers of Arabic with a Galilee dialect participated in this study: four males and one female with a mean age 30 years, 2 months (range: 25 years 10 months – 36 years 2 months). Participants were right-handed with normal hearing and normal or corrected-to-normal vision. All participants had been exposed to MSA in Arab schools in Israel starting in first grade.

The derived ERPs from the continuous EEG recordings obtained during this study revealed N400 responses to the semantic anomaly conditions but P600/SPS responses to the sentences that involved code-switching between the two language varieties. These findings suggest that native speakers of Palestinian Arabic process lexical switching between Standard and spoken Arabic in a manner that is similar to the processing of cross-linguistic code-switches in speakers of Spanish and English. They did not show evidence that switching into or out of Standard Arabic constituted a change in register or a switch between variants of a single linguistic system. Khamis-Dakwar & Froud interpret this as preliminary evidence that the two language varieties in Arabic require the representation of distinct lexical stores, one for each variety. This raised the question whether switching between varieties has neurophysiological consequences that are measurable at levels other than lexical.

This question was the focus in a second series of experiments that looked at the ERP responses to phonological code-switching between varieties of Arabic (Khamis-Dakwar, Froud, & Boudelaa, 2009). For this series, the ERP targeted was the MMN, understood as an index of language-specific phonological memory traces (Näätänen, 1999). Two different oddball paradigms were used. The first switched phonemes within a single language variety, changing only the semantics of the stimuli (standard: /ħaʔʔ/ ("right" in Levantine); deviant: /ħadd/ ("border" in Levantine)). The second kept the semantics of the stimuli the same, but switched between language varieties (standard: /ħaʔʔ/ ("right" in Levantine); deviant: /ħaqq/

("right" in MSA)). The contrast between brain responses to these two oddball paradigms, it was hypothesized, would permit identification of neurophysiological consequences over and above the phonemic category change when a switch across varieties is involved.

Study participants were 17 students and affiliates of American universities who were native speakers of Levantine Spoken Arabic. All shared a pronunciation of the classical Arabic *qaf* /q/ as glottal stop /ʔ/in their spoken dialect. There were 12 females and 5 males, 1 left handed and 16 right handed, 6 Egyptians, 3 Lebanese, and 8 Palestinians with a mean age of 31 years (range 25 years 7 months – 36 years 2 months). All participants reported having no history of neurological or psychological disorders, had normal hearing and corrected-to-normal vision, and had begun learning MSA at school in their homeland in first grade. They participated in a passive listening task, watching a silent movie of their choice while the syllables were presented to them through earphones and high-density EEG data were recorded.

Data analysis revealed MMN responses to the deviants in both conditions, but the MMN response was significantly enhanced when the phonemic change resulted in a switch between language varieties. Based on the hypothesis that MMN is an index of phonological memory traces, these results suggest that switching between varieties – even at the phonemic level – does have consequences at the neurophysiological level. These data strongly support the view that lexical representations are distinct between the standard and spoken language varieties.

Investigations of the syntactic systems at play in diglossia have not yet been undertaken, and no studies have yet been published that use the P600 as a means to evaluate the nature of the interactions between Standard and spoken varieties. However, in our lab there is pilot work under way that uses P600 as an index of morphosyntactic processing with respect to grammaticality violations that are (a) shared between MSA and spoken varieties, and (b) specific to MSA (such as word order violations, passive structures). This experimental work will focus on the grammaticality judgments of Heritage learners of Standard Arabic who speak a spoken variety, compared to non-Heritage learners who have had no other exposure to a spoken dialect. It is hoped that this work will permit an examination of the interrelationships between the two language varieties at the syntactic level.

To summarize, there are currently very few papers in existence that make use of neurocognitive methodologies for investigation of diglossia in Arabic. We contend that EEG provides a valuable means for examining, objectively and with millisecond timing resolution, the interplay between different levels of representation and processing in native speakers of Arabic. Behavioral measures do not provide the same kinds of insight, since linguistic knowledge is preconscious and not available for introspection. Although much groundwork has been laid for a descriptive framework within which we can examine the richness and variety of Arabic language varieties, neurocognitive approaches have the promise and potential to provide an additional means for rigorous and objective evaluation of hypotheses generated within existing theoretical frameworks.

3. Discussion: Advantages of breaking with tradition

Here we have provided a brief review of neurophysiological studies of code-switching between the two language varieties, along with some background information about the methods involved. It is our hope that this approach will continue to grow and develop. Alongside developmental studies of diglossic knowledge (e.g. Khamis-Dakwar, Gordon, Froud, 2012; Khamis-Dakwar & Makhoul, 2014), and the continuing work from various fields of linguistics, we believe that neurocognitive investigations of diglossia can provide us with evidence that the subsystems of the two language varieties are interacting with one another in observable ways – directly and indirectly – during development and in adulthood.

The effects of distance between high and low varieties of Arabic in the diglossic situation have been a matter of debate for many years, but without concurrent examination of the mental representation of the two language varieties, full understanding of the representation and processing of these two language varieties would be incomplete. We propose that due to its high temporal resolution, ERP has the potential to contribute to studies addressing unresolved issues and to advance this emerging field of research enabling the study of neural events associated with different sub-processes of language comprehension and production. With respect to the broader discussion of representation in Arabic diglossia (e.g., Badawi, 1973; Eid, 1990; Hary, 1996) we suggest that a reframing of the question is in order. It does not seem appropriate to continue to ask whether diglossic language systems involve one or two (or more) systems. Language is not a monolithic entity, and even the subsystems of language are appropriately broken down into component parts for rigorous examination and the furtherance of our understanding. Any approach to the neural representation of language must be framed at an appropriate level of abstraction. In our preliminary investigative work so far, we have asked questions rather broadly since the focus for much work on Arabic to date has been the question of how many linguistic systems are represented in diglossia. This has meant, for example, that the findings at the level of lexical representation have been interpreted to support a view of two distinct systems. A new direction for interpretation within a more fine-grained approach is clearly called for, alongside the integration of different sources of evidence and the interpretation of experimental data at all levels of linguistic representation and processing.

For example, we have reported here on evidence from neurophysiological investigations of Arabic diglossia at different levels of linguistic abstraction. These findings (the few available to date) reveal that, at the level of lexical representations, there appears to be a distinction between the Standard and spoken Arabic. If we combine these observations with previously available work on the development of diglossic competence (e.g., Khamis-Dakwar, Gordon, Froud, 2012; Khamis-Dakwar & Makhoul, 2014), we can hope to reveal more information about the organization of the two language varieties in native speakers of Arabic. The available data suggest a juxtaposition within the cognitive model of the two language varieties in which the two lexicons for the two language varieties are separate, but the grammatical system could be shared. In other words, we can combine evidence from neurophysiological investigations and studies of diglossic language acquisition to generate a view of a system of distinct levels of representation that interact with one another in very circumscribed ways (see figure 2 below for a highly simplified example). It could be the case that each language variety is represented at the lexical level as a distinct system, but that these interact with a syntactic component. Other configurations are empirically possible, but as a first step, a separate lexicon/shared syntax model could be used to catalyze further systematic studies of the interactions between linguistic systems in Arabic diglossia.

The idea of a continuum between the different language varieties has often been raised and may also have empirical utility as this research direction unfolds. However, one consideration should be that an empirically useful hypothesis is one that can be falsified (Popper, 1959). Whichever theoretical framework is adopted to inform our future investigations, it must be tested systematically through different experimental approaches, and validated – or falsified – through converging evidence from multiple investigations.

Further studies are needed to test the suggested model, especially with respect to the lack of separateness between the two language varieties at the grammatical level, and to evaluate the notion of a distinction between language varieties at the lexical level. Experimentally, we can envision methods that would permit elucidation of the unitary structure proposed for the grammatical system – for example,

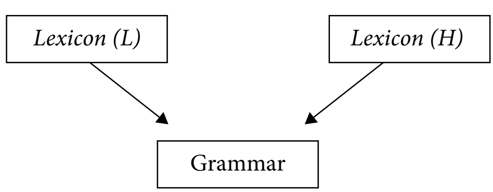


Figure 2. Simplified representational model of the two language varieties in Arabic.

recording brain responses to code-switched sentences where morphological and lexical features of the switched items are manipulated orthogonally. This kind of approach could permit us to evaluate whether grammatical code-switching has a neurophysiological effect beyond that associated with lexical code-switching, described above. We would also recommend the expansion of neurophysiological studies in this domain, to support further understanding of the neural processes associated with diglossic variation. This approach would supplement behavioral findings from language development, aphasia (lesion deficit studies), and code-switching in adults and children, and could offer a useful unifying research avenue for work with multiple populations: Arabic speaking adults in the Arab world, monolingual Arabic-speaking children learning MSA, and second language learners of Arabic who are being trained within different educational approaches. Ultimately, the findings from neurocognitive investigations should be mapped onto speech community variables, including language dominance, frequency of use, length of exposure, and data that have been collected revealing insights into individual and group attitudes towards the different varieties.

In this paper, we have described the beginnings of a research program that has the promise to delineate the neural underpinnings of what appears to be an unusually complex sociolinguistic situation that has resisted analysis in many ways and for many years. We suggest, based on the early findings within this new research domain for diglossia, that the diversity of research directions that are currently being applied to the complex and multilayered issues of linguistic representation and processing in Arabic diglossia can ultimately be unified. A deep understanding of the representation of, and interactions between, the two language varieties of Arabic is crucial, not only for greater understanding of Arabic itself, but because diglossia represents an outlier in terms of the human language experience. By delineating the representation and neural underpinnings of diglossia, within a unified yet diversified research program that incorporates every available research domain, we can as a research community make a very real contribution to our knowledge of human language, human interaction, and human nature.

References

American Speech-Language-Hearing Association. (2007). *Childhood apraxia of speech* [Technical report]. Available fromwww.asha.org/policy

Badawi, E. (1973). *Mustawayaat al-ʿarabiyya al-muʿaaṣira fi Miṣr* [Levels of contemporary Arabic in Egypt]. Cairo: Dar Al-Ma'aarif

Boudelaa, S., Pulvermüller, F., Hauk, O., Shtyrov, Y., & Marslen-Wilson, W. D. (2010). Arabic morphology in the neural language system: A mismatch negativity study. *Journal of Cognitive Neuroscience*, 22, 998–1010. DOI: 10.1162/jocn.2009.21273

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, 1, 351–353. DOI: 10.1038/1561

Conboy, B. T., Rivera-Gaxiola, M., Silva-Pereyra, J. F., & Kuhl, P. K. (2008). Event-related potential studies of early language processing at the phoneme, word, and sentence levels. In A. D. Friederici, & G. Thierry (Eds.), *Early language development: bridging brain and behavior*. Trends in language acquisition research series, (Vol. 5, pp. 24–64). Amsterdam: John Benjamins.

Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience*, 12, 635–647. DOI: 10.1162/089892900562390

Dehaene-Lambertz, G.(1997). electrophysiological correlates of categorical phoneme perception in adults. *Neuroprt*,8, 919–924

Eid, M. (1990). Arabic linguistics: The current scene. In M. Eid (Ed.), *Perspectives on Arabic Linguistics I* (pp. 3–37). Amsterdam: John Benjamins. DOI: 10.1075/cilt.63.03eid

Ferguson, C. (1959). Diglossia. *Word*, 15, 325–340.

Friedrici, A.D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6, 78–84. DOI: 10.1016/S1364-6613(00)01839-8

Friederici, A.D., Hahne, A., & Saddy, D. (2002). Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. *Journal of Psycholinguistic Research*, 31, 45–63. DOI: 10.1023/A:1014376204525

Froud, K., & Khamis-Dakwar, R. (2012). Mismatch negativity responses in childhood apraxia of speech. *American Journal of Speech-Language Pathology*, 21, 302–312. DOI: 10.1044/1058-0360(2012/11-0003)

Hahne, A., & Friedrici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, 11, 194–205. DOI: 10.1162/089892999563328

Hary, B. (1996). The language continuum in Arabic multiglossia. In: A. Elgibali (Ed.), *Understanding Arabic: Essays in contemporary Arabic Linguistics in honor of El-Said Badawi* (pp. 69–90). Cairo: American University in Cairo Press.

Jackson, G., Swainson, R., Mullin, A., Cunnington, R., & Jackson, S. (2004). ERP correlates of receptive language-switching tasks. *The Quarterly Journal of Experimental Psychology, 57A*, 223–240. DOI: 10.1080/02724980343000198

Jackson,G.M., Swainson, R., Cunnington, R. & Jackson, S.R. (2001). ERP correlates of executive control during repeated language switching. *Bilingualism: Language and Cognition*, 4, 169–178

Kappenman, E. S., & Luck, S.J. (2012). ERP components: the ups and downs of brainwave recordings. In S. J. Luck, & E. S. Kappenman (Eds.), *Oxford handbook of event-related potential components* (pp.1–30). New York, NY: Oxford University Press.

Khamis-Dakwar, R., & Froud, K. (2007). Lexical processing in two language varieties: An event related brain potential study of Arabic native speakers. In M. Mughazy (Eds.), *Perspectives on Arabic Linguistics XX*. Amsterdam: John Benjamins.

Khamis-Dakwar, R., Froud, K., & Boudelaa, S. (2009). Lexical processing in diglossic codeswitching between Modern Standard Arabic and Palestinian Colloquial Arabic: An event-related brain potential study of Arabic native speakers. Poster presented at Cognitive Neuroscience Symposium. San Francisco, March 2009.

Khamis-Dakwar, R., Froud, K., & Gordon, P. (2012). Acquiring diglossia: Mutual influences of formal and colloquial Arabic on children's grammaticality judgments. *Journal of Child Language*, 39, 1–29. DOI: 10.1017/S0305000910000784

Khamis-Dakwar, R., & Makhoul, B. (2014). Arabic diglossia and its implications for language learning disability assessment. In Saeigh-Haddad, & Joshi, M. (Eds.), *Handbook of Arabic Literacy: Insights and Perspectives s:* (pp. 279–303). Netherlands. New York, NY: Springer.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203–205. DOI: 10.1126/science.7350657

Kutas, C., Van Petten, C., & Kluender, R. (2006). Psycholinguistics electrified II: 1994–2005. In M. Traxler, & M.A. Gernsbacher (Eds), *Handbook of psycholinguistics* (2nd edn, pp. 659–724). New York: Elsevier. DOI: 10.1016/B978-012369374-7/50018-3

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, 9, 920–933. DOI: 10.1038/nrn2532

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

Luck, S.J., & Kappenman, E. S. (Eds.). (2012). *Oxford handbook of event-related potential components*. New York, NY: Oxford University Press.

Moreno, E., Federmeier, K., & Kutas, M. (2002). Switching languages, switching palabras (words): An electrophysiological study of codeswitching. *Brain and Language*, 80, 188–207. DOI: 10.1006/brln.2001.2588

Moreno, E. M., & Kutas, M. (2005). Processing semantic anomalies in two languages: An electrophysiological exploration in both languages of Spanish-English bilinguals. *Cognitive Brain Research*, 22, 205–220. DOI: 10.1016/j.cogbrainres.2004.08.010

Näätänen, R. (1999). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, 38, 1–21. DOI: 10.1111/1469-8986.3810001

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Houtilainen, M., Iivonen, A., Vainio, A., Alku, P., Ilmoneiemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385, 432–434. DOI: 10.1038/385432a0

Näätänen, R., Kujala, T., & Winkler, I. (2010). Auditory processing that leads to conscious perception: A unique window to central auditory processing opened by the mismatch negativity and related responses. *Psychophysiology*, 48(1): 4–22 DOI: 10.1111/j.1469-8986.2010.01114.x

Näätänen, R. & Kreegipuu, K. (2012). The Mismatch Negativity (MMN). In S. J Luck, & E.S. Kappenman (Eds.), *Oxford handbook of event-related potential components* (pp. 143–158). New York, NY: Oxford University Press.

Ojimas, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of profeciency. *Journal of Cognitive Neuroscience*, 17 , 121–1228.

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 786–803. DOI: 10.1037/0278-7393.20.4.786

Popper, K. (1959). *The logic of scientific discovery*. Reissued 2002. London, UK: Routledge.

Pulvermuller, F., Kujala, T., Shtyrov, Y., Simola, J., & Tiitinen, H. (2001). Memory traces for words as revealed by the Mismatch Negativity (MMN). *NeuroImage*, 14, 607–616. DOI: 10.1006/nimg.2001.0864

Rugg, M. D. (1997). *Cognitive neuroscience*. Hove, East Sussex: Psychology Press.

Sharma, A., & Dorman, M. F. (2000). Neurophysiological correlates of cross-language phonetic perception. *Journal of the Acoustical Society of America*, 107, 2697–2703. DOI: 10.1121/1.428655

Simon, G., Bernard, C., Lalonde, R., & Rebai, M. (2006). Orthographic transparency and grapheme–phoneme conversion: An ERP study in Arabic and French readers. *Brain Research*, 1104, 141–152. DOI: 10.1016/j.brainres.2006.05.071

Swaab, T.Y., Ledoux, K., Christine, C., & Boudewyn, M. (2012). Language-related ERP components. In S. J. Luck, & E. S. Kappenman, (Eds.), *Oxford handbook of event-related potential components* (pp. 397–440). New York, NY: Oxford University Press.

Weber-Fox, C., & Neville, H. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience*, 8, 231–256. DOI: 10.1162/jocn.1996.8.3.231

Index