

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

## Afroasiatic. Data and perspectives

### This is the author's manuscript

*Original Citation:*

*Availability:*

This version is available <http://hdl.handle.net/2318/1660634> since 2018-02-23T17:27:22Z

*Publisher:*

John Benjamins

*Terms of use:*

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

AFROASIATIC

**PAGE PROOFS**  
© JOHN BENJAMINS PUBLISHING COMPANY

**2nd proofs**

## CURRENT ISSUES IN LINGUISTIC THEORY

AMSTERDAM STUDIES IN THE THEORY AND HISTORY  
OF LINGUISTIC SCIENCE – Series IV

ISSN 0304-0763

General Editor

JOSEPH C. SALMONS  
University of Wisconsin–Madison  
jsalmons@wisc.edu

Founder & General Editor (1975-2015)

E.F.K. KOERNER  
Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin  
efk.koerner@rz.hu-berlin.de

*Current Issues in Linguistic Theory* (CILT) is a theory-oriented series which welcomes contributions from scholars who have significant proposals that advance our understanding of language, its structure, its function and especially its historical development. CILT offers an outlet for meaningful contributions to current linguistic debate.

A complete list of titles in this series can be found on <http://benjamins.com/catalog/cilt>

### *Editorial Board*

Claire Bowern (New Haven, Ct.)  
Sheila Embleton (Toronto)  
Elly van Gelderen (Tempe, Ariz.)  
Matthew Juge (San Marcos, Tex.)  
John E. Joseph (Edinburgh)  
Martin Maiden (Oxford)  
Martha Ratliff (Detroit, Mich.)  
E. Wyn Roberts (Vancouver, B.C.)  
Klaas Willems (Ghent)

**Volume 339**

Mauro Tosco (ed.)

*Afroasiatic. Data and perspectives*

**2nd proofs**

# AFROASIATIC

DATA AND PERSPECTIVES

*Edited by*

MAURO TOSCO

*University of Turin*

JOHN BENJAMINS PUBLISHING COMPANY  
AMSTERDAM & PHILADELPHIA

**2nd proofs**



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.

DOI 10.1075/cilt.339

**Cataloging-in-Publication Data available from Library of Congress:**  
LCCN 2017045536 (PRINT) / 2017050051 (E-BOOK)

ISBN 978 90 272 0012 9 (HB)  
ISBN 978 90 272 6457 2 (E-BOOK)

© 2018 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Company · <https://benjamins.com>

# Table of contents

## INTRODUCTION

- Afroasiatic: Fresh insights from an “old” language family 1  
*Mauro Tosco*

## Part I. Afroasiatic: Classification and typology

- Did Proto-Afroasiatic have marked nominative  
or nominative-accusative alignment? 11  
*Helmut Satzinger*

- The limits and potentials of cladistics in Semitic 23  
*Petr Zemánek*

- Lexicostatistical evidence for Ethiosemitic, its subgroups, and borrowing 41  
*Grover Hudson*

## Part II. Forms and functions

- Reconsidering the ‘perfect’–‘imperfect’ opposition  
in the Classical Arabic verbal system 61  
*Michal Marmorstein*

- The imperfective in Berber: Evidence of innovated forms and functions 85  
*Mena Lafkioui*

- Condition, interrogation and exception: Remarks on particles in Berber 105  
*Catherine Taine-Cheikh*

- The semantics of modals in Kordofanian Baggara Arabic 131  
*Stefano Manfredi*

## Part III. Predication and beyond

- Insubordination in Modern South Arabian: A common isogloss  
with Ethiosemitic? 153  
*Olga Kapeliuk*

Possessive and genitive constructions in Dahālik (Ethiosemitic) <i>Marie-Claude Simeone-Senelle</i>	167
The characterization of conditional patterns in Old Babylonian Akkadian <i>Eran Cohen</i>	185
Locative predication in Chadic: Implications for linguistic theory <i>Zygmunt Frajzyngier</i>	203
Unipartite clauses: A view from spoken Israeli Hebrew <i>Shlomo Izre'el</i>	235
The Interaction of state, prosody and linear order in Kabyle (Berber): Grammatical relations and information structure <i>Amina Mettouchi</i>	261
Index	287

## INTRODUCTION

### Afroasiatic

#### Fresh insights from an “old” language family

Mauro Tosco

University of Turin

This volume contains a selection of papers originally presented at the 14th Italian Meeting of Afroasiatic Linguistics in Turin (15–18 June, 2011). They have been chosen in order to give the interested reader an updated (although by necessity incomplete) comparative view of most branches of Afroasiatic and of the breadth of theoretical and empirical research being carried on. The articles are intended therefore to be representative of a whole gamut of interests which focus on Afroasiatic, from the presentation of new data, often from scarcely known varieties (be it Semitic – as in the case of the Kordofanian Baggara Arabic – Berber, or Chadic) to a sophisticated linguistic analysis of old debates (such as the value of the Classical Arabic verbal forms).

I have grouped the articles into three broad areas of interest: the family as such, in its classificatory but also typological aspects; the analysis of the intricate morphology of Afroasiatic and its developments; and syntax in a wide sense, from the clause to the sentence and beyond. Many years ago, Hans-Jürgen Sasse remarked how Semitic linguistics developed a strong typological orientation very early, to the detriment of reconstruction, and he lamented the fact that the concept of Semitic appears to be more of a typological than a genetic unit in the eyes of many scholars (Sasse 1981: 131); in those very same years, Sasse published his still unsurpassed reconstruction of East Cushitic phonology (Sasse 1979). While we do not engage in this work on phonological reconstruction, classification and reconstruction play a big role, especially in the first part of the book.

The volume opens with Helmut Satzinger's article on the syntactic alignment of the protolanguage. It is also the only contribution specifically addressing Proto-Afroasiatic, and this is all the more interesting as work at the macro-family level has been at a standstill for many years now. Interest in the marked nominative character of Afroasiatic has instead been revived in recent years, with König (2006) and especially Frajzyngier & Mettouchi (2013), who expanded and corrected Sasse's (1984) seminal work on the Afroasiatic case. On the basis of a thorough



analysis of the pronominal series of the different family groups, Satzinger's conclusion is that a nominative-absolutive alignment can still be reconstructed for all the families of Afroasiatic with the exception of Semitic. Satzinger notices how "the nominative-accusative one is the most widely spread among all languages, followed by the ergative-absolutive alignment. The nominative-absolutive alignment is extremely rare in comparison". The author hypothesizes that "[T]he nominative-absolutive alignment of Afroasiatic may be as old as Proto-Afroasiatic, or it may have developed from an ergative-absolutive alignment" while Semitic nominative-accusative is a later development "from the Afroasiatic nominative-absolutive alignment, as it still contains conspicuous remnants of it".

Classificatory problems and methodologies lie at the core of Petr Zemánek's "The Limits and Potentials of Cladistics in Semitic". The author explores the possibility of exclusively using grammar data in the construction of a phylogenetic tree of Semitic and its visualization as NeighborNet networks, and he does so using lists of grammatical features by Faber (1997), by Gai (1994), and by himself. The results, though mixed, are promising: areality plays a big and possibly disturbing role, with four areas: Mesopotamia, Syro-Palestine, South Arabia, and Ethiopia. The position of Arabic within the Central Semitic group is confirmed and given new strength, while the results are less clear-cut in the case of Modern South Arabian. In general, the languages of the Arabian Peninsula (with the exception of Arabic, which consistently points more to the North than to the South) show an "unstable behavior".

The necessity to integrate grammatical markers and the lexicon is recognized by Zemánek; in the following article, "Lexical Evidence for Ethiosemitic, its Subgroups, and Borrowing", Grover Hudson builds upon his recent volume (Hudson 2013) on the lexicon of the Semitic languages of the Horn of Africa (variously called Ethiosemitic, Afrosemitic, or Northeast African Semitic) to further elaborate on their internal classification on the basis of their lexical stock. The author uses a time-honored (and much criticized) lexicostatistical framework (both on the basis of a 98-word list and a longer 250-word list), but enriching it with a sophisticated analysis of internal borrowing. His results show that Hetzron's (1972) internal classification of Ethiosemitic, although basically confirmed, needs revision in a few lower-level branches. The problems concentrate, as might be expected, within the tightly-knit cluster of the Gurage languages, and they prompt the author to propose a revised classification. Another important result of the study goes much beyond internal classificatory problems, insofar as it provides "helpful quantification of, and little support for, the traditional idea that ES (Ethiosemitic) has been unusually influenced by Agaw". In the light of the commonly held view (cf. Thomason & Kaufman 1988), which sees in Ethiosemitic an example of extreme contact-induced change, Hudson's results, although limited to basic vocabulary, cannot be easily dismissed. They further seem to go hand in hand with recent work

on the shared features of the languages of the Horn, which, after Tosco's (2000) critique of Ferguson's (1976) establishment of Ethiopia as a language area, has seen a recent resurgence of studies on more regional language areas (as already in Sasse 1986 for the Dullay-Konsoid – and partially Omotic – interaction in southwest Ethiopia) or fine-grained analyses of common grammatical features and developments (as in Crass & Meyer 2007).

The second part of the volume delves on the morphosyntax of specific Afroasiatic varieties; this section opens with Michal Marmorstein's “Reconsidering the ‘perfect’–‘imperfect’ Opposition in the Classical Arabic Verbal System”. The author successfully copes with a discussion as ancient as Arabic studies – the ‘core’ value of the verbal forms. Through a thorough analysis of the use of the two paradigms in different syntactic environments, their compatibility with different particles, word-order facts (with the opposition between the verbal and the nominal clause), lexical classes, and textual domains, Marmorstein shows that the opposition between perfect *fa'ala* and imperfect *yaf'alu* is not obtained in any environment and, where it does, it serves to indicate several semantic distinctions. Often, the opposition applies between *fa'ala*, on the one hand, and not only *yaf'alu*, but also *qad fa'ala* and the participle on the other. The author concludes that the complexity of the system, which, besides *fa'ala* and *yaf'alu*, consists of many other forms, cannot be reduced to a single temporal or aspectual dichotomy and a single label or ‘core value’.

Morphological change lies at the heart of Mena Lafkioui's “The Imperfective in Berber: Evidence of Innovated Forms and Functions”, which analyzes developments in the Berber verbal system with a focus on innovations in the North (Tarifit of Northern Morocco) and the South (Tuareg). The core of the argument is that mutual or external contact can safely be excluded: changes were system-internal and driven mainly by functional parameters and the morphological expression of pragmatic or semantic distinctions (such as habituality for punctual verbs and durative/intensive values with non-punctual verbs).

The second contribution on Berber is Catherine Taine-Cheikh's “Condition, Interrogation, and Exception: Remarks on Particles in Berber”, which offers both an overview and a detailed analysis of the particles used to express condition in different Berber languages. Areal convergence and variation is observed, and different grammaticalization paths are detailed. Convergence is detected in five regions of the Berber domain, and this may be expressed with a classical wave propagation model. Semantic convergence is found in the frequent connection of the particles introducing conditionals with those used in interrogative clauses and those expressing exception. The author remarks on how Berber brings evidence for three well-known different patterns for the grammaticalization of conditional particles (namely, from a copula, a marker of polar questions, and a temporal marker).

Stefano Manfredi, in “The Semantics of Modals in Kordofanian Baggara Arabic”, brings us to a largely unknown Sudanese Arabic dialect. Kordofanian Baggara Arabic follows Eastern Arabic dialects in its use of *b(i)=* with imperfective verbal forms when non-modal and in its absence from modal contexts. The paper aims at drawing a polysemantic account of modal auxiliaries in light of the participant-oriented approach proposed by Van der Auwera & Plugian (1998). We find in this dialect the morphologization of *bukūn* “he will” from the 3sg.M imperfective form of the verb ‘be’ preceded by preverbal *b(i)=* (*\*b=i-kūn* “he is, he will be”) and its grammaticalization to the expression of epistemic necessity (‘must’): this fully conforms to the path ‘future’ > ‘epistemic necessity’ proposed by van der Auwera & Plugian. But Kordofanian Baggara Arabic also displays a semantic specification from general participant-external possibility to deontic possibility rather than a semantic generalization the other way round. As is often the case, one is reminded here of the inherent weakness of too many typological generalizations, which are often based upon insufficient data sets.

The clause, the sentence, and the text are the focus of the last section, which opens with Olga Kapeliuk’s “Insubordination in Modern South Arabian: A Common Isogloss with Ethiosemitic?” The South Arabian and Ethiosemitic type of insubordination addressed here is found in the common use of imperfect or perfect verbal forms, subordinated by the relative particle *d-* and acting as main verbs. Following an insight by Pennacchietti (1993), this insubordinate use is interpreted in Modern South Arabian as implying the presence of a zero copula, while an overt copula is always present in the corresponding Ethiosemitic examples. The neat parallel between the Semitic languages on both sides of the Red Sea forces Kapeliuk to question the role of the Cushitic substrate in the very shaping of Ethiosemitic. The reader is immediately reminded of the low number of Cushitic loanwords in the basic vocabulary of Ethiosemitic discussed earlier in this volume by Hudson, and of the latter’s remark against the traditional and commonly-held hypothesis which sees in Ethiosemitic “a secondary population in northeast Africa”. Obviously, this tallies well with the frequently noted fact that linguistic diversity within Afroasiatic in the Horn of Africa is so great that this area could well be the cradle of the whole family.

Geographically akin is Marie-Claude Simeone-Senelle’s “Possessive and Genitive Constructions in Dahālik (Ethiosemitic)”. A common feature of Modern South Arabian and Dahalik is the restriction and possible fossilization of the Semitic Construct State (the direct annexation of the possessum followed by the possessee, which can be determined). This, of course, is just another instance of a trend widely attested in Semitic, and in Ethiosemitic in particular. Following again another attested tendency, we witness the rise of an ‘analytic’ construction with a relator between the two terms. The relator itself takes different forms (much as it

happens with the analytical genitives of Spoken Arabic, and against the unitary treatment of *d-* or *-d* in Aramaic) and is generally in use for alienable possession. Soqotri is highly original in having the analytic construction only with pronominal possession, and with the reversed order Poss N.

Eran Cohen’s “The Characterization of Conditional Patterns in Old Babylonian Akkadian” investigates the use of the connective particle *=ma* in conditional clauses in Old Babylonian. *=ma* creates an asymmetrical connection between two clauses which show otherwise what Cohen calls a “modal congruence”. The syntactic patterns investigated by the author are examples of those paratactic conditionals which by definition are seemingly devoid of specific characterization as conditionals and in which the connective does not contribute any specific meaning. Syntactic and semantic features of the paratactic construction having conditional value are singled out and compared with the markedly different paratactic circumstantial construction. A sound description of this specific domain of Akkadian syntax is presented by Cohen as an important step toward the cross-linguistic understanding of paratactic conditional patterns.

In “Locative Predication in Chadic: Implications for Linguistic Theory”, Zygmunt Frajzyngier provides a detailed account of locative expressions in languages belonging to Central, West, and East Chadic, in order to demonstrate locative predication as a category in Proto-Chadic. In general, if the clause does not have an inherently locative predicate, a locative predicator is used, or a serial verb construction, or still other means. For a non-inherently locative complement, a generic locative preposition is used. Frajzyngier details the presence of locative predicators (defined as “a predicate whose sole function is to serve as a locative predicate when the predicate of the clause is not inherently locative and the clause aims to convey the locative predication”) in all three branches of Chadic. In so doing, he further proves that elements such as Hausa and Mupun *a* are not prepositions but locative predicators, and can arrive at an independent, not contact-induced explanation of the presence of serial verb constructions in a language such as Lele with non-inherently locative verbs. Frajzyngier finally postulates the presence of a locative predicate in Proto-Chadic on the reasoning that “it is less likely that languages from three branches have independently grammaticalized locative predication, which is otherwise typologically rare, than the possibility that some languages from each branch have retained a function from the Proto-Chadic”.

Shlomo Izre’el’s contribution is much more than its name implies. “Unipartite Clauses: A View from Spoken Israeli Hebrew” is a neat, coherent presentation of a revolutionary approach to the ‘sentence’ in spoken human language. Building on his previous paper (Izre’el 2012), and recognizing his debt to much French linguistic thought (foremost to Tesnière), the author defines a sentence as a unit consisting minimally of a predicate. Unipartite sentences, which are the object of the article,

consist of a predicate only. They convey new information and carry sentential information load and modality. The plethora of one-word utterances so common in spontaneous spoken language and usually regarded as elliptical, reduced or concise syntactic structures, are finally given their proper status as full-fledged sentences in Izre'el's account. This is done on the basis of their intonational behavior, as they build intonation groups (or units) of their own. Building on the *Corpus of Spoken Israeli Hebrew*, Izre'el further proposes a thorough classification of unipartite sentences, starting from their status as anchored ("in referential expressions beyond the sentence domain") or unanchored, and he exemplifies possible types. This brief synopsis barely hints at the enormous potential of Izre'el's theory if and when it is tested on other languages, as suggested by the author himself in his conclusions.

In "The Interaction of State, Prosody, and Linear Order in Kabyle (Berber): Grammatical Relations and Information Structure", Amina Mettouchi ideally connects to Frajzyngier's chapter in assuming a strong non-aprioristic view of grammar and with Izre'el's chapter in her attention to prosodic structures; the author successfully brings together information structure, prosody, and morphosyntax, showing that

- (a) the state opposition in itself does not mark grammatical relations; (b) coreference in gender and number between the noun and the bound pronoun, in itself, is not transparent for the encoding of grammatical relations; (c) word order in itself does not mark grammatical relations. However, the interaction of state, word order, and prosodic grouping allows the computation of grammatical relations for nouns.

Functions are therefore marked by the interaction of a plurality of coding jointly marking a value.

The Afroasiatic language family is in many aspects unique: more than Indo-European, it brings together a minority of languages whose records date from the beginnings of written history, languages with a unique time-span of continuous data, encompassing at times several millennia (such as Egyptian and Aramaic) and, on the other hand, a bewildering number of still scarcely investigated languages, all too often spoken by dwindling communities (as is the case of much of Chadic). The balance between these two extremes is difficult to strike: data, methodologies, and the whole mind frames of the specialists are too often different.

It is no surprise, therefore, that this brief overview does not do justice to the richness of the volume: but its sheer amount of otherwise unknown or scarcely accessible data, and of mind-provoking concepts and insights, is there, an apt reminder of how much still lies ahead waiting for discovery and appraisal in this, at the same time oldest and very new, language family.

## References

- Crass, Joachim & Ronny Meyer (eds.). 2007. *Deictics, Copula and Focus in the Ethiopian Convergence Area*. Köln: Rüdiger Köppe.
- Faber, Alice. 1997. “Genetic Subgrouping of the Semitic Languages”. *The Semitic Languages*, ed. by Robert Hetzron, 3–15. London: Routledge.
- Ferguson, Charles A. “The Ethiopian Language Area”. *Language In Ethiopia*, ed. by M. Lionel Bender, J. Donald Bowen, R. L. Cooper & Charles A. Ferguson, 63–76. Oxford: Oxford University Press.
- Frajzyngier, Zygmunt & Amina Mettouchi. 2013. “A Previously Unrecognized Typological Category: The State Distinction in Kabyle (Berber)”. *Linguistic Typology* 17:1.1–20. doi:10.1515/lity-2013-0001
- Gai, Amikaim. 1994. “Reflections on the Classification of the Semitic Languages”. *Orbis* 37.135–144. doi:10.2143/ORB.37.0.2012770
- Hetzron, Robert. 1972. *Ethiopian Semitic: Studies in Classification*. (Journal of Semitic Studies monograph 2). Manchester: Manchester University Press.
- Hudson, Grover. 2013. *Northeast African Semitic: Lexical Comparisons and Analysis*. Wiesbaden: Harrassowitz.
- Izre’el, Shlomo. 2012. “Basic Sentence Structures: A View from Spoken Israeli Hebrew”. *Penser les langues avec Claire Blanche-Benveniste*, ed. by Sandrine Caddéo, Marie-Noëlle Roubaud, Magali Rouquier & Frédéric Sabio, 215–227. Aix-en-Provence: Presses Universitaires de Provence. Corrected version can be downloaded at [http://www.tau.ac.il/~izreel/publications/SentenceStructure\\_BlancheMem.pdf](http://www.tau.ac.il/~izreel/publications/SentenceStructure_BlancheMem.pdf)
- König, Christa. 2006. “Marked Nominative in Africa”. *Studies in Language* 30:4.655–732. doi:10.1075/sl.30.4.02kon
- Pennacchietti, Fabrizio A. 1993. “Le forme verbali pseudorelative: isoglossa strutturale del semitico sud-occidentale”. *Semitica – Serta Philologica Constantini Tsereteli Dicata*, ed. by Riccardo Contini, Fabrizio A. Pennacchietti & Mauro Tosco, 213–225. Torino: Zamorani.
- Sasse, Hans-Jürgen. 1979. “The Consonant Phonemes of Proto-East-Cushitic (PEC): A First Approximation”. *Afroasiatic Linguistics* 7.1.
- Sasse, Hans-Jürgen. 1981. “Afroasiatisch”. *Die Sprachen Afrikas*, ed. by Bernd Heine, Thilo C. Schadeberg & Ekkehard Wolff, 129–148. Hamburg: Helmut Buske.
- Sasse, Hans-Jürgen. 1984. “Case in Cushitic, Semitic and Berber”. *Current Progress in Afro-asiatic Linguistics: Papers of the Third International Hamito-Semitic Congress*, ed. by James Bynon, 111–126. Amsterdam: John Benjamins. doi:10.1075/cilt.28.08sas
- Sasse, Hans-Jürgen. 1986. A Southwest Ethiopian Language Area and Its Cultural Background. *The Fergusonian Impact*, vol. 1 ed. by Joshua A. Fishman, Andrée Tabouret-Keller, Michael Clyne, Bh. Krishnamurti & Mohamed Abdulaziz, 327–342. Berlin: Mouton de Gruyter.
- Thomason, Sarah & Terrence Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- Tosco, Mauro. 2000. “Is There an ‘Ethiopian Language Area’?” *Anthropological Linguistics* 42:3.329–365.
- Van der Auwera, Johan & Vladimir A. Plungian. 1998. “Modality’s Semantic Map”. *Linguistic Typology* 2.79–124. doi:10.1515/lity.1998.2.1.79

**PAGE PROOFS**  
© JOHN BENJAMINS PUBLISHING COMPANY

**2nd proofs**

PART I

## **Afroasiatic**

Classification and typology

**PAGE PROOFS**  
© JOHN BENJAMINS PUBLISHING COMPANY

**2nd proofs**



**PAGE PROOFS**

© JOHN BENJAMINS PUBLISHING COMPANY

**2nd proofs**

Did Proto-Afroasiatic  
have marked nominative  
or nominative-accusative alignment?

Helmut Satzinger  
Universität Wien

The case system of Berber and Cushitic displays nominative-absolutive alignment, or a marked nominative system, a feature intermediary between ergative-absolutive and nominative-accusative alignment. Comparison of the cases of the nouns with the paradigms of the personal pronoun:

- A. independent / predicative (etc.) pronoun,
- B. dependent / object (etc.) pronoun,
- C. suffix / genitival (etc.) pronoun.

The two paradigms of the Semitic pronoun (independent and suffix pronoun) are a reduction of an original tripartite system, the suffix pronoun assuming functions of the dependent pronoun.

The numerous paradigms of the Berber and Chadic pronouns can be reduced to the same three basic paradigms. They can on their part be reduced to two, insofar as the independent and the dependent pronouns are ultimately of the same origin.

Correlation of noun cases and pronoun paradigms:

Proto-Afroasiatic		
Noun	Absolutive case	Nominative case
Pronoun	1. Absolute pronoun (A),	1. Suffix pronoun (C),
	2. dependent pronoun (B)	2. conjugation morpheme

**Keywords:** Afroasiatic cases, syntactic alignment, nominative-absolutive, alignment, marked nominative, paradigms of personal pronoun

## 1. Cases in Berber and Cushitic

The Berber noun appears in two forms: one being called the ‘absolute’ state, viz. (sing.:) m. *a-*, fem. *ta-*, and the other the ‘annexed’ state, viz. (sing.:) m. (*w*)*u-*, f. *t-*. Traditionally, Berberologists regard the functional difference between them as one of status: „In besonderer syntaktischer Umgebung, wenn das Nomen in engem Zusammenhang mit dem voranstehenden Wort steht, nimmt es die Form des ‚status annexus‘ (im Gegensatz zu ‚status absolutus‘ = Zitierform) an...“ (Wolff 1981: 179).<sup>1</sup>

However, Hans-Jürgen Sasse, in his seminal 1984 study, pointed out a different interpretation. In light of the Cushitic noun, the two forms appear to be two different cases, i.e., morpho-syntactic phenomena, rather than prosodic features, as are statuses. The function of the “absolute state” form is very similar to that of the Cushitic absolutive case, whereas the “annexed state” form corresponds by and large to the Cushitic nominative case; Lipiński (2001: 35, 179, 260ff.) wrongly calls it an **ERGATIVE CASE** (cf. Satzinger 2005). The use of a noun in the absolute state, or absolutive case, rather than the annexed state, or nominative case, for a subject that is positioned to the right of the verb is not due to the prosodic status of the noun, but rather to its being fronted, or topicalized.

## 2. The nominative-absolutive alignment, or marked nominative system

There are several types and varieties of syntactic alignment for basic constituents of the sentence (Comrie 2013). The alignment that is more familiar to us, and also the most frequently found among the world’s languages is nominative-accusative alignment (such as in most European and Semitic languages). Somewhat rarer, and extremely rare in Europe, is ergative-absolutive alignment, where the subject of an intransitive verb appears in the same case form (called **ABSOLUTIVE CASE**) as the object of a transitive verb, whereas there is a special case for the subject of transitive verbs (**ERGATIVE CASE**), the marked form. In Europe, only Basque has this system, while most Australian aboriginal languages and several American languages (like Mayan) do, along with Tibetan, Sumerian, and others, not to mention languages with ‘split ergativity’, like Iranian languages. A still much rarer variety has a portmanteau case (called **ABSOLUTIVE CASE**) for most syntactic functions, including nominal predicate and direct object, though not for the subject of the intransitive verb. Furthermore it has a marked case (the **NOMINATIVE**) for the subject of the verb, regardless whether it is transitive or intransitive), called nominative-absolutive

1. “In a particular syntactic environment, when the noun is closely related to the preceding word, it assumes the form of the ‘status annexus’ (in contrast to ‘status absolutus’ = citation form) ...”

alignment, or marked nominative system. Orin Gensler first showed that both Berber and Cushitic were perfect representatives of this marked nominative type. In defence and support of Sasse’s (1984) paper, he presented this analysis several times (Gensler 2000), though the paper was never published.<sup>2</sup>

In an ergative-absolutive case system, the ergative is functionally marked. In a nominative-accusative system, the accusative is functionally marked. In a nominative-absolutive system, or marked nominative system (Dixon 1994: 63ff.), the nominative, rather than the accusative, is the marked member in terms of function. This means *inter alia* that the accusative is the ‘citation form’, which includes for Gensler “all forms for which case marking can be seen as irrelevant (... fronted topic; emphatic in-situ...)” and we would add to this the predicative function, and the focus.

- (1) Nominative-Accusative: John saw Bill<sup>marked</sup>. John ran.  
Nominative-Absolutive: John<sup>marked</sup> saw Bill. John<sup>marked</sup> ran.  
Ergative-Absolutive: John<sup>marked</sup> saw Bill. John ran.

The following table presents an overview comparison of the functions of the three alignments.

Table 1. Syntactic alignment in comparison

Ergative-Absolutive:	Nominative-Absolutive:	Nominative-Accusative:
<b>Absolutive</b> (unmarked) quotation, address, predicate, focus, topic subject of intransitive verbs object of transitive verbs adverbial form	<b>Absolutive</b> (unmarked) quotation, address, predicate, focus, topic object of transitive verbs adverbial form	<b>Accusative (marked)</b> object of transitive verbs adverbial form
<b>Ergative (marked)</b> agent (subject) of transitive verbs	<b>Nominative (marked)</b> subject of intransitive verbs, agent (subject) of transitive verbs	<b>Nominative (unmarked)</b> quotation, address, predicate, focus, topic subject of intransitive verbs, agent (subject) of transitive verbs
E.g., Basque	Berber, East Cushitic Otherwise, e.g., Central Nilotic	Semitic Otherwise, e.g., Indo-European

2. I am grateful to the author for bringing to my attention the manuscript and the accompanying data in July 2011, after the 14th Italian Meeting of Afroasiatic Linguistics, Turin, June 15–17, 2011.

The nominative-absolutive alignment of the noun has been preserved in Cushitic (plus Omotic) languages and in Berber. Case distinction is lacking in Egyptian, in Chadic, in some Eastern Berber varieties, in some ancient (Hebrew, Aramaic) and nearly all modern Semitic languages. Nevertheless, the alignment can be determined by other features, in particular, by the pronominal system.

Most remarkably, the nominative-absolutive alignment was converted into a nominative-accusative alignment in Semitic (Akkadian, Classical Arabic, Geez), although some noticeable features can best be explained as being remnants of the old system (Sasse 1984).

### **3. The personal pronoun in languages with nominative-absolutive alignment**

The above deals with nouns, but the situation of personal pronouns seems quite different. Many varieties of Berber have case marking on nouns (two cases, otherwise called *STATUSES*; marked by change of prefix vowel), though some kinds of Eastern Berber, like Awjili or Ghadamsi, do not (Kossmann 2013; van Putten 2014; others, like Nefusi, Siwi, Foqahi, Sokni, mark cases only by accentual movement: Brugnatelli 1986). Nevertheless, all Berber varieties have the same personal pronoun paradigms, whether nouns indicate case forms or not. Where there is case marking, there are exactly two cases, as mentioned above. Nevertheless, there are some six paradigms of the personal pronoun. Hence, there is no clear-cut one-to-one relationship between noun cases (numbering two) and pronoun paradigms (numbering six). But they can be correlated.

### **4. The personal pronoun in Afroasiatic: Egyptian, Cushitic**

Afroasiatic has basically three pronominal paradigms (Satzinger 1991: 129ff.; 2003a):

- A. independent / predicative (etc.),
- B. dependent / object (etc.),
- C. suffix / genitival (etc.).

This is found in Egyptian, Cushitic and generally in Chadic. Only Semitic has reduced them to two, namely (A) independent, and (C) suffix.

For example, the personal pronoun in Egyptian (Table 2) and in the Cushitic language Saho (Table 3).

Table 2. The personal pronoun in Egyptian

	A. Independent pronoun (old series)	B. Dependent pronoun	C. Suffix pronoun
1sc	<i>ink</i> [ʔa'nak] *ʔanáV	<i>wi</i> *yuwa ?	=i [-i] *-i / -ya ?
2sm	<i>twt</i> [cu'wat]? *kuwátV?	<i>tw</i> [cuw]? *kuwa ?	=k [-ak] *-ka
2sf	<i>tmt</i> [ci'mat]? *kimátV?	<i>tn</i> [cin]? *kima ?	=t [-ic] *-ki
3sm	<i>swt</i> [su'wat]? *suwátV?	<i>sw</i> [suw]? *suwa ?	=f [-af] *-fu ? < *-hu ?
3sf	<i>stt</i> [si'tat]? *sitátV?	<i>sy</i> [siy]? *siya ?	=s [-as] *-sa
...	Quotation form; predicate ( <i>ink pw</i> "it is I") partly, subject of nominal predicate ( <i>ink sn=k</i> "I'm your brother") augens ( <i>m pr=i ink</i> "in my own house") ~ absolutive case	subject of adjectival predicate ( <i>nfr wi</i> "I am good") object ( <i>sdm wi</i> "hear me!") with thetic elements ( <i>m=k wi</i> "here I am!"; <i>nn sw</i> "he does not exist") ~ absolutive case	possessive=genitive ( <i>pr=i</i> "my house") prepositional ( <i>hr=f</i> "on him") subject of verbal predicate ( <i>sdm=i</i> "I shall hear") ~ nominative / genitive case

As an example of the personal pronoun in Cushitic:

Table 3. The personal pronoun in Saho (East Cushitic; Banti 2003: 22; also cf. Lamberti 1999)

	A. Long non-subject form	B. Short non-subject form	C. Subject form
1sc	<i>yoo/yowa/yoyya/yetta</i>	<i>yi</i>	<i>anu</i>
2sc	<i>kowa/koyya/kotta</i>	<i>ku</i>	<i>atu</i>
3sm	<i>kaa/kayya</i>	<i>kaa</i>	<i>usuk</i>
3sf	<i>teya/teyya</i>	<i>tee</i>	<i>ishi/ishe</i>
...	~ absolutive case	~ absolutive case	~ nominative / genitive case

## 5. The personal pronoun in Berber

The Berber languages have some six pronominal paradigms. They can, however, be reduced to the original three.

Table 4. The personal pronouns in Tashelhit = Shilha

	A. Independent	B. Direct object	C. Complement of prep.	D. Indirect object	E. Possessive (a)	F. Possessive (b)
1sc	<i>nkk(in)</i>	<i>-yyi</i>	<i>-i</i>	<i>-yyi</i>	<i>-Ø<sup>3</sup></i>	<i>(i)nu</i>
2sm	<i>kyy(in)<sup>4</sup></i>	<i>-k</i>	<i>-k</i>	<i>-ak</i>	<i>-k</i>	<i>nn-k</i>
2sf	<i>kmm(in)</i>	<i>-km</i>	<i>-m</i>	<i>-am</i>	<i>-m</i>	<i>nn-m</i>
3sm	<i>ntta(n)<sup>5</sup></i>	<i>-t</i>	<i>-s</i>	<i>-as</i>	<i>-s</i>	<i>nn-s</i>
3sf	<i>nttat<sup>6</sup></i>	<i>-tt</i>	<i>-s</i>	<i>-as</i>	<i>-s</i>	<i>nn-s</i>
1pc	<i>nkkwni</i>	<i>-a(n)y</i>	<i>-ney</i>	<i>-a(n)y</i>	<i>-t-ny</i>	<i>nn-y</i>
2pm	<i>kwnni</i>	<i>-kwn</i>	<i>-wen</i>	<i>-awn</i>	<i>-t-un</i>	<i>nn-un</i>
2pf	<i>kwnnimti</i>	<i>-kwnt</i>	<i>-k<sup>w</sup>ent</i>	<i>-awnt</i>	<i>-t-unt</i>	<i>nn-unt</i>
3pm	<i>n(it)tni</i>	<i>-tn</i>	<i>-sen</i>	<i>-asn</i>	<i>-t-sn</i>	<i>nn-sn</i>
3pf	<i>n(i)tnti</i>	<i>-tnt</i>	<i>-sent</i>	<i>-asnt</i>	<i>-t-snt</i>	<i>nn-snt</i>

Paradigm C (complement of prepositions) clearly represents the Afroasiatic suffix pronoun. The indirect object suffix (D) is based on a (prepositional?) element *(y)a-* (in Tuâreg: *ha-*) plus suffix pronoun. Of the possessive suffixes, there are forms that are immediately attached to nouns (e.g., *baba-s* “his father”); they are very similar to the C forms.

Others display a connective element *-n-* before the suffix pronoun. Apart from minor deviations, the pronominal elements are identical among these four paradigms (CDEF). A few conspicuous features set them apart from paradigms AB (Gensler 2000).

- The base of the third person forms which is *-s-* in CDEF, as compared to the *-t-* of the others.
- The second person feminine has *m* in CDEF, but *km* in AB.
- The third person singular has gender distinction in AB, though not in CDEF.

Sasse (1981: 143ff.) posits only two formally diverging Afroasiatic basic paradigms of pronouns that had originally been free. Their distribution depends on case, paradigm (1) representing the absolutive case, paradigm (2) the subject case (hence, the nominative). As the absolutive forms were frequently cliticized, possessive and object suffixes have come into existence, according to Sasse. Similarly, Blažek (1995: 2)

3. Tamasheq (Tuâreg) *-(h)i*; Taqbaylit (Kabyle) *-(i)w*; Nefusi *-Ø*, but with a different accent pattern (Beguinet 1942: 118); etc.

4. Tamasheq (Tuâreg) *kay*; Taqbaylit (Kabyle) *kečč*; Nefusi *šek*; etc.

5. Zenaga: *ntu*, *netta*, *nenta*, *nentahu*; Tamasheq (Tuâreg) *enta*; Nefûsi *nît*; etc.

6. Zenaga: *ntuhet*, *ntheadatt* *ntadatt*, *nentahadd*; Tamasheq (Tuâreg) *enta*; Nefûsi *niyet*; etc.

assumes that the Chadic pronouns “can be reconstructed in two sets ... The set A represents the independent forms, the set B is reconstructed on the basis of the object and possessive forms.” Similarly, “The original AA system of personal pronouns was represented by the same opposition of the set A = subject case (independent) vs. B = absolutive case (object and possessive).” Note that these latter categories, object and possessive pronouns, have been merged here. The independent pronoun is not so much a subject pronoun as a predicate pronoun, except for Semitic.

However, when taking into account the general evidence of the families (other than Semitic), it does not seem that dependent pronoun (‘object’) and suffix pronoun (‘possessive’) have a common origin (be it in the absolutive pronoun or elsewhere).

An older study (Satzinger 1991: 129f.) concluded that paradigms B and C (dependent and suffix pronoun, respectively), though obviously related, are definitely distinct from each other. A later study would “suggest to regard the dichotomy of the Afroasiatic ‘object pronoun’ ... and ‘possessive pronoun’ ... not as a secondary feature, restricted to some branches, but rather as a basic feature of the original system” (Satzinger 2004: 492).

## 6. Chadic: The personal pronouns in Hausa

The system of Hausa paradigms (Newman 2000: 476–487) has some resemblance with the Berber system.

Table 5. The personal pronoun in Hausa

	Independent	Strong object	Weak object	Indirect object	Free possessive (m/f)	Bound possessive (m/f)
1sc	<i>nī</i> <sup>7</sup>	<i>ni</i>	<i>nì</i>	<i>mini, mìn, manì</i>	<i>nāwa / tāwa</i>	<i>-nā / -tā</i>
2sm	<i>kai</i>	<i>ka</i>	<i>kà</i>	<i>makà, mā, m̃a</i>	<i>nākà / tākà</i>	<i>-nkà / -rkà</i>
2sf	<i>kē</i>	<i>ki</i>	<i>kì</i>	<i>miki</i>	<i>nāki / tāki</i>	<i>-nki / -rki</i>
3sm	<i>shī</i>	<i>shi</i>	<i>shì</i>	<i>masà, mishi, m̃as, m̃ār</i>	<i>nāsà / tāsà, nāshi / tāshi</i>	<i>-nsà / -rsà, -nshi / -rshi</i>
3sf	<i>ita</i>	<i>ta</i>	<i>tà</i>	<i>matà</i>	<i>nātà / tātà</i>	<i>-ntà / -rtà, nātà / tātà</i>
1p	<i>mū</i>	<i>mu</i>	<i>mù</i>	<i>manà</i>	<i>nāmù / tāmù</i>	<i>-nmù / -rmù</i>
2p	<i>kū</i>	<i>ku</i>	<i>kù</i>	<i>mukù</i>	<i>nākù / tākù</i>	<i>-nkù / -rkù</i>
3p	<i>sū</i>	<i>su</i>	<i>sù</i>	<i>musù</i>	<i>nāsù / tāsù</i>	<i>-nsù / -rsù</i>

7. Hausa vowels: low tone is marked by a grave accent, high tone is unmarked; *a* – high, short; *ā* – high, long; *à* – low, short; *ā* – low, long.



I do not deal here with the “heavy subject pronoun” (Newman 2000: 486, § 1.2.3), being the conjugation of the Completive Tense (= past/perfect), like *Mūsā yā kōmā* “Musa returned/has returned,” nor the ‘light subject pronoun’, being a combination of pronoun and TAM marker, used as prefix of the other TAM paradigms, like progressive *i-nā zuwā* “I am coming.” Characteristic forms are *yā* (etc.) for the third singular, masculine, and an additional “impersonal person” for the indefinite subject, *ā* (etc.) (Newman 2000: 486, § 1.2.2).

- There is an independent pronoun: predicate and subject of the nominal sentence, focus, topic; object of prepositions (*sic* – cf. French *pour moi*); object ofthetic negation, pronoun for *babu* “there is not” (from the evidence in Berber and Egyptian, the ‘object pronoun’ would be expected). It can be correlated to the independent pronouns of Berber, Cushitic, and Egyptian.
- Dependent pronoun: direct object of a verb; complement of *akwai* “there is ...”, “... exists”, and of *ga* “here ... is”, “there ... is” (cf. Berber, Egyptian; Satzinger 2005). It can be correlated to the dependent pronouns of Berber, Cushitic, and Egyptian.
- Pronoun for indirect object: composed of *mV*- and dependent pronoun (but 3sm. -*sa* besides -*shī*).
- Free possessive pronoun, composed of gender/number-marked bases *nā*- (m., pl.) and *tā*- (f.) and the dependent pronoun (but 1sc -*wā*, rather than -*nī*; 3sm. -*sa* besides -*shī*): *nāwa* “mine (Masculine and Plural)” (French « le mien », « les miens/miennes »), *tāwa*, “mine (Feminine)” (French « la mienne »).
- Bound possessive pronoun, composed of gender/number markers -*n*- (m., pl.) and -*t*- (f., before vowel), -*r*- (f., before consonant) and dependent pronoun (but 1sc -*ā*, rather than -*nī*; 3sm. -*sa* besides -*shī*): *gida-n-sa* “his house”; *mata-r-sa* “his wife” – the same as the latter, though clitic and syncopated.

In these possessive forms that deviate from the dependent pronoun vestiges of an original suffix pronoun may be seen. Hence, they may be correlated to the suffix pronouns of Berber, Cushitic, and Egyptian.

Table 1  Correlation of pronouns in Egyptian, Berber, and Hausa

Egyptian	Independent	Dependent	Suffix
Berber	Independent	Direct object	1. Complement of preposition 2. Possessive (a)
Hausa	Independent	Object	Possessive (?)

For a general overview of the pronominal paradigms in Chadic languages (comprising “Independent – Object – Possessive”) see Blažek (1995: 5ff.).

## 7. Originally only two paradigms of the personal pronoun?

The number of three paradigms – A (independent), B (dependent), and C (suffix) – can be reduced by one, though not by deriving B (dependent) and C (suffix) from the same origin, but rather by deriving A (the independent pronoun) from the B pronoun. It has been shown that “there is no original set of A pronoun forms” (Satzinger 2004: 492). The independent pronouns are quite diverse through the various families and languages (cf. Sem. Cush. 2sm *\*ʔanta* with Eg. *ṯwt* < *\*kwt* [ki'wat (?)], Tuâreg Hausa etc. *kay*, Ber. *kečč* etc.; cf. Zaborski 1998) and attest to glaring innovations (cf. 2sm Eg. *intk* [ʔan'tak], Beja *barûk* / *batûk*). It is not possible to reconstruct a uniform Proto-Afroasiatic paradigm of the independent pronoun. Generally speaking, the A pronouns are either

- derived from, or identical with, B pronouns: e.g. Eg. *swt* (A), *sw* (B); Sem. *\*šū(a)*; or
- built on a base *\*ʔan*, with stative endings or other pronominal elements: *\*ʔanâku*, *\*ʔanâ* / *\*ʔanî*, *\*ʔanta*, *\*ʔantî*, etc.; or
- built on nouns, with suffix pronouns added: Eg. 2sm *intk* [ʔan'tak] < *\*int* [ʔali'ta] “essence” (or sim.; in absolutive case) plus *-k* (Satzinger 1991: 122ff.; note that the element *int*, probably derived from the preposition (*i*)*n* ~ Sem. *\*li-* (?), is essentially different from the base *\*ʔan* of Semitic and Cushitic, as noted); Beja 2sm *barûk* (m.), 2sf *batûk* (f), hence *\*bat-û-* / *\*bat-t-û-*, plus *-k*.

Berber varieties show a nominative-absolutive syntactic alignment, and display several paradigms of the personal pronoun. How can we correlate them? “The pronominal expression corresponding to the noun in the nominative is, on the one hand, the suffix pronoun (genitive, prepositional), on the other, the conjugation morphemes of the verb (subject). The pronoun expression corresponding to the noun in the absolutive case is, on the one hand, the absolute pronoun (predicate, focus, topic, citation form), on the other, the so-called object pronoun, or dependent pronoun (object, rhematic noun in thetic expressions).” (Satzinger forthcoming.)

This can be generalised to Afroasiatic as a whole, insofar as it has the aforementioned alignment, i.e., with the exception of Semitic. In the nominative-absolutive alignment, the functions of the absolutive case are citation, predicate, focus, topic, furthermore the object of transitive verbs. On the pronominal level, these are the functions of the absolute or independent pronoun (A), except for the object function, which is one of the dependent pronoun (B).

8. Correlation of noun cases and pronoun paradigms

The main function of the nominative is that of a subject of the verb (transitive and intransitive). The subject function of personal pronouns is covered by the conjugation morphemes (prefixed or suffixed) of the verb in most languages. In Egyptian, however, the suffix pronouns serve as subject of the verb in all but the stative (old perfective) conjugations, either directly joining a verbal noun (?), or an auxiliary element, like *-n*, or a verb of saying or thinking, viz. *i* (plus *-n*), *hr*, or *k3* (Satzinger 2003b).

In addition, “some Semitic languages have secondarily (i.e., much later than the suffix and prefix conjugations) developed comparable structures. In Gé‘ez, verbal nouns in the adverbial accusative (as *qatīl-a* “while/when killing”, or the like) may be conjugated by means of the suffix pronoun: *qatīl-ō* ( < *\*qatīl-a-hū*) “when he killed”. A further comparable feature is the circumstantial expressions formed by adjectives that are in concord with their referent: (“you [nominative], or your, or of you [genitive] ... *tekūz-e-ka* being sad”; “you [accusative] ... *tekūz-a-ka* being sad” (Satzinger 1968; cf. Kapeliuk 1998)). In Syriac, the suffix pronouns are employed in the new perfect *qṭīl-leh* which has been compared with the Egyptian *sdm-n=f* form; note, however, two important differences: first, the passive participle is in concord with the object of the construction (only in Neo-Aramaic can this concord be absent); second, the suffix pronoun functions as a copy pronoun for a substantival subject; it is even present if the subject is nominal: *N. qṭīlā-leh* “N. has killed (her)” = Egyptian *sm3-n N.* (this latter argument also applies to the Ethiosemitic constructions mentioned)” (Satzinger 2002: 249f.; cf. Satzinger 2004: 488f.). “Berber: in the Ayt Ziyān dialect, the Kabyle stative conjugation (representative of the Afroasiatic suffix conjugation) is replaced by the stative form of the adjective, with the suffix pronoun as subject expression: Ait Ziyān *zggagg-iyi, zggagg-ik, zggagg-ikm, etc.*, ‘I am red, etc.’, as compared with Great Kabylia *zggwagg-ag, zggwagg-ed, etc.* (Galand 1990: 129; cf. Aikhenvald 1995: 51f.)” (Satzinger 2004: 488). Zaborski (2001) has even argued that this type of verbal construction dates back to common Proto-Afroasiatic.

Table 8. to-Afroasiatic Case and Personal Pronoun

Proto-Afroasiatic		
Noun	Absolutive case	Nominative case
Pronoun	1. Absolute pronoun, 2. dependent pronoun	1. Suffix pronoun, 2. conjugation morpheme

- Subject of verb (Nominative): expressed by conjugation; or by suffix pronoun.
- Object of verb (Absolutive): expressed by the dependent pronoun.
- Nominal predicate (Absolutive): expressed by the independent pronoun.
- Subject of nominal predicate (Absolutive): primarily expressed by the dependent pronoun, which is, however, often supplanted by the independent pronoun.

Of the three alignments, the nominative-accusative is the most widespread, followed by the ergative-absolutive alignment. The nominative-absolutive alignment is extremely rare in comparison: The *World Atlas of Language Structures* (WALS) gives the Figures 46 for nominative-accusative and 32 for ergative-absolutive, to which should be added the 4 of active-inactive; only 6 show nominative-absolutive alignment (“marked nominative”). The nominative-absolutive alignment of Afroasiatic may be as old as Proto-Afroasiatic, or it may have developed from an ergative-absolutive alignment (cf. Satzinger 2001; 2005). Semitic nominative-accusative alignment has obviously developed from the Afroasiatic nominative-absolutive alignment, as it still contains conspicuous remnants of it (Sasse 1984).

## References

- Aikhenvald, Alexandra Yu. 1995. “Split Ergativity in Berber Languages”. *St. Petersburg Journal of African Studies* 4.39–68.
- Banti, Giorgio. 2003. “Grammatical Notes”. Vergari and Vergari 2003:7–36.
- Beguïnot, Francesco. 1942. *Il berbero Nefusi di Fassato*. 2<sup>a</sup> edizione. Roma: Istituto per l’Oriente.
- Blažek, Václav. 1995. “The Microsystem of Personal Pronouns in Chadic, Compared with Afroasiatic”. *Studia Chadica et Hamitosemitica*, ed. by Dymitr Ibriszimow, Rudolf Leger, 1–22. Köln: Köppe.
- Brugnatelli, Vermondo. 1986. “Alternanze accentuali e morfo-sintassi nominale nel berbero orientale”. *Contributi di orientalistica, glottologia e dialettologia*, 61–72. Milano: Cisalpino-Goliardica.
- Comrie, Bernard. 2013. “Alignment of Case Marking of Full Noun Phrases”. *The World Atlas of Language Structures Online*, ed. by Matthew S. Dryer & Martin Haspelmath, chapters 98–99. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at <http://wals.info/chapter/98>
- Dixon, Robert Malcolm Ward. 1994. *Ergativity*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511611896
- Galand, Lionel. 1990. “Du nom au verb d’état. Le témoignage du berbère.” *Proceedings of the Fifth International Hamito-Semitic Congress*, ed. by Hans G. Mukarovsky, I, 123–138. Wien: Veröffentlichung der Institute für Afrikanistik und Ägyptologie der Universität Wien (Beiträge zur Afrikanistik, Band 40).
- Gensler, Orin. 2000. “Proto-Afroasiatic as a ‘Marked Nominative’ Language.” Paper presented at: NACAL 28 (Portland, OR), March 10; WOCAL 3 (Lomé, Togo), August 22; *Linguistisches Kolloquium, Humboldt-Universität Berlin*, December 12.
- Kapeliuk, Olga. 1998. “The Ethio-Semitic Possessive Pronouns as Predicalizers in Historical Perspective.” *Aethiopica* 1.148–163.
- Kossmann, Maarten. 2013. *A Grammatical Sketch of Ghadames Berber (Libya)*. Köln: Köppe.
- Lamberti, Marcello. 1999. “The Pronouns of the 1st and 2nd Person in Cushitic and Their Reconstruction”. *Afroasiatica Terestina. Papers from the 9th Italian Meeting of Afro-Asiatic (Hamito-Semitic) Linguistics*, ed. by Marcello Lamberti and Livia Tonelli, 347–361. Padova: Unipress.
- Lipiński, Edward. 2001. *Semitic Languages: Outline of a Comparative Grammar* (Orientalia Lovaniensia Analecta 80), 2nd edn. Leuven: Peeters.

- Newman, Paul. 2000. *The Hausa Language: An Encyclopedic Reference Grammar*. New Haven: Yale University Press.
- Sasse, Hans-Jürgen. 1981. "Afroasiatisch". *Die Sprachen Afrikas*, ed. by Bernd Heine, Thilo C. Schadberg & Ekkehard Wolff, 129–148. Hamburg: Buske.
- Sasse, Hans-Jürgen. 1984. "Case in Cushitic, Semitic and Berber". *Current Progress in Afro-asiatic Linguistics: Papers of the Third International Hamito-Semitic Congress*, ed. by James Bynon, 111–126. Amsterdam: John Benjamins.
- Satzinger, Helmut. 1968. „Äthiopische Parallelen zum ägyptischen s/dt". *Mitteilungen des Deutschen Archäologischen Instituts* 23:163–166.
- Satzinger, Helmut. 1991. "Structural Analysis of the Egyptian Independent Personal Pronoun". *Proceedings of the Fifth International Hamito-Semitic Congress 1987*, vol. 2 ed. by Hans G. Mukarovsky, 121–135. Wien: Veröffentlichungen der Institute für Afrikanistik und Ägyptologie der Universität Wien.
- Satzinger, Helmut. 2001. "On Ergativity in Egyptian". *New Data and New Methods in Afroasiatic Linguistics. Robert Hetzron in Memoriam*, ed. by Andrzej Zaborski, 173–182. Wiesbaden: Harrassowitz.
- Satzinger, Helmut. 2002. "The Egyptian Connection: Egyptian and the Semitic Languages". *Israel Oriental Studies* 20:227–264.
- Satzinger, Helmut. 2003a. "Statuses and Cases of the Afroasiatic Personal Pronoun". *Egyptian and Semito-Hamitic (Afro-Asiatic) Studies in Memoriam W. Vycichl*, ed. by Gábor Takács, 487–498. Leiden: Brill.
- Satzinger, Helmut. 2003b. "The Egyptian Conjugations within the Afroasiatic Framework." *Egyptology at the Dawn of the Twenty-First Century: Proceedings of the Eighth International Congress of Egyptologists, Cairo, 2000*, vol. 3 ed. by Zahi Hawass & Lyla Pinch Brock, 392–400. Cairo: The American University in Cairo Press.
- Satzinger, Helmut. 2004. "Some Remarks on the Afroasiatic Case System". *Wiener Zeitschrift für die Kunde des Morgenlandes* 94:177–183.
- Satzinger, Helmut. 2005. "On the Assumed Ergativity of the Berber Language(s)". *Proceedings of the 10th Meeting of Hamito-Semitic (Afroasiatic) Linguistics (Florence, 18–20 April 2001)*, ed. by Pelio Fronzaroli & Paolo Marrassini, 381–389. Firenze: Università di Firenze, Dipartimento di Linguistica.
- Satzinger, Helmut. Forthcoming. "Zero Article, Bare Noun, Absolutive Case." In *a Festschrift*.
- Van Putten, Marijn. 2014. *A Grammar of Awjila Berber (Libya)*. Köln: Köppe.
- Vergari, Moreno, and Roberta Vergari. 2003. *A Basic Saho-English-Italian Dictionary, with an English-Saho Index and Grammatical Notes*, by Giorgio Banti and Moreno Vergari. Asmara: Sabur Print Services.
- Wolff, Ekkehard. 1981. "Berber". *Die Sprachen Afrikas*, ed. by Bernd Heine, Thilo C. Schadberg & Ekkehard Wolff, 171–185. Hamburg: Buske.
- World Atlas of Language Structures (WALS)*. Available online at <http://wals.info/chapter/98>
- Zaborski, Andrzej. 1998. "Personal Pronoun Systems and their Origin in Some Languages of Ethiopia". *Languages & Linguistics* 1:65–86.
- Zaborski, Andrzej. 2001. "The Origin of the Suffix 'Conjugations' in Afroasiatic Languages". *Sprawozdania z Posiedzeń Komisji Naukowych* 45:59–61.

# The limits and potentials of cladistics in Semitic

Petr Zemánek

Charles University, Prague

Classificational methods based on cladistics are increasingly used in comparative and historical linguistics, including the classification of the Semitic languages. The main data type used in such studies is lexical (especially Swadesh lists); in comparison, grammatical features have been introduced rather slowly.

This contribution examines the possibilities of using grammatical data for phylogenetic tree construction and visualization with NeighborNet techniques. Three datasets with grammatical data are examined both individually and in combination for the two procedures, i.e., constructing phylogenetic trees and networks visualizing the distances among languages.

The results show great variation in trees constructed on the basis of grammatical data by phylogenetic methods, especially for datasets with less rigorous choice of features, but they provide interesting visualizations when the datasets are used with NeighborNet tools. We have extracted the following signals from the models: there seem to be four regions where the Semitic languages resided, the position of Arabic appears stable within the Northwestern languages, and the positions of Sayhadic and Modern South Arabian require further examination, but they may constitute a separate Peninsular region (without Arabic).

**Keywords:** Semitic classification, grammatical features, cladistics, phylogenetic trees, network techniques

## 1. Introduction

In recent years, the application of cladistic methods in linguistics has become popular among evolution-oriented linguists as well as evolutionary biologists and other researchers from outside linguistics. Methods developed in the natural sciences, e.g., for biological evolution, have been applied to modeling language evolution, and thus the circle that begun in the early stages of the history of comparative linguistics, when the two disciplines shared the concept of the evolutionary tree, has been closed once again.



One may ask how adequate the application of methods developed in biological sciences can be for linguistic classification and how far we can go in considering language a biological unit. Such a metaphor appeared already in the early history of comparative linguistics, but we also know well that such projections have their limits: linguistic data differ from biological ones. On the other hand, one should not overlook at least partial overlap in the methodology of the two disciplines; e.g., for a Semitist familiar with the work of Robert Hetzron or Alice Faber, the concept of shared innovations is one of the accepted methods for language classification, one which is widely used outside Semitic as well. There are also many advantages of using cladistics: among others, one can take advantage of the mathematical apparatus developed for such purposes, allowing consistent treatment of the available data. On the other hand, one can also observe that, for many evolutionists from various natural sciences, linguistic data form a welcome test of general cladistic methods.

Almost any introduction to cladistic analysis (e.g., Kitching et al. 1998) will remind a linguist that many concepts used in cladistics are shared with comparative linguistics and language classification based on genealogy, although the terminology may vary slightly.

One of the first language families studied using cladistic tools was the Indo-European family (e.g., Ringe et al. 2002; Rexová et al. 2003; Gray & Atkinson 2003; Nicholls & Gray 2006; Gaillard-Corvaglia et al. 2007).

Other families followed shortly, one of the first being Bantu (e.g., Holden 2002; Holden & Gray 2006; Rexová et al. 2006), and other studies aiming at general problems of language evolution and classification appeared as well (e.g., Atkinson et al. 2005, 2008; Delmestri & Cristianini 2010).

Many of these studies make use of the Swadesh list (both in its 100-word and 200-word versions), as such data are generally easily available; cf. for Semitic linguistics the classification study of Kitchen et al. 2009, likewise based on the Swadesh list.

A wave of cladistics-oriented studies in linguistics followed the publication of the *World Atlas of Language Structures* (WALS 2005). This monumental work provided plenty of easily accessible data not only from lexical but also from other linguistic domains, and these data have rapidly been exploited for cladistic purposes. One can ask about the motivation of such studies: there are certainly other interests than purely linguistic ones, as the authors of such contributions come from a variety of other fields as well, ranging from natural sciences to statistics to psychology, to mention just some.

Whatever the case, one important change was brought about by WALS 2005: whereas earlier studies relied heavily if not exclusively on lexical data, later studies regularly included data from phonology, morphology, and syntax. It is true that the lexicon is an important part of a linguistic system (or 'organism', if we are to use the analogy when dealing with methods developed in biology), but an evolutionary

picture based purely on lexical data must inevitably be skewed, as would be any representation of, e.g., an animal species based only on the characteristics of, say, its skin. The data from WALS 2005 allowed researchers to gain new insights into the evolution of not only languages as such, but also of other linguistic subsystems: cf. Ben Hamed et al. 2005 for vowels, or Lupyan & Dale 2010 for connections with social structure.

Data from both lexicon and grammar (in a broad sense) has been used in several studies (e.g., for Bantu, Holden 2002); however, one aim of the present study is to examine the utility of purely grammatical features. In this connection, it is also of crucial importance to understand the nature of linguistic data: in the linguistic context, types of evolution other than purely genealogical, i.e., external changes due to language contact, can be much more influential than in the case of biological units. I show that language contact causes many changes in various languages, and its role in the case of Semitic classification is particularly important.

## 2. Methodologies, techniques

### 2.1 Methodologies

Two main types of analysis will be examined here: phylogenetic analysis and analysis of neighborhood networks (NeighborNet) based on distances.

Phylogenetic analysis is at its core similar to the comparative method used for the classification of languages, and it aims at constructing an evolutionary sequence of units and groups of units, usually displayed in the form of a tree. It is usually sensitive to the nature of data and requires one to input data of exclusively phylogenetic type (bearing phylogenetic signal), i.e., such that it really shows a new development in the genealogical sequence.<sup>1</sup> As such, the aims of phylogenetic analysis directly correspond to the aims of the purely genealogical classification of languages. The analysis stems from statistical methods based on so-called Bayesian statistics (cf., e.g., Ghosh et al. 2006 or Jackman 2009), the most commonly applied procedures being Maximum Parsimony and Maximum Likelihood.

The other method tested here, NeighborNet networks (sometimes called also Network Joining), works with sets of (binary) data and is less sensitive to the nature of the data, at least in the sense that data without a clear phylogenetic signal still do

---

1. A possible example of a methodological mistake would be the criterion “existence of indefinite article”, which could be applied to the Semitic languages, but would merge two distinct evolutionary phases: the older one in classical Arabic, and the later formation in Arabic dialects and some Ethiopian languages.



contribute to the overall picture of the relations among languages and the result is not as distorted as in the case of phylogenetic analysis. It is also originally intended for phylogenetic analysis, and it can be understood as simply another projection of a phylogenetic tree in the form of a network, without necessarily reflecting the history of the branch divisions. NeighborNet networks can be used especially when the underlying evolutionary history is not necessarily treelike or when it is for some other reasons difficult to determine a unique tree. Unlike the phylogenetic tree, it also provides a more exact image of the distance between various units (languages). The algorithm takes an arbitrary distance matrix and constructs a phylogenetic tree in a network-like projection. The network can also be understood as an unrooted tree, i.e., no language is chosen as the central (or ancestral) one in the tree. The distinction between genotypic and phenotypic types of information is, however, much less clear than with the tree-like models. More information on the method can be obtained from Huson & Bryant 2006, Levy & Pachter 2011, and Heggarty et al. 2010 (including the application in linguistics).<sup>2</sup>

## 2.2 Data characteristics

Three sets of data are used in the present study, all of them consisting of data taken from grammar. The first set is based on the study of Alice Faber (1997), which belongs to one of the accepted classifications of the Semitic languages. Second is a study by Amikaim Gai (1994), where we can find another set of data, not as ordered as in the case of Faber, but reminding us of several interesting features that connect some Semitic languages across the accepted classificational boundaries (so-called cross isoglosses). The third set has been constructed by the author of this study (referenced here as Zemánek 2017) and contains rather randomly chosen features that link some of the Semitic languages, not necessarily in the same sense as Gai 1994, but reflect structural differences and different evolutionary phases of the Semitic languages; some of these features have already been used in other types of the classification of the Semitic languages (e.g., Kienast 2001: 19 or Diakonoff 1965: 11–12). All the data have been coded as binary.<sup>3</sup> A list of the data extracted from each dataset is given below.

---

2. See also <http://www.languagesandpeoples.com/Eng/SupplInfo/AnttilaNeighborNet.htm> (20.12.2011).

3. Another option is scalar coding, which usually reflects some evolutionary mechanism. It is often pointed out that coding does influence the results. However, in our case, when we not only experiment with different datasets and their combination, but also compare phylogenetic trees with neighborhood networks, we considered it best to use the same data for both procedures, and as the NeighborNet analysis by SplitsTree4 requires binary data, the choice has been done for us.

- Data extracted from Faber 1997: General data valid for Semitic: Suffix conjugation as past tense; Prohibitive marker *\*'al(la)* “don’t”; Pharyngealization as a secondary articulation; Non-geminate prefix conjugation for non-past; Intraparadigmatic generalization of vowels in prefix conjugation; Generalization of *-t-* suffix conjugation verbs; Development of compound negative marker *\*bal*; Generalization of *-k-* in suffix conjugation verbs; Generalization of *\*(')al* as verbal negative. Other features describe intragroup differentiation, such as the change *w>y* in Central Semitic, etc.
- Data extracted from Gai 1994: Assimilation of unvocalized *n* to the following consonant; Doubling of the second radical in the imperfect; Internal passive; Passive stems (exclusivity of N or T stems in certain languages); Case neutralization of the construct state; Inflection of the nominal predicate; Preceding referential pronoun; The nature of the attributive sentence (= dependent clause); Direct subordination of a sentence to a preposition; The case systems (expressed by etymologically related means). Other features, covering only development internal to a particular language (e.g., differences between Classical Arabic and its dialects, or between phases of Hebrew) are not taken into consideration in the present study, as they fall outside the limitations of the chosen procedures.
- Data chosen by Zemánek (2017): Number of laryngeals; Number of “*s*” sibilants; Emphatic ejective articulation; Presence of voiced emphatics; Assimilation of “*n*” at pronouns; *s/h* in 3rd person pronouns; *s/h/* in derived verbs; Coexistence of *-k* and *-t* in suffix conjugation; Existence of two forms of prefix conjugation; Existence of purely adverbial cases; Existence of definite article (prefixed and suffixed); Existence of indefinite article (suffixed *-n*).

The datasets are chosen according to the following criteria: Faber 1997 serves as the starting and reference point – as one possible solution of the classification puzzle that has been produced in the form of a tree, it offers feedback on the methods used. If the methods do not convey an acceptable tree, it is obvious that we need to go on searching for more suitable methods. Gai 1994 is another collection of data, all important for the understanding of the relations among the Semitic languages; however, this dataset is not as compact as the previous one. No result in the form of a tree or graph has so far been produced on the basis of this dataset. Zemánek 2017 is yet another collection, comparable to both Faber 1997 and Gai 1994 in size, and it contains features that connect at least two Semitic languages or represent a form of binary coding of data used otherwise in other classifications. No consistency rules have been applied during the construction of this dataset.

All of the datasets are relatively small. This is somewhat unusual with regard to other studies which work with datasets containing hundreds of features. In our case, only the combination of all three datasets gets closer to the size of datasets used in other studies. This is caused by several factors. First, two of the datasets (Faber

1997 and Gai 1994) contain tens of features, which also determines the size of the last dataset. Second, grammatical features will never offer anything like the large numbers of features available in the lexicon. Third, the similar size of the datasets ensures that no one will prevail over the others when combined in the experiments to be described below.

### 2.3 Software used

For most of the procedures connected with the cladistic analysis as well as with the analysis of NeighborNet networks, there are many computer programs available. After some research, we used Mesquite 2.7 (cf. Maddison & Maddison 2010) for phylogenetic analysis and SplitsTree4 (cf. Huson & Bryant 2006 and <http://www.splitsree.org>) for network projections. Both programs offer a wide variety of possible methods for dataset analysis; the number of modules suitable for constructing trees or networks is very large.

### 2.4 Languages represented in the graphs

The number of Semitic languages is relatively high, especially if we count the Ethiopian languages or if different stages of languages like Aramaic or Arabic are understood as independent languages. However, our aims are somewhat different – we do not want to create a new tree classifying the Semitic languages but to test the methods developed for biology on data from the Semitic languages. Our choice was driven not only by previous classifications but also by our effort both to retain the contrast among various groups and not to overload the resulting picture.

Therefore, we need rather a closed and less numerous set of languages which will cover all the main branches of the Semitic languages. For these reasons, the languages included in our sample are as follows: Akkadian and Eblaite for North-East Semitic; Arabic, Ugaritic, Aramaic and Hebrew for Central Semitic; Sayhadic (Epigraphic South Arabian) and Modern South Arabian for the Yemeni regions; and Ethiopian Semitic represented by Geez, Amharic and Harari, and Tigre together with Tigrinya. It is clear however that Ethiopian Semitic languages are especially underrepresented.

Some of these pairs represent another testing device: it has been shown in a number of studies that the differences between Amharic and Harari, Tigre and Tigrinya or Akkadian and Eblaite are not very significant, and this small distance is represented by minimal or no differences in our data. For judging the performance of various methods, this measure proved to be very useful.<sup>4</sup>

---

4. The software produces large numbers of trees, only a few of which are acceptable. The production of trees is very well controllable by the fact that the program uses a rigorously applied

### 3. Projections of data to the models

#### 3.1 Constructing phylogenetic trees

The first tree was constructed on the Faber 1997 dataset. As indicated above, this set has been chosen as the basic one, as it provides feedback to its desired form. The tree in Faber 1997 constructed by means of comparative linguistic methods can thus serve as a reference point. The tree that exhibited basic structural similarity to the referential one (modeled by comparative linguistic methods) was produced using the Maximum Parsimony method,<sup>5</sup> which then served as the model procedure for the analysis of all the other datasets as well as their combinations. The result is shown in Figure 1.

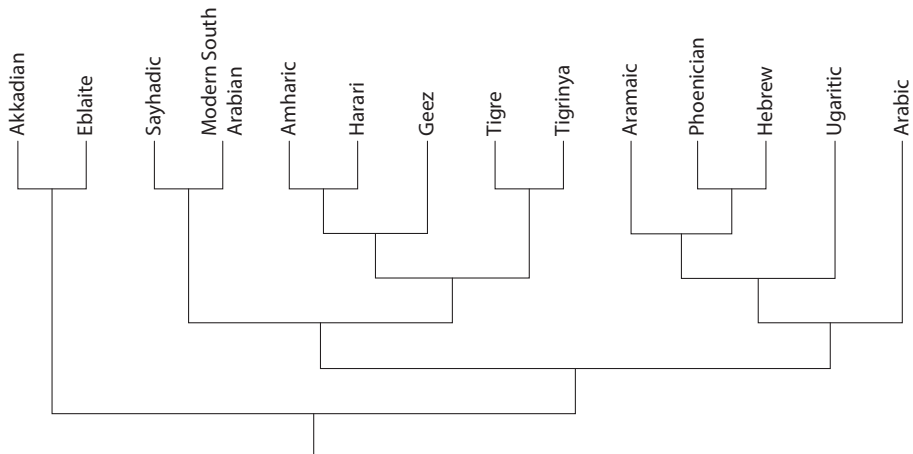


Figure 1. Phylogenetic tree based on the dataset of Faber 1997

The structural properties of both trees, that constructed by means of comparative linguistics methods (cf. Faber 1997) and that based on phylogenetic methods (shown here in Figure 1), are very similar. The basic divisions can be found in both trees, the only differences being the grouping of Sayhadic and Modern South Arabian in one group and connecting Geez as an outgroup with Amharic and Harari (instead of pairing it with Tigre and Tigrinya). For this level of analysis, it

method. For finding the suitable method, the concept of referential tree was adopted. This concept was combined with the concept of “unbreakable” pairs: since pairs of closely related languages mentioned above should be classified together, and trees which separate or “break” these pairs are clearly to be refused.

5. This method is often mentioned as one suitable for linguistic data, and in our case this was confirmed both by the concept of referential tree and by ‘unbreakable’ pairs. One should however keep in mind that there are many other methods used in phylogenetic studies.

can be said that the features chosen by Faber are truly phylogenetic in their nature and perfectly serve their purpose.

The collection of features taken from Gai 1994 (Figure 2) does not aim at constructing a tree, so it would be unfair to expect a tree similar to the one above in Figure 1. From the article it is clear that the author's intention was to show that there are some conflicting signals which should be dealt with in future classifications of Semitic languages. That is why Gai's data can be used as an additional dataset to see how the model will handle them.

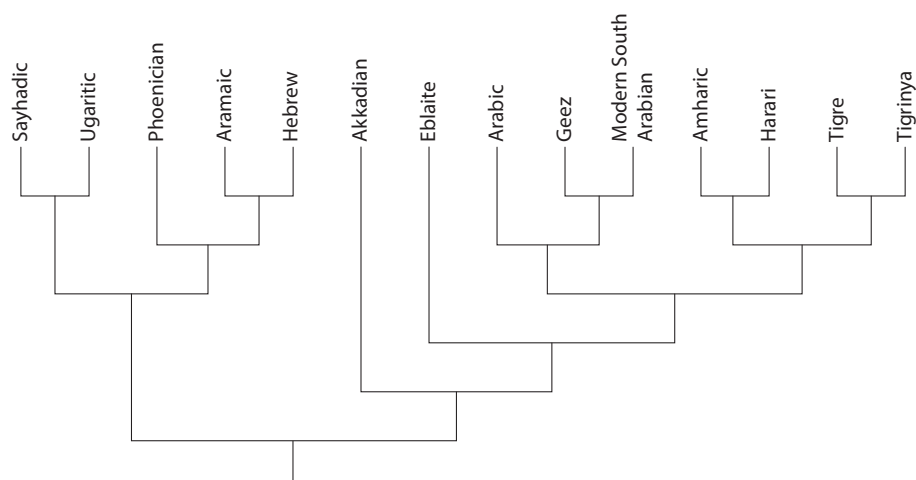


Figure 2. Phylogenetic tree based on the dataset of Gai 1994

Some of the divisions are unacceptable, and probably every Semitist would find pairs or branches for which he or she could spend a long time explaining why such a division cannot work; on the other hand, every Semitist would also agree that there are at least some criteria according to which bringing this or that language into a closer grouping is justifiable. E.g., the right branch, connecting the languages of the Arabian Peninsula and Ethiopia, is reflected in the classification by Moscati et al. 1964: 13–15 although their connection with Eblaite is more than disputable. Generally, the idea of cross isoglosses mentioned in Gai 1994 does find its reflection in this graph (surprising connections); however, the resulting tree as a whole is hardly acceptable, and even one of the 'unbreakable' test pairs (Akkadian and Eblaite) is broken in this tree. The phylogenetic method, especially with samples consisting of few features, is very demanding on the input data and their phylogenetic nature.

The results based on the dataset of Zemánek 2017 (Figure 3) exhibit the same or very similar characteristics as the previous one (Figure 2). No acceptable tree has been produced, and again, as in the case of Figure 2, the right branches agree with Moscati et al. 1964: 13–15, this time being connected with Phoenician (!). Other

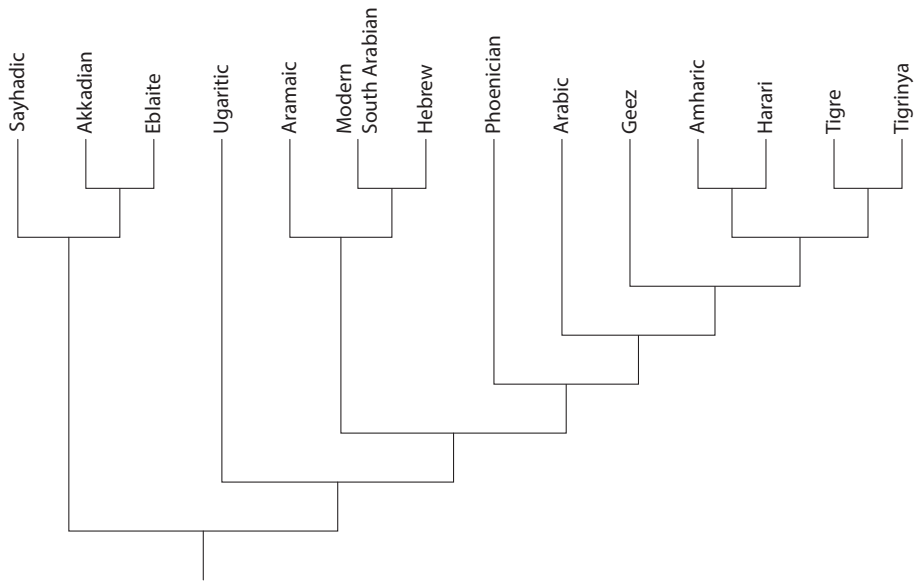


Figure 3. Phylogenetic tree based on the dataset of Zemánek 2017

groups also offer some unexpected combinations. It is clear that the dataset as such does not possess the correct combination of features bearing phylogenetic signals and hence does not work as a whole.

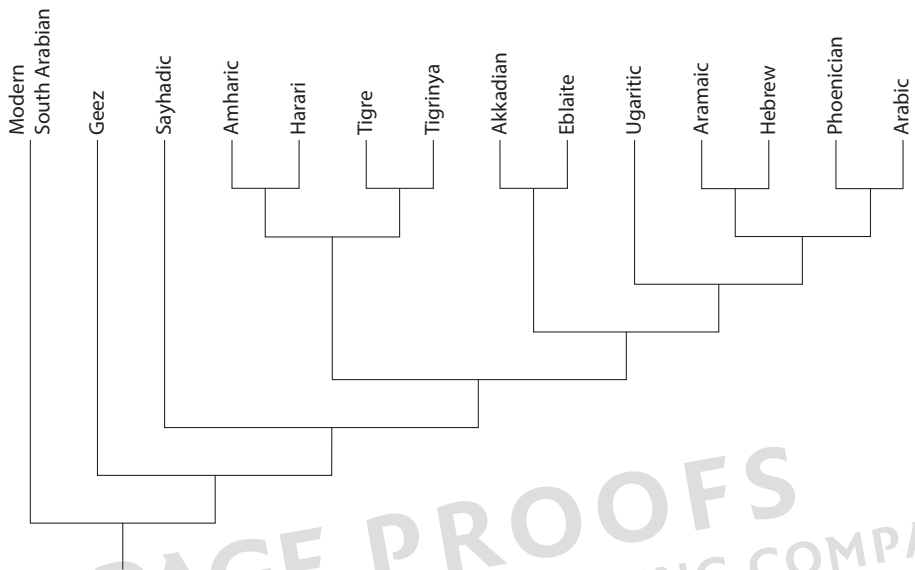


Figure 4. Phylogenetic tree based on the combined datasets of Faber 1997 + Gai 1994 + Zemánek 2017

The tree produced from data combined from the three datasets is reproduced in Figure 4. This tree demonstrates that a collection of rather randomly chosen data will lead to a failure if such a rigorous method is applied. It is different from the three previous datasets, and few new insights can be observed. Interestingly, the model projects Modern South Arabian as the first outgroup (the group that first separated from the rest),<sup>6</sup> followed by Geez as the second outgroup and Sayhadic as the third. The rest is then divided into Ethiopian languages and the remaining group, divided in turn into northwestern languages (with Arabic) and Akkadian and Eblaite (which make up the real first outgroup in other classifications).

Generally, the classification of Semitic languages based on Faber 1997 has remained the same, but the additional data have changed this picture in unforeseen ways, and the cladistic methods did not bring new insights which would complement the Faber model (1997). It is obvious that the phylogenetic method is very demanding on the quality of data collection and on its phylogenetic content. While data collection of Faber 1997 has clearly been sampled by its author with the final image of the tree in mind, other datasets that do not share these characteristics disturb that tree and offer no unambiguous improvements. On the other hand, this does not exclude the possibility of building an acceptable tree based on data chosen with more consistency, although it would most probably necessitate leaving the dataset of Gai 1994 aside, without answering the questions raised by his data.

It is also possible that the method chosen on the basis of Faber's dataset does not suit the new collections of data; this, however, is made less likely by the high quality and consistency of the Faber's dataset, as has been proved above in Figure 1.

### 3.2 The NeighborNet networks

The other method of visualization treats the input data slightly differently, as it does not really distinguish between genotypic and phenotypic data, and also measures the distances among a set of units, in our case languages. The datasets were the same in both instances.

The network constructed on the Faber's dataset (Figure 5) provides a rather schematic picture of the system, where the distance to the core is large in the case of Northeast Semitic and Central Semitic, while the southern branches (divided into two) are rather closer to the center. In this projection, no further differentiation within the individual groups is possible. This set processed with the network

---

6. One cannot resist mentioning the DNA studies that genetically connect the populations of the Afroasiatic peoples from Africa with the populations of Yemen (e.g., Cabrera et al. 2009); however, few linguistic arguments can be adduced in support of this thesis, e.g., pre-Arabian toponymy in Dhofar and Oman of possible Modern South Arabian lineage.

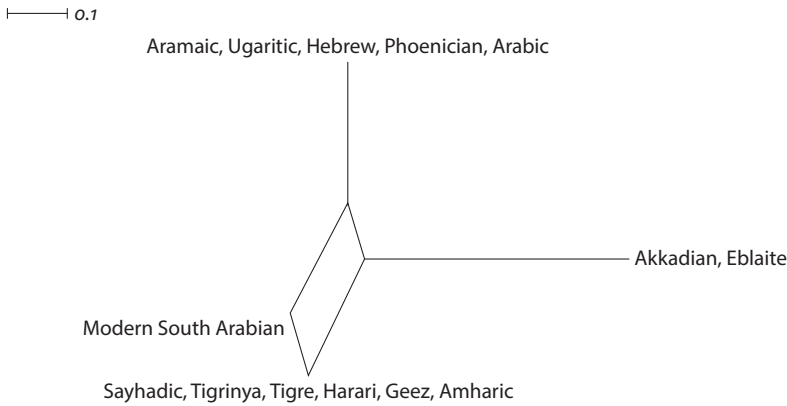


Figure 5. NeighborNet network based on the dataset of Faber 1997

algorithm seems to position the center of the Semitic languages rather in the vicinity of the southern languages, while the northern and western languages seem to be placed rather farther away from the possible center, which probably runs against the intuition of many Semitists. Interestingly, there are four groups of languages, most of them in a reasonable geographical distribution (the only exception being Sayhadic). It is clear that the amount of data provided by Faber 1997 is not sufficient for this type of analysis; on the other hand, the basic picture is clear, and the signal dividing the Semitic languages into four groups is rather strong.

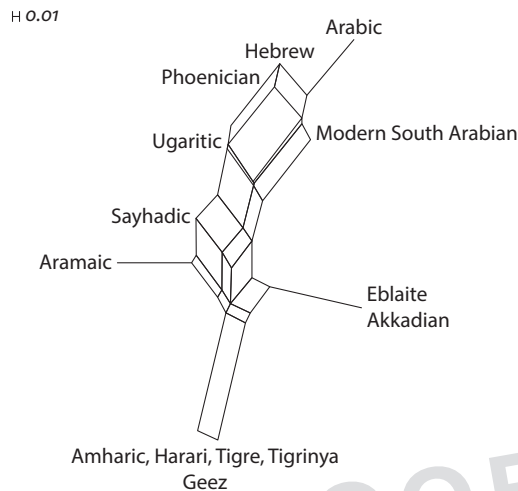
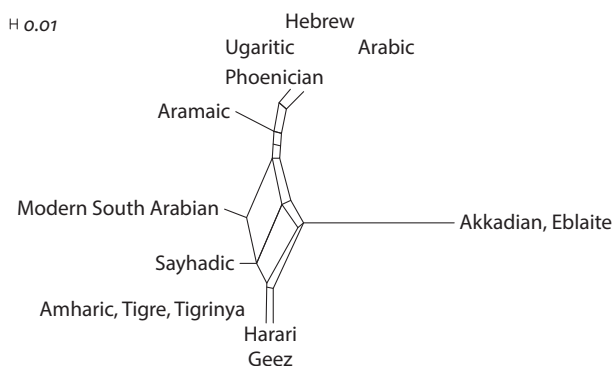


Figure 6. NeighborNet network based on the dataset of Gai 1994



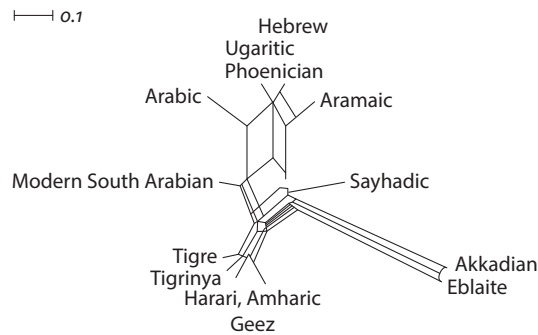
It is interesting that in spite of the nature of the Gai's contribution, the aim of which was to point out isoglosses that cut across the accepted classifications, without necessarily offering a new overall picture of the classification, the network in Figure 6 shows that the basic division into Northeast Semitic, Central Semitic and Southern Semitic is retained even in this collection of data. The position of Sayhadic and Modern South Arabian may not be accepted by everyone, but we should also bear in mind that this dataset did not primarily aim at constructing a consistent classification of the Semitic languages. This model offers three groups, with great variation within the northwestern languages: this means that the data are well formed for these languages but fail to reflect the existing distances, which is visible at Ethiopian Semitic.<sup>7</sup>



**Figure 7.** NeighborNet network based on the combination of the datasets of Faber 1997 + Gai 1994

Figure 7 shows the network built upon the joint data of Faber 1997 and Gai 1994. We can see that the image from Figure 5 (Faber 1997) has similar structural characteristics: four groups of languages. Here we have put Sayhadic and Modern South Arabian into one group, although other distinctions are possible, e.g., a common South Arabian and Ethiopian group or Modern South Arabian in one group and the rest in another. The data of Gai are reflected in better distinctions within the northern group.

7. On the other hand, there are two evolutionary paths directed toward this group, which means that there is some distinction among the Ethiopian Semitic languages, but with current data the projection cannot distinguish among the nodes at the end of these paths.



**Figure 8.** NeighborNet network based on the combination of the datasets of Faber 1997 + Gai 1994 + Zemánek 2017

The network in Figure 8 is a result of uniting all three datasets we have used in this study. There are at least four groups, the only problems having to do with Modern South Arabian and Sayhadic (both seem to be independent within the peninsular region). Some (very subtle) distance between Akkadian and Eblaite is observable. Within Ethiopian Semitic, the distinction between the northern and southern languages is blurred, with Geez positioned in the southern group as in the phylogenetic tree based on the data of Faber (Figure 1). On the other hand, the northern languages differentiate three subgroups, Arabic, Aramaic, and Canaanite (including Ugaritic), which corresponds to the views of many Semitists.

If we adopt an areal point of view, we can identify the Mesopotamian, Ethiopian, and Syro-Palestinian<sup>8</sup> regions, and two subregions in the south of the Arabian Peninsula. This division would also correlate with chronological lines, again with the exception of the two Arabian language groups. Generally, the distances among languages are smaller when compared to the other projections, with the exception of Akkadian and Eblaite.

It is obvious from these results that more data can indeed offer more insights, and also that this method is not as sensitive to the nature of the data as the methods aiming at construction of a rooted phylogenetic tree.

8. Interestingly, Arabic remains in contact with the northwestern languages in all the network projections (Figures 5–8).

#### 4. Discussion

The results of the construction of the phylogenetic trees in Section 3.1 do not offer many new insights, as the main observation is that the tree constructed with the Maximum Parsimony method based on the data of Faber cannot be reasonably extended with additional data from the other two datasets. In other words, more data did not bring new insights or improvements on the original tree; on the contrary, it destroyed the generally accepted groupings identified in Figure 1, such as surprising shifts in the first outgroup or pairing Phoenician with Ethiopian Semitic; introduction of other methods (e.g., Maximum Likelihood) would break the link between the basic model and its improvements and the entire model of Faber would have to be abandoned.

The NeighborNet method, however, offers several points (usually called “signals” in phylogenetic studies) that are worthy of discussion.

First, in all the models, it projected four groups of languages in the regions of Mesopotamia, Syro-Palestine, Ethiopia, and Southern Arabia, and it showed a strong geographical bias in the grouping of the languages.

Second, less sensitivity to the distinction between genotypic and phenotypic features connected with interesting stimuli can be interpreted in favor of models that include areal points of view.

Third, the distances of the Mesopotamian region from the remaining languages are the biggest; then comes the distance between Syro-Palestine and Ethiopia, with the South Arabian region lying in between.

Fourth, the position of Arabic is very stable in all the projections, showing strong links with the Canaanite languages and Aramaic, and rather weak links with the remaining languages of the Arabian Peninsula (not to mention the distance to Akkadian); in other words, the links of Arabic with the North seem stronger than those with the South.

Fifth, the peninsular languages, with the exception of Arabic, show unstable behavior; yet, if an areal point of view is applied, they can be considered as members of one region, although the differences among them are far from negligible.

Although these observations are mere signals that can only serve as supporting arguments, we find parallels for many of them in recent Semitic scholarship. The first and fourth signal agree rather well with the conception of Huehnergard & Rubin (2011), where the areal influences assume an important role and the idea of regions is very similar, with the exception of the Arabian areal, of which only the southern regions are confirmed as relatively independent in our projections (the fifth signal). The third signal can be also found in the old idea of Otto Rössler to divide Semitic into *Altsemitisch* and *Jungsemitisch* (Ancient Semitic

and Neosemitic), the latter being further divided into Frühjungsemitisch and Spätjungsemitisch (Early Neosemitic and Late Neosemitic; cf. Kienast 2001: 19); however, the distance between the latter two divisions is also considerable. The fourth signal, a closer link between Arabic and the Syro-Palestinian regions, can be supported by data especially from Safaitic, where at least some features connect it with the northern regions (cf. Al-Jallad 2015: 10–15). The rather unclear and unstable position of Modern South Arabian in the model (the fifth signal) is also reflected in the recent discussion on these languages (esp. Rubin 2010), although its ‘divorce’ from Arabic can be found in all of these studies. Finally, both Modern South Arabian and Sayhadic seem not to pair very well with both Ethiopian languages and the languages of the Syro-Palestinian region.

## 5. Conclusions

From our data, it seems that the construction of a phylogenetic tree is not very suitable for the Semitic languages, at least as a model based on grammatical data. This is certainly not a definite conclusion, as differently constructed datasets, possibly with another method, may eventually lead to interesting results. However, inclusion of grammatical features together with lexical data is probably inevitable; from the other side, a similar conclusion has been reached in a recent volume on lexicostatistics (Kogan 2015: 602).

The NeighborNet method based on grammatical data seems to detect important signals that go along with current developments in our knowledge of the Semitic languages. This suggests that while purely genealogical trees have problems with creating an accurate picture of the development of the Semitic languages, the networks created by this method are able to make use of greater amounts of data, including data without a clear phylogenetic signal.

For future research, it will certainly be possible to expand the data and to concentrate on other subdomains, such as dialectal variation within some languages (especially Aramaic and Arabic, but possibly also Akkadian and, of course, Ethiopian Semitic). For all of these domains, both methods can lead to interesting insights. Based on our results, the NeighborNet method is preferred for the Semitic languages with data based on grammatical features, but phylogenetic trees are not entirely to be excluded for specific classification tasks.

## References

- Al-Jallad, Ahmad. 2015. *An Outline of the Grammar of the Safaitic Inscriptions*. Leiden: Brill. doi:10.1163/9789004289826
- Atkinson, Quentin, Geoff Nichols, David Welch & Russell Gray. 2005. "From Words to Dates: Water into Wine, Magic or Phylogenetic Inference?" *Transactions of the Philological Society* 103.193–219. doi:10.1111/j.1467-968X.2005.00151.x
- Atkinson, Quentin, Andrew Meade, Chris Venditti, Simon J. Grenhill & Mark Pagel. 2008. "Languages Evolve in Punctuational Burst". *Science* 319.588. doi:10.1126/science.1149683
- Ben Hamed, Mahé, Pierre Darlu & Nathalie Vallée. 2005. "On Cladistic Reconstruction of Linguistic Trees through Vowel Data". *Journal of Quantitative Linguistics* 12.79–109. doi:10.1080/09296170500055467
- Cabrera, Vicente M, Khaled K. Abu-Amero, José M. Larruga & Ana M. González. 2009. "The Arabian Peninsula: Gate for Human Migrations Out of Africa or Cul-de-Sac? A Mitochondrial DNA Phylogeographic Perspective". *The Evolution of Human Populations in Arabia: Paleoenvironments, Prehistory and Genetics*, ed. by Michael D. Petraglia and Jeffrey I. Rose, 79–87. Dordrecht: Springer.
- Delmestri, Antonella & Nello Cristianini. 2010. *Linguistic Phylogenetic Inference by PAM-like Matrices*. University of Trento, Department of Information Engineering and Computer Science. Technical Report #DISI-10-058. Trento 2010.
- Diakonoff, Igor M. 1965. *Semito-Hamitic Languages*. Moscow: Nauka.
- Faber, Alice. 1997. "Genetic Subgrouping of the Semitic Languages". *The Semitic Languages*, ed. by Robert Hetzron, 3–15. London & New York: Routledge.
- Gai, Amikaim. 1994. "Reflections on the Classification of the Semitic Languages". *Orbis* 37.135–144. doi:10.2143/ORB.37.0.2012770
- Gaillard-Corvaglia, Antonella, Jean-Léo Leonard & Pierre Darlu. 2007. "Testing Cladistics on Dialect Networks and Phyla (Gallo-Romance and Southern Italo-Romance)". *Proceedings of Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology*, 23–30. Prague: ACL. doi:10.3115/1626516.1626520
- Ghosh, Jayanta K., Mohan Delampady & Tapas Samanta. 2006. *An Introduction to Bayesian Analysis: Theory and Methods*. New York: Springer. doi:10.1007/978-0-387-35433-0
- Gray, Russell D. & Quentin D. Atkinson. 2003. "Language-Tree Divergence Times Support the Anatolian Theory of Indo-European Origin". *Nature* 426.435–439. doi:10.1038/nature02029
- Heggarty, Paul, Warren Maguire & April McMahon. 2010. "Splits or Waves? Trees or Webs? Network Analysis of Language Divergence?" *Philosophical Transactions of the Royal Society B* 12.3829–3843. doi:10.1098/rstb.2010.0099
- Holden, Clare J. 2002. "Bantu Language Trees Reflect the Spread of Farming across Sub-Saharan Africa: A Maximum Parsimony Analysis". *Proceedings of the Royal Society B* 269.793–799. doi:10.1098/rspb.2002.1955
- Holden, Clare J. & Russell D. Gray. 2006. "Rapid Radiation, Borrowing and Dialect Continua in the Bantu Languages". *The Phylogenetic Methods and the Prehistory of Languages*, ed. by Peter Forster and Collin Renfrew, 19–31. Cambridge: The McDonald Institute for Archaeological Research.
- Huehnergard, John & Aaron D. Rubin. 2011. "Phyla and Waves: Models of Classification of the Semitic Languages". *The Semitic Languages: An International Handbook*, ed. by Stefan Weninger in collaboration with Geoffrey Khan, Michael P. Streck & Janet C. E. Watson, 259–278. Berlin & Boston: Walter de Gruyter. doi:10.1515/9783110251586.259

- Huson, Daniel H. & David Bryant. 2006. "Application of Phylogenetic Networks in Evolutionary Studies". *Molecular Biology and Evolution* 23:2.254–267. doi:10.1093/molbev/msj030
- Jackman, Simon. 2009. *Bayesian Analysis for the Social Sciences*. Hoboken, NJ: Wiley. doi:10.1002/97804706866621
- Kienast, Burkhart. 2001. *Historische semitische Sprachwissenschaft*. Wiesbaden: Harrassowitz.
- Kitchen, Andrew, Christopher Ehret, Shiferaw Asseffa & Connie J. Mulligan. 2009. "Bayesian Phylogenetic Analysis of Semitic Languages Identifies an Early Bronze Age Origin of Semitic in the Near East". *Proceedings of the Royal Society B* 276.2703–2710. doi:10.1098/rspb.2009.0408
- Kitching, Ian J., Peter L. Forey, Christopher J. Humphries & David M. Williams. 1998. *Cladistics: The Theory and Practice of Parsimony Analysis*. Oxford: Oxford University Press.
- Kogan, Leonid. 2015. *Genealogical Classification of Semitic. The Lexical Isoglosses*. Berlin: De Gruyter. doi:10.1515/9781614515494
- Levy, Dan & Lior Pachter. 2011. "The Neighbor-Net Algorithm". *Advances in Applied Mathematics* 47.240–258. doi:10.1016/j.aam.2010.09.002
- Lupyan, Gary & Rick Dale. 2010. "Language Structure Is Partly Determined by Social Structure". *PLoS ONE* 5:1.e8559. doi:10.1371/journal.pone.0008559
- Maddison, W. P. & D. R. Maddison. 2010. *Mesquite: A Modular System for Evolutionary Analysis*. Version 2.73 <http://mesquiteproject.org>
- Moscatti, Sabatino, Anton Spitaler, Edward Ullendorf & Wolfram von Soden. 1964. *An Introduction to the Comparative Grammar of the Semitic Languages: Phonology and Morphology*. Wiesbaden: Harrassowitz.
- Nicholls, Geoff K. & Russell D. Gray. 2006. "Quantifying Uncertainty in a Stochastic Model of Vocabulary Evolution". *The Phylogenetic Methods and the Prehistory of Languages*, ed. by Peter Forster and Collin Renfrew, 161–171. Cambridge: The McDonald Institute for Archaeological Research.
- Rexová, Kateřina, Yvonne Bastin, & Daniel Frynta. 2006. "Cladistic Analysis of Bantu Languages: A New Tree Based on Combined Lexical and Grammatical Data". *Naturwissenschaften* 93.189–194. doi:10.1007/s00114-006-0088-z
- Rexová, Kateřina, Daniel Frynta, and Jan Zrzavý. 2003. "Cladistic Analysis of Languages: Indo-European Classification Based on Lexicostatistical Data". *Cladistics* 19.120–127. doi:10.1111/j.1096-0031.2003.tb00299.x
- Ringe, Don, Tandy Warnow, & Ann Taylor. 2002. "Indo-European and Computational Cladistics". *Transactions of the Philological Society* 100.59–129. doi:10.1111/1467-968X.00091
- Rubin, Aaron D. 2010. *The Mehri Language of Oman*. Leiden: Brill.
- WALS. 2005. Haspelmath, Martin, Matthew S. Dryer, David Gil & Bernard Comrie. 2005. *World Atlas of Language Structures*. Oxford: Oxford University Press.
- Zemánek, Petr. 2017. Complementary Dataset for Semitic Languages Classification [Data set]. Zenodo. doi:10.5281/zenodo.1012680

**PAGE PROOFS**  
© JOHN BENJAMINS PUBLISHING COMPANY

**2nd proofs**

# Lexicostatistical evidence for Ethiosemitic, its subgroups, and borrowing

Grover Hudson

Michigan State University

Wordlists as short as 100 words fail to provide the fullest lexicostatistical evidence for subclassification of a compact and numerous language group such as Ethiopian-Eritrean Semitic (ES). Analysis of a 250-word comparative wordlist of ES languages (Hudson 2013) provides new evidence on the subclassification of the family and the extent of ES borrowing from Agaw and East Cushitic. Prior studies on the subclassification of ES are only partly supported by the 250-word comparisons, where numbers of lexemes unique to subgroups provide new evidence for ES itself and its generally recognized subgroups but no evidence for traditional South ES, Hetzron's 'Outer South', and 'Transversal South' groups. Nor is there evidence for the long-supposed extensive ES borrowing from Agaw.

**Keywords:** wordlists, subclassification, lexicostatistics, Ethiopian Semitic, borrowing

## 1. Subclassification of ES languages

Subclassification increases in difficulty with the number of nodes separating the highest proto-language and languages targeted for subclassification at the bottom of a family tree, because each node represents a language which transmits to its descendants its innovations, borrowings, and tendencies of 'drift' (systematically favored changes), all of which have to be distinguished from innovations of the lower subgroups, which are the critical evidence of their validity (Greenberg 1957: 49). A few grammatical traits argued to be innovations (as by Hetzron 1972) are seldom satisfactory evidence for subclassification, nor is lexicostatistics based on as few as 98 comparisons (Bender 1971). A subclassification supported by more comparisons, including lexical, will be more convincing.

There is much discussion of the merits of shorter or longer wordlists in lexicostatistical research (Heggarty 2010: 314–315). A list of as few as 100 words is



preferred for glottochronology, a controversial type of lexicostatistics the goal of which is dating and for which a constant rate of change is hypothesized for fewer words having better and hypothetically regular rates of retention. But the relatively closely related ES languages (as compared, for example, with Ethiopian Omotic or Cushitic groups) may differ by so small a number of cognates as to be less than the margin of error, so “the greater resolution offered by more data points” (Heggarty 2010: 315) is needed; furthermore, a wordlist with “meanings outside the most stable core should be a positive benefit, for these are by definition the ones most likely to be variable, and thus informative of finer language relationships”.

## 2. A 250-word list as evidence for subclassification

The 250 ‘words’ on which the present research is based are listed below. The goal of the list is words and phrases (phrases such as ‘domestic animal’, ‘carry on the back’, and ‘dry season’) eliciting confident translation equivalents across ES languages.

*able, be ~ (v); all; animal, domestic ~; animal, wild ~; ashes; ask (v); back (of body); be, become (v); bee; big; bird; birth, give ~ (v); bite (v); bitter, be ~ (v); black; blood; bone; boy, child; break (v); breast, teat; breath; bridge; brother; burn (v); bury (v); butter; call (v); carry (v); carry (baby) on back (v); cat; cheek; chest (of body); chicken; chin; cloud; cold; come (v); country; cow; dawn; day; dew; die (v); do (v); dog; donkey; dove; draw water, pour (v); dream; drink (v); dry (v); dry season; dung; ear; earth, soil; eat (v); egg; elbow; elephant; empty; enemy; enough, be ~ (v); enter (v); evening meal; eye; face; fall (v); far; fast (n); fat (n); father; fear (v); fertile; finger; fingernail; fire; fish; fist; flame; flea; flee (v); flour; flower; fly (n); fly (v); foot; forbid, prohibit (v); forehead; forest; forget (v); four; full; give (v); go (v); go out (v); goat; good; grandfather; grandmother; grass; grind (v); hair; half; hand; head; hear (v); heart; heavy, be ~ (v); here; hide, skin; highland; honey; horn; hot; house; how many/much?; hundred; hunger; hunting; husband; hyena; kidney; kill (v); kiss (v); knee; know (v); lame; last year; laugh (v); leaf; left(side); lightning; lion; lip; liver; load (v); lost, be ~ (v); louse; lowland; make, work (v); male; man; many; marrow; measure (v); meat; midday meal; milk (n); money; moon; mother; mountain; mourn, be sad (v); mouse; mouth; mud; name; navel; near; neck; new; night; nose; now; old, grow ~ (v); one; open (v); other, another; palm; plant (v); plow (v); put on (clothing), dress (v); python; rain; rainy season; raw; red; relative, family member; return (v); rib(s), side of body; right(side); river; road, way; root; saliva; salt; sand; say (v); scratch (v); see (v); seed; seek, want (v); seize, hold (v); send (v); set (of sun) (v); shadow; sharp; shave (v); sheep; sister;*

*sky; sleep (v); smallpox; smoke; snake; span (of hand); spear; spring (of water); stand (v); star; stick; stomach; belly; stone; suck (v); sun; sweat (n); swim (v); tail; take (v); tame (v); tear (of eye); tell (v); thirst; this year; three; tie (v); tire (v); today; tomorrow; tongue; tooth; touch (v); town, village; tree; two; urine; watch, guard (v); water; what?; when?; where?; white; who?; wind (n); wing; without; woman; wound (n); year; yesterday*

The list is an expansion of those (typically 100) words at first used in glottochronological research (Hymes 1960: 6) and widely used in lexicostatistical work generally, such as Bender 1971. However, it excludes grammatical words (for example pronouns), which have high retention rates and grammatical meanings and usage specific to their paradigms, and other words problematically eliciting translation equivalents in ES languages, for example *bark* (of a tree), *long*, and *sit*. Such words are replaced here, and the list is expanded by others I have thought almost as basic and found reliably elicited, for example *carry on the back*, *dawn*, and *rainy season*.

Table 1 presents the first four comparisons of the 250-word list across 14 ES languages (Hudson 2013: 35). In the first column are two-letter abbreviations of the language names, from top to bottom Tigre, Tigrinya, Ge'ez, Gafat, Soddo, Mesqan, Muher, Chaha, Inor, Silt'e, Zay, Harari, Argobba, and Amharic. Words and roots of three proto-languages appear in the last three rows: Semitic, Agaw, and East Cushitic.

This top-to-bottom order of the 14 ES languages in Table 1 follows a generally geographical order of north to south from Tigre to Silt'e, then east and north as Zay, Harari, Argobba, and Amharic, keeping together languages generally thought more closely related, including (1) Tigre, Tigrinya, and Ge'ez; (2) Gafat, a group of one; (3) Soddo, Mesqan, Muher, Chaha, and Inor; (4) Silt'e, Zay, and Harari, and (5) Argobba and Amharic (Hetzron 1972: 4–8, 119–122).

Following are aspects of Table 1 relevant for appreciating the nature and validity of the comparisons. For sources of the data see Hudson (2013: 68).

**GAPS IN THE DATA.** Cells having '–' (for example 'animal, domestic ~' in the row of Gafat) are those for which no word was found or, in the case of the three proto-languages, for which no cognate has been reconstructed. ES language gaps must be few to provide reliable comparison of numbers of cognates.

Gaps were only 70 or 2% of the 3500 ( $14 \times 250$ ) comparisons, and Gafat contributed 46 or 18.4% of 250 (in Table 1 *animal, domestic ~*). Gafat went extinct some 65 years ago, and knowledge of it is limited (Leslau 1945, 1956), so, to make its contribution comparable to that of the other languages, Gafat numbers are increased by 18.4%. Gaps for the other 13 languages totaled only 24 or .7% of 3250 ( $13 \times 250$ ) and were ignored in the counts as probably insignificant.

Table 1. First four comparison sets of the 250-word list

	<i>able, be ~ (v)</i>	<i>all</i>	<i>animal, domestic ~</i>	<i>animal, wild ~</i>
Te	kāhalä <sup>1</sup> , šaqmä	kəl~kullu <sup>1</sup>	ʔənsus <sup>1</sup>	ʔaw <sup>1</sup> , wänän
Tn	kāzalä <sup>1</sup>	kʷəllu~kullu <sup>1</sup>	ʔənsəsa <sup>1</sup>	ʔabay <sup>1</sup> , ʔarawit <sup>2</sup> , ʔarä <sup>2</sup>
Ge	kəhlä <sup>1</sup>	kʷəllu <sup>1</sup>	ʔənsəsa <sup>1</sup>	ʔarwe <sup>2</sup>
Ga	färäkä <sup>2</sup>	ɣəl(al)ho~ɣəlhʷä~ɣəl(l)-əm <sup>1</sup>	–	awre <sup>2</sup>
So	čalä <sup>1</sup>	kull-əm <sup>1</sup> , guggər~guggur <sup>2</sup>	ənsəsa <sup>1</sup> , gəzat <sup>2</sup> , kəbt <sup>3</sup>	awre <sup>2</sup>
Mq	xarä <sup>1</sup>	ənnə-m <sup>1</sup>	waga <sup>4</sup>	awre <sup>2</sup>
Mu	xänä <sup>3</sup>	ənnə-m <sup>1</sup> , guggər~guggur~guggʷər <sup>2</sup>	waga <sup>4</sup>	awi <sup>1</sup> , or <sup>2</sup>
Ch	xärä <sup>3</sup>	ənnə(-m) <sup>1</sup>	waga <sup>4</sup>	awi <sup>1</sup>
In	čälä <sup>1</sup> , xärä <sup>3</sup>	ənnə(-m)~ənʔä <sup>1</sup>	waga <sup>4</sup>	awi <sup>1</sup> , or <sup>2</sup>
Si	aqätälä <sup>4</sup>	hullu <sup>1</sup>	gəzät <sup>2</sup> , dinät	bisāwä <sup>3</sup>
Za	aqätälä <sup>4</sup>	hullə(-m) <sup>1</sup>	gəzat <sup>2</sup>	bənensa
Ha	färäka <sup>2</sup>	kullu <sup>1</sup> , jammiʔ	aḡäbäri	ūri <sup>2</sup>
Ar	čä(h)ala <sup>1</sup> , färräha <sup>2</sup>	diyyu(-mm)	gizi <sup>2</sup>	awre <sup>2</sup>
Am	čalä <sup>1</sup>	hullu <sup>1</sup>	ənsəsa <sup>1</sup> , kəbt <sup>3</sup>	awre <sup>2</sup>
Se	*khl <sup>1</sup> , *kwn <sup>3</sup>	*kullV <sup>1</sup>	–	*ʔarway <sup>2</sup> , *bʔs <sup>3</sup>
Ag	–	–	–	–
Ec	–	–	–	*ʔarw~ʔarb- <sup>2</sup>

PHONETIC WRITING follows the source dictionaries but is regularized with consideration to typical practice of ES linguistics: glottalized ejective velar *q* (other glottalized ejectives as *t'*, *s'*, etc.), alveopalatal affricates *č* and *ʃ*, long consonants as doubled consonants (*tt*, etc.), long vowels with a macron (e.g., *ā*), mid central vowel *ä*, and high central vowel *ə*.

JUDGMENT OF COGNATES is shown by the superscript numbers of each column. These are mostly those of the sources, often Leslau 1979 and 1987. Others are the author's, of words related in form and meaning consistent with expected phonetic and semantic change and informed by knowledge of known ES sound changes and geography of the languages.

PROTO-LANGUAGE RECONSTRUCTIONS. The last three rows of Table 1 are reconstructed words and roots of Proto-Semitic, Proto-Agaw including Proto-North Agaw, the branch of four of the five Agaw languages, and Proto-East Cushitic including Proto-Highland East Cushitic, a five-language branch of East Cushitic at the southern ES periphery. The major sources for Semitic are Leslau 1987 and Militarev & Kogan 2000 and 2005, for Agaw Appleyard 2006, and for East Cushitic and Highland East Cushitic Dolgopolsky 1983, Hudson 1989, and Sasse 1979 and 1982. The proto-language words and roots are reconstructions thought cognate with ES words of their column, and may have meanings not of the 250-word list.

VARIANT FORMS are shown by parentheses and ‘~’, in Table 1 as in Chaha *ənnə(-m)* “all” abbreviating *ənnə* and *ənnə-m*, and ‘~’ as in Gafat *yəl(al)ho~yəl-h<sup>w</sup>ä~yəl(l)-əm* “all”.

SYNONYMS seem necessarily accepted in the comparisons, which are data derived from dictionaries, which don’t fully distinguish different meanings or primary and secondary usage. These are seen as multiple words in a cell of the table, for example in the column of “able, be ~ (v)” Argobba *čä(h)alä* and *färräha*. Excluded were synonyms for which the list meaning seemed clearly secondary or the synonyms plainly metaphoric, but the dictionaries don’t often provide a confident basis for such exclusion. There being a Semitic etymology, for example Argobba *čä(h)alä* from Se *\*khl*, is no basis, by itself, for excluding *färräha* in the same meaning. There are after all well-balanced synonyms like English *small* and *little* (Swadesh 1955: 129) differentiated by usage and other words the meanings of which are sometimes ambiguously matched to those of the list, for example “hair”, which can mean hair of the head or of the body, and “hide, skin”, which can mean human or animal skin. It seemed appropriate to keep synonyms where doing otherwise seemed arbitrary.

Of course languages with better dictionaries contribute more synonyms, and the more synonyms a language contributes, the greater opportunity that language has to share cognates. With its lengthy dictionary by Leslau (1987), Ge’ez contributed 152 or 19% of the 798 synonyms. However, each synonym is an opportunity for other languages to contribute a cognate to it, and synonyms having no cognates have no effect on findings of the present research, which sought numbers of shared cognates. The 798 synonyms were 19% of the 4,228 ES comparisons ( $250 \times 14 = 3500 - 70 \text{ gaps} = 3,430 + 798 = 4,228$ ). Inclusion of synonyms seems not to have prevented sensible results as numbers of cognates shared by members of ES groups.

BORROWINGS are a thorny problem. The necessary practice of glottochronology, followed in some lexicostatistical studies, requires that borrowed words be disallowed as not properly cognate. And while borrowing is less expected of basic vocabulary, it is more likely between closely related languages. Leslau 1987 thought Ge’ez *bašor~basor* “meat” to be borrowed from Hebrew and not descended from Semitic *\*bVšar* “flesh” (Militarev & Kogan 2000: 38), in contrast to *bäsär* “meat” of eight South ES languages. In glottochronological counting, if Leslau is right the Ge’ez word shouldn’t be counted as cognate with the South ES words. But such judgments invite doubt, particularly because of CONTAMINATION, by which a word appears to be borrowed but is only reshaped on the model of its cognate in another language. The pronunciation of Ge’ez *bašor~basor* may be only contamination by the Hebrew word, and the South ES words suggest that a cognate native word could have existed in Ge’ez. The particular influence of Amharic by contamination of the lexicon of its ES neighbors must be expected, even in basic vocabulary, and this would often be hard to distinguish from borrowing.

Because of contamination, native words may be wrongly identified as borrowed, but borrowed words may be unrecognized because of NATIVIZATION, by which a borrowed word is made to conform to borrowing-language phonology, as where South ES replaces Ge'ez *s'* with *t'*, perhaps in Amharic *t'ərs* "tooth". So, unwilling to simply assert borrowing without presenting the lengthy and speculative arguments to support the claim, in counting cognates I have made little attempt to identify and exclude any but obvious borrowings.

3. Percentages of shared cognates in a 98-word list

Table 2 presents the percentages of shared cognates in Bender's work (1971: 173), using a 98-word list for all pairs of 14 ES varieties, ordered as in Table 1 but with Welane (We) and Gyeta (Gy) for Table 1's Silt'e and Muher. Welane is perhaps mutually intelligible with Silt'e, and Gyeta is very like Muher.

Shaded cells of Table 2 show the five ES subgroups distinguished above: Tigre-Tigrinya-Ge'ez, Gafat, Soddo-Mesqan-Muher-Chaha-Inor, Silt'e-Zay-Harari, and Amharic-Argobba. As discussed by Bender (1971: 177–179), basic evidence of valid groups is their ingroup percentages greater than percentages for in- and out-group members. In Table 2 the single exception to this is the pair Soddo-Amharic, which at 66% equals the least percentage within the group of Soddo, Mesqan, Gyeta, Chaha, and Inor. Bender (1974: 67) termed Soddo "an intractable case", as "closely related to both Amharic and some of the so-called Gurage languages". In Table 2 other ingroup numbers exceed their group's outgroup numbers usually by good margins, as next discussed.

Table 2. Percentage of shared cognates; 98-word list (Bender 1971)

	Te	Tn	Ge	Ga	So	Mq	Gy	Ch	In	We	Za	Ha	Ar
Am	52	56	62	65	66	56	53	55	50	60	59	61	79
Ar	45	53	56	57	64	57	52	55	53	57	56	56	
Ha	47	48	53	52	61	52	52	54	54	64	70		
Za	49	47	54	53	61	55	55	58	55	79			
We	52	51	61	59	63	59	58	62	55				
In	43	43	46	49	66	70	83	81					
Ch	44	43	51	52	70	80	89						
Gy	44	43	48	51	69	76							
Mq	43	45	52	54	69								
So	47	49	54	62									
Ga	46	49	51										
Ge	71	68											
Tn	64												

#### 4. Rate of error in counting cognates

A ‘good margin’ is greater than the margin of error. Consider three studies as evidence of margin of difference in judgments of lexical cognates, which may be considered evidence of margin of error.

(1) In ES glottochronological research, Fleming (1968) reported for each pair of languages two percentages of their number of shared cognates, with and without words he judged ‘doubtful cognates’. The two percentages differed on average by 4.3%. (2) Results of Cohen (1961: 71, 74) for comparisons of pairs of ten ES languages using a 116-word list, and of Bender (1971: 173) for the same languages using a 98-word list differed in judgment of numbers of cognates by average 8.8%. (3) Bender’s judgments of percent cognates for the same four ES languages in his 1966 (using a 200-word list), 1968 (100-word list), and 1971 research (98-word list) differed by 14% from 1966 to 1968 and 4% from 1968 to 1971. See Table 3 (with four percentages corrected from Hudson 2013: 54). Bender (1968: 2–3) suggested differences of method which might explain generally greater percentages in 1968: less cautious “application of cognate evaluation”, greater “care in choice of items”, and “less arbitrary application of comparative techniques”.

Table 3. Comparison of percent cognates in Bender 1966, 1968, & 1971

	Ge'ez			Tigre			Tigrinya		
	1966	1968	1971	1966	1968	1971	1966	1968	1971
Amharic	55	70	62	43	58	52	59	66	56
Tigrinya	56	72	68	48	64	64			
Tigre	56	70	71						

Bender himself (1974: 8) acknowledged a likelihood of 4–6% of error in Table 2, and Militarev (2000: 268) guessed the probability of errors in evaluating possible cognates in Semitic, Berber, and Egyptian comparisons at 5–10%.

On such evidence we can approximate the rate of error in counting cognates in ES comparisons as 5%. Then not only is the 66% for outgroup pair Soddo-Amharic of Table 2 equal to that for ingroup Soddo-Inor, but within a 5% margin of error, and so are Soddo percentages with Argobba (64), Harari (61), Zay (61), Welane (63), and Gafat (62). Coherence of other groups also involves some percentages between ingroup and outgroup members not greater than a 5% margin of error, especially concerning Welane, now somewhat apart from its Silt'e sibling (Meyer 2006: 16–17), and neighbor to Soddo and Amharic. Borrowing would increase its cognates with its neighbors and decrease those with Silt'e of its close family.

5. Numbers of shared cognates in the 250-word list

Table 4 presents numbers of shared cognates of the 250-word list (with its synonyms) for pairs of 14 ES languages.

Table 4. Number of shared cognates; 250-word list with synonyms

	Te	Tn	Ge	Ga	So	Mq	Mu	Ch	In	Si	Za	Ha	Ar
Am	160	209	198	167	187	145	147	134	130	134	147	130	228
Ar	140	180	170	167	180	142	142	134	130	138	152	139	
Ha	119	126	138	117	143	135	130	124	121	165	169		
Za	122	131	133	128	171	153	150	138	133	193			
Si	111	122	128	116	162	180	164	161	156				
In	107	121	131	130	175	214	238	258					
Ch	110	126	132	130	181	236	259						
Mu	115	132	139	139	205	240							
Mq	112	130	137	140	202								
So	128	153	155	164									
Ga	114	146	146										
Ge	236	259											
Tn	228												

Some considerations favoring the evidence of Table 4 are as follows:

1. Although they employ the different measures of percentage of cognates in a 98-word list vs. the raw number of shared cognates in the 250-word list with synonyms, Tables 2 and 4 give evidence (shaded cells) for the same five sub-groups, but, being based on more data, Table 4 may be thought to provide more reliable evidence, as there is less possibility that the evidence for the five groups is due to chance.
2. “The only generally accepted criterion for subgrouping is shared innovation” (Campbell 1999: 170), and, compared with shorter lists of more basic words, the 250-word list includes words which, having lesser retention rates, are more likely to provide evidence for innovations.
3. Hetzron’s (1972: 36) “Transversal South ES”, which combines Silt’e and/or Welane and Zay plus Harari with Argobba and Amharic, is unsupported in Tables 2 and 4, but more evidently in the larger numbers of Table 4. In Table 2, 15 percentages for members of the group and outsiders are greater than the group’s internal low of 56 (for Zay-Argobba). See these in the rows of Amharic, Argobba, Harari, Zay, and Welane. In Table 4, with Silt’e for Welene, 31 such numbers are greater than the group’s internal low of 130.
4. It was mentioned that the margin of error is perhaps 5%, so that in Table 2 the margin of error is 5 (5% of 98) and in Table 4 approximately 15 (5% of 307;



250 + avg. number of synonyms 57). The latter, larger number is less likely skewed toward wrong interpretation. (The actual number of comparisons for a pair of languages depends on the numbers of synonyms in the comparisons, which varies but not widely for the languages, except Ge'ez.)

Thus in Table 2 it can be seen that, if percentages for Welene with Soddo, Mesqan, Gyeta, and Chaha erred positively by 5% (to 68, 64, 63, and 67), Welene's membership in a group with these would seem as likely as its preferred membership with Zay and Harari. Such skewing over four comparisons is unlikely, of course, but the same skewing in a more critical case, such as Soddo-Gafat, would yield 67 (from 62), greater than for Soddo-Inor, and make less evident Gafat's probably correct membership apart from the others.

## 6. Lexical evidence in the ES family tree

In a family tree of languages, languages of a lower branch are more recently separated, with less time to diverge than are those from languages of higher branches, so their numbers of shared cognates are expected to be greater than their numbers with languages of higher branches. Thus another way to see lexical evidence for subclassification is in the family tree, here that of Hetzron (1977: 17), with percentages from Table 2 added between neighbor languages in Figure 1 and numbers from Table 4 added between neighbor languages in Figure 2.

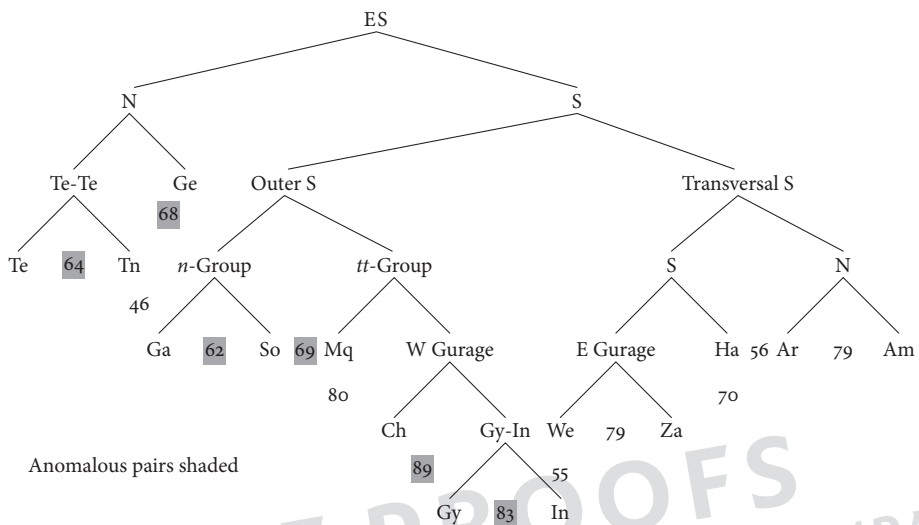
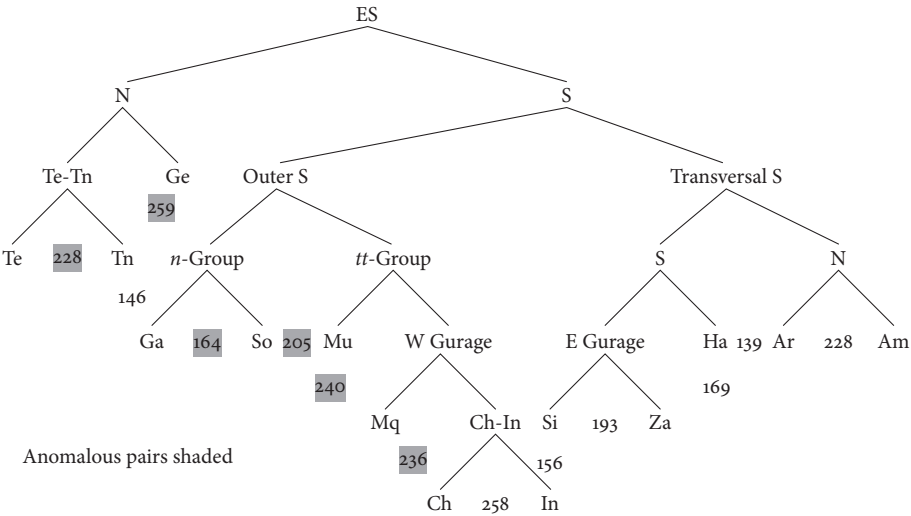


Figure 1. Hetzron's ES tree with Bender's percentages of shared cognates; 98-word list



Shaded in Figure 1 are three pairs of percentages inconsistent with our expectations of the family tree: higher branch 68 of Ge'ez-Tigrinya greater than lower-branch 64 of Tigre-Tigrinya; higher branch 69 of Soddo-Mesqan greater than lower-branch 62 of Gafat-Soddo; and 89 of Chaha-Gyeta greater than 83 of Gyeta-Inor. Also, percentages for Ge'ez-Tigrinya and Tigre-Tigrinya, differing by 4, are less than the 5% margin of error for the 98-word list. Swadesh (1954: 326) thought 81% of shared cognates indicative of dialects, so in Figure 1 Chaha-Gyeta (89%) and Gyeta-Inor (83%) appear to be dialects and expected to minimally differ. In fact, if Chaha and Inor switch places in Figure 1, according to the branching rejected by Hetzron (1972: 72), the third inconsistency disappears: the Mesqan-Inor percentage is 70, Inor-Gyeta 83, and the lowest branch Gyeta-Chaha 89.

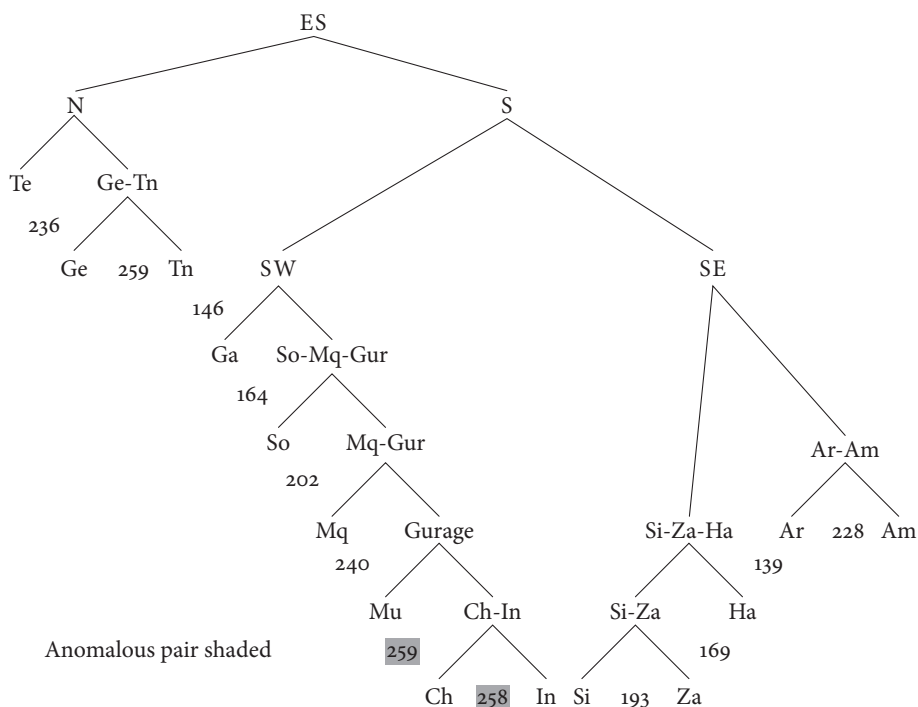


**Figure 2.** Hetzron's ES tree with numbers of shared cognates; 250-word list with synonyms

In Figure 2 again three pairs of numbers (shaded) are inconsistent with relations of the tree: 228 of lower branching Tigre-Tigrinya less than 259 of Ge'ez-Tigrinya; 164 of lower-branching Gafat-Soddo much less than 205 of Soddo-Muher; and 236 of Mesqan-Chaha less than 240 of Muher-Mesqan. The latter pair is well inside the approximate margin of error 15 for the number of comparisons.

Figure 3 is a restructured tree differing from Figures 1 and 2 in three ways to better model lexical relationships which the Table 4 numbers suggest. Ge'ez and Tigrinya are paired against Tigre as in Hetzron (1972: 119). As argued by Hudson (2013: 46), rejecting the criterion of *n*-Group vs. *tt*-Group which links Gafat and Soddo, the Gafat-Soddo pair are separated to become part of a continuum of rising

numbers in the north to south branching. And Muher and Mesqan switch positions as argued by Girma (2001: 75). With changes in the tree come changes in node labels: “Southwest ES” for Hetzron’s ‘Outer South’ and “Southeast ES” for his ‘Transversal South’. Numbers from Table 4 again appear between neighboring languages of the tree, but here only one pair (shaded) is inconsistent with branching, Muher and Chaha (259) and Chaha and Inor (258), and only by one, as is perhaps reasonable for the most recent separations of closely related languages and maybe even dialects. All other pairs differ by more than the approximate margin of error 15.



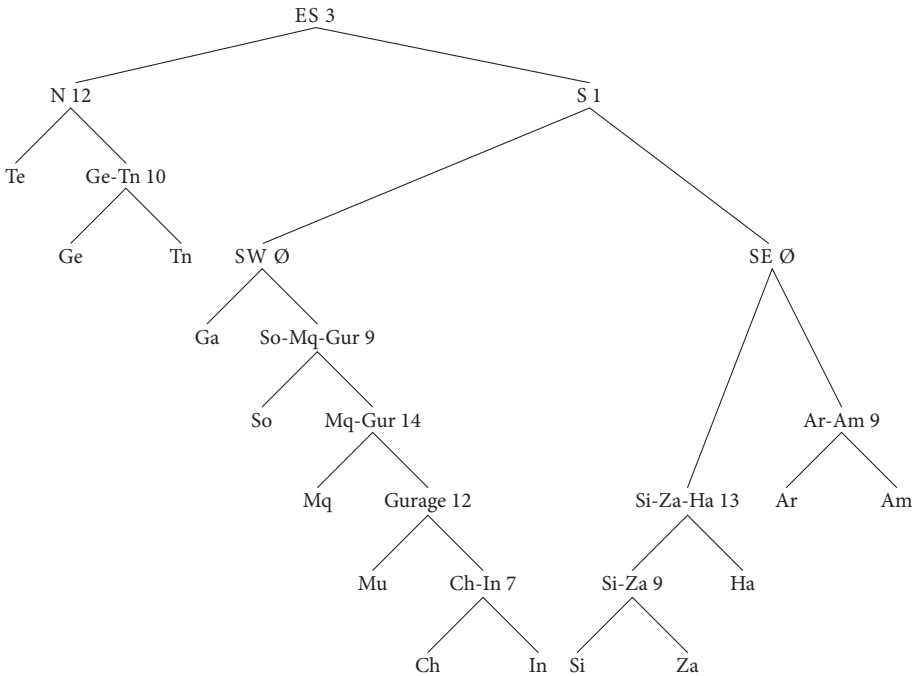
**Figure 3.** Hetzron’s tree revised, with numbers of shared cognates; 250-word list with synonyms

## 7. Number of lexemes unique to ES and its subgroups

The 250-word list provides other evidence for ES subclassification not available with shorter lists: numbers of LEXEMES (sets of cognates) unique to groups, many fewer of which are discoverable using shorter lists of words more resistant to replacement and lacking “the greater resolution offered by more data points” (Heggarty 2010: 315).

In Figure 4 shown at each node of the tree is the number of lexemes unique to each branching but excluding lexemes which have a cognate in any of the three proto-languages. (For the list of all lexemes of the 250-word list, see Hudson 2013: 254–272). As exclusive of lexemes with proto-language cognates, numbers of lexemes in Figure 4 are less likely to be the result of Semitic inheritance or borrowings from Agaw or East Cushitic than to be innovations and as such evidence for validity of the groups. This likelihood is only favored, however, with the possibility of partial retention of innovations and borrowings of higher groups.

However, as arguments for groups it seems significant that only probable groups have numbers of lexemes as great as 7 and improbable groups, including those taken at random, have no more than 6. Thus the Southwest ES group of Figure 3 has no lexemes, and the Gafat-Soddo pair of Figure 2 has only 4. Except for ES itself, South ES, Southwest ES, and Southeast ES, other groups of Figure 4 are both supported in previous research and here by having 7 or more lexemes. The case of the group of ES as a whole is discussed below.



**Figure 4.** Revised tree with numbers of lexemes unique to groups; 250-word list with synonyms

Lexemes unique to South ES (one) and its immediate descendants Southwest and Southeast ES (none) are remarkably minimal. Hudson 2007 and Voigt 2009 argued that the grammatical evidence of Hetzron 1972 for South ES consists of retentions, not innovations valid for subclassification, and the single lexeme unique to South ES is other evidence against this group.

Of course the bigger the group the smaller the likelihood of lexemes unique to the group, because each language of the group is an opportunity to lose a cognate. But while Southwest ES with 6 members and Southeast ES with five members have no unique lexemes, the group of Soddo-Mesqan-Gurage with five members has 9. North ES is well supported by 12 unique lexemes (contrary to the argument against this group by Bulakh & Kogan 2010).

Differing in Figures 2 and 3 is the grouping of Mesqan or Muher with Chaha and Inor. The latter grouping (argued by Girma 2001) of Figure 3 (and Figure 4) has 12 lexemes vs. Mesqan with Chaha of Figure 2, which has only 3. Probably also relevant is that the Figure 3 (and 4) group of Muher, Chaha, and Inor, plus Chaha dialects Ezha, Gura, and Gumer and Inor dialect Gyeta, corresponds well to the historical alliance of *Sebat Bet Gurage* “Seven Houses of Gurage”, to which the term ‘Gurage’ should probably be limited (Hudson 2013: 20–21).

All 14 ES languages uniquely share only three lexemes, which is both interestingly greater (perhaps) than the zeros of Southeast and Southwest ES, and consistent with the unlikelihood of retention of cognates by all 14 languages. In fact, for 13 ES languages there are another four uniquely shared lexemes; for 12 of the languages another five; and for 11 of the 14 another seven. And notice that 11 is the number of South ES languages, for which there is only one uniquely shared lexeme. So ES seems indeed well supported by its number of unique and almost unique lexemes.

Additional evidence for ES is lexical reconstructions of ES, some 50 of Kogan 2005 and 85 of Hudson (2013: 251–254). Figure 5 is the five-branch ES family tree without South, Southwest, and Southeast ES subgroups.

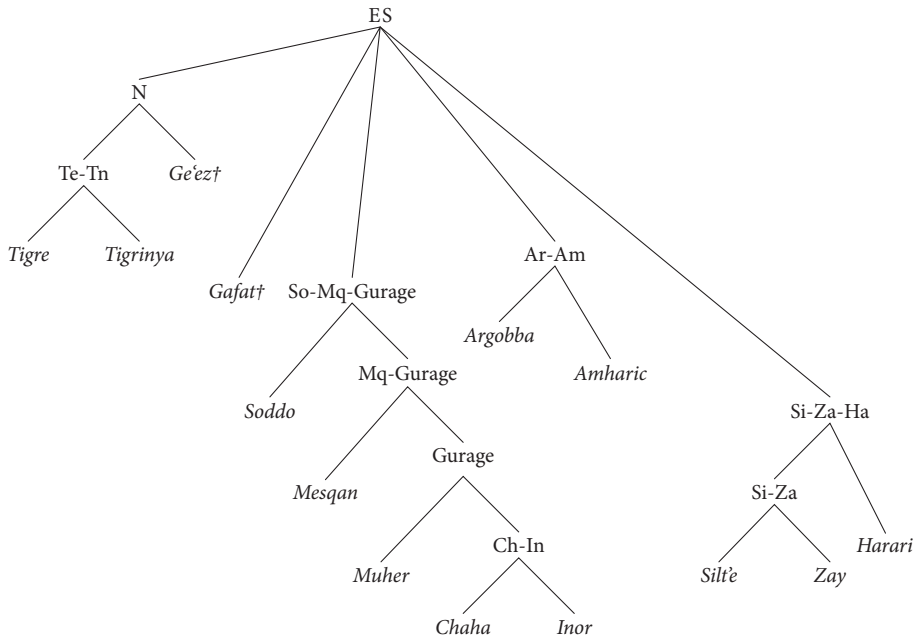


Figure 5. ES family tree of five subgroups on evidence of the 250-word list with synonyms

## 8. ES cognates with proto-languages

The 250 comparisons provided 769 lexemes of two or more cognates. Table 5 presents for the 14 ES languages numbers of these lexemes which include a proto-languages cognate. Let \*Se abbreviate Proto-Semitic, \*(N)Ag Proto-Agaw and Proto-North Agaw, and \*(H)Ec Proto-East Cushitic and Proto-Highland East Cushitic.

Table 5. Numbers of lexemes including a proto-language, by ES language; 250-word list with synonyms

	TE	TN	GE	GA	So	MQ	MU
*Se	182	194	236*	137	152	137	155
*(N)Ag	35	41	45	31	30	25	26
*(H)Ec	49	60	54	54	57	56	59
	CH	IN	SI	ZA	HA	AR	AM
*Se	143	135	130	137	132	152	164
*(N)Ag	23	23	19	18	23	32	36
*(H)Ec	54	57	68	62	64	54	59

\* Twenty-seven of the 236 are the lexemes unique to the pair Ge and \*Se.

The totals of proto-language (reconstructed) words appearing in the 250 comparisons are \*Se 282, \*(H)Ec 124, and \*(N)Ag 62. Even allowing for our better knowledge of the \*Se lexicon, numbers of \*Se cognates are markedly greater than numbers of \*(N)Ag and \*(H)Ec, providing a helpful quantification of the traditional idea that ES has been unusually influenced by borrowing from Cushitic, particularly Agaw.

Bender (1966: 6) found that “the ‘folk-linguistic’ belief that Agaw has profoundly influenced Amharic and differentiated it sharply from the other Ethiopian Semitic languages surely cannot be based on basic vocabulary”, and Bender (1971: 211) found that “the influence is mainly from Ethiosemitic to Agaw”. In Table 5 the three North ES languages (Te, Tn, Ge) share most with their neighbor Agaw, but across Table 5 \*(N)Ag cognates are always fewer than those of \*(H)Ec, and they decrease with increase of the \*(H)Ec numbers and with distance from Agaw territory (Zay being farthest). This seems somewhat inconsistent with the theory that ES is a secondary population in northeast Africa which spread south through Agaw territory, collecting Agaw borrowings even in its basic vocabulary (Ehret 2011: 173–174).

Further reducing the evidence for ES borrowing from Agaw is that of the 62 ES-\*(N)Ag lexemes: 23 also have \*Se cognates, and 22 have \*(H)Ec cognates, so some proportion of these must eventually be attributed to ES inheritance of Afroasiatic roots surviving in \*Se, \*(N)Ag, and \*(H)Ec. In fact, the 11 lexemes having cognates in all three of \*Se, \*(N)Ag, and \*(H)Ec (see the list in Hudson 2013: 296–297) also have cognates in Afroasiatic roots proposed by Orel & Stolbova (1995).

## References

- Appleyard, David. 2006. *A Comparative Dictionary of the Agaw Languages* (Kuschitische Sprachstudien / Cushitic Language Studies 24). Köln: Köppe.
- Bender, M. Lionel. 1966. “Notes on Lexical Correlations in Some Ethiopian Languages”. *Journal of Ethiopian Studies* 4:1.5–16.
- Bender, M. Lionel. 1968. “Remarks on Glottochronology of Northern Ethiopian Semitic Languages”. *Journal of Ethiopian Studies* 6:1.1–11.
- Bender, M. Lionel. 1971. “The Languages of Ethiopia: A New Lexicostatistic Classification and Some Problems of Diffusion”. *Anthropological Linguistics* 13:5.165–288.
- Bender, M. Lionel. 1974. “The Role of Phonological Innovations in the Lexicostatistic Subgroupings of Ethiopian Languages”. *Third Annual Conference on African Linguistics*, ed. by Erhard Voeltz, 39–68. Bloomington, IN: Indiana University Press.
- Bulakh, Maria & Leonid. Kogan. 2010. “The Genealogical Position of Tigre and the Problem of North Ethio-Semitic Unity”. *Zeitschrift der Deutschen Morgenländischen Gesellschaft (ZDMG)* 160.273–302.
- Campbell, Lyle. 1999. *Historical Linguistics, an Introduction*, 2nd edn. Cambridge, MA: MIT Press.

- Cohen, David. 1961. "Le vocabulaire de base sémitique et le classement des dialectes méridionaux (matériaux pour un premier essai de glottochronologie)". *Semitica* 11.53–84.
- Dolgopolsky, Aron B. 1983. "Semitic and Cushitic Sound Correspondences and Cognate Sets". *Ethiopic Studies Dedicated to Wolf Leslau on the Occasion of his Seventy-fifth Birthday*, ed. by Stanislav Segert & András Bodrogligeti, 123–142. Wiesbaden: Harrassowitz.
- Ehret, Christopher. 2011. *History and the Testimony of Language*. Berkeley: University of California Press.
- Fleming, Harold C. 1968. "Ethiopic Language History: Testing Linguistic Hypotheses in an Archaeological and Documentary Context". *Ethnohistory* 15.353–388. doi:10.2307/480417
- Girma A. Demeqe. 2001. "The Ethio-Semitic Languages (Re-examining the Classification)". *Journal of Ethiopian Studies* 34:2.57–93.
- Greenberg, Joseph. 1957. *Essays in Linguistics*. Chicago: University of Chicago Press.
- Heggarty, Paul. 2010. "Beyond Lexicostatistics: How to Get More out of 'Word List' Comparisons". *Diachronica* 27:2.301–324. doi:10.1075/dia.27.2.07heg
- Hetzron, Robert. 1972. *Ethiopian Semitic: Studies in Classification* (Journal of Semitic Studies monograph 2). Manchester: Manchester University Press.
- Hetzron, Robert. 1977. *The Gunnän-Gurage Languages*. Napoli: Istituto Orientale di Napoli.
- Hudson, Grover. 1989. *Highland East Cushitic Dictionary* (Kuschitische Sprachstudien / Cushitic Language Studies 7). Hamburg: Buske.
- Hudson, Grover. 2007. "North and South Ethiopian Semitic". *From Beyond the Mediterranean (Akten des 7 internationalen Semito-äthiopistenkongresses, Berlin, 2004)*, ed. by Rainer Voigt, 339–348. Aachen: Shaker.
- Hudson, Grover. 2013. *Northeast African Semitic: Lexical Comparisons and Analysis*. Wiesbaden: Harrassowitz.
- Hymes, Dell. 1960. "Lexicostatistics So Far". *Current Anthropology* 1.3–44. doi:10.1086/200074
- Kogan, Leonid. 2005. "Common Origin of Ethiopian Semitic: The Lexical Dimension". *Varia Aethiopica in Memory of Sevir B. Chernetsov (Scrinium 1)*, ed. by V. M. Lourié & D. A. Nosnitsin, 367–395. St. Petersburg: Byzantinorossica.
- Leslau, Wolf. 1945. *Gafat Documents: Records of a South-Ethiopic Language*. New Haven: American Oriental Society.
- Leslau, Wolf. 1956. *Étude Descriptive et Comparative du Gafat (Ethiopien Méridional)*. Paris: Librairie C. Klincksieck.
- Leslau, Wolf. 1979. *Etymological Dictionary of Gurage (Ethiopic)*, 3 vols. Wiesbaden: Harrassowitz.
- Leslau, Wolf. 1987. *Comparative Dictionary of Ge'ez (Classical Ethiopic)*. Wiesbaden: Harrassowitz.
- Meyer, Ronny. 2006. *Wolane: Descriptive Grammar of an East Gurage Language*. Cologne: Köppe.
- Militarev, Alexander. 2000. "Towards the Chronology of Afrasian (Afroasiatic) and Its Daughter Families". *Time Depth in Historical Linguistics*, vol. 1 ed. by Colin Renfrew, April McMahon & Larry Trask, 267–307. London: McDonald Institute for Archaeological Research.
- Militarev, Alexander & Leonid Kogan. 2000. *Semitic Etymological Dictionary, vol. 1, Anatomy of Men and Animals*. Münster: Ugarit-Verlag.
- Militarev, Alexander & Leonid Kogan. 2005. *Semitic Etymological Dictionary, vol. 2, Animal Names*. Münster: Ugarit-Verlag.
- Orel, Vladimir E. & Olga V. Stolbova. 1995. *Hamito-Semitic Etymological Dictionary: Materials for a Reconstruction*. Leiden: Brill.
- Sasse, Hans-Jürgen. 1979. "The Consonant Phonemes of Proto-East-Cushitic (PEC): A First Approximation". *Afroasiatic Linguistics* 7:1.

- Sasse, Hans-Jürgen. 1982. *An Etymological Dictionary of Burji* (Kuschitische Sprach-studien / Cushitic Language Studies 1). Hamburg: Buske.
- Swadesh, Morris. 1954. "Perspectives and Problems of Amerindian Linguistics". *Word* 10.306–322. doi:10.1080/00437956.1954.11659530
- Swadesh, Morris. 1955. "Towards Greater Accuracy in Lexicostatistic Dating". *International Journal of American Linguistics* 21.121–137. doi:10.1086/464321
- Voigt, Rainer. 2009. "North vs. South Ethiopian Semitic". *Proceedings of the 16th International Conference of Ethiopian Studies*, ed. by Svein Ege, Harald Aspen, Birhanu Teferra & Shiferaw Bekele, 1375–1386. Trondheim: Norwegian University of Science and Technology (NTNU).



**PAGE PROOFS**  
© JOHN BENJAMINS PUBLISHING COMPANY

**2nd proofs**

PART II

## Forms and functions

**PAGE PROOFS**  
© JOHN BENJAMINS PUBLISHING COMPANY

**2nd proofs**

**PAGE PROOFS**  
© JOHN BENJAMINS PUBLISHING COMPANY

**2nd proofs**

# Reconsidering the ‘perfect’–‘imperfect’ opposition in the Classical Arabic verbal system

Michal Marmorstein

The Hebrew University of Jerusalem

A basic premise that prevails in Western grammatical descriptions is that the Classical Arabic verbal system is based on an asymmetrical opposition between two basic components: the ‘perfect’ *fa’ala* and the ‘imperfect’ *yaf’alu*. The present study re-examines the validity of this premise, in view of several paradigmatic and syntagmatic characteristics of the verbal forms, in particular their syntactic distribution at the level of the clause and at the level of the text, their compatibility with modifying particles and with (verbal or nominal) clause patterns, and their interaction with various lexical classes. The study shows that the opposition between *fa’ala* and *yaf’alu* is restricted to only some environments that, besides these two forms, comprise other verbal forms as well. The (a priori) postulation of an invariable meaning of each form and a fixed opposition between them is replaced by the delineation of various syntactic environments (‘contexts’) that include or preclude an opposition between the forms.

**Keywords:** Classical Arabic, verbal system, tense, aspect, contextual analysis

## 1. Introduction

Western grammatical descriptions are characterized by an ongoing controversy regarding the ‘real essence’ of the verbal forms of Classical Arabic, focusing in particular on whether they are marked for tense or aspect. However, one premise has become consensus in this literature and has acquired the status of a self-evident truth; namely, the Classical Arabic verbal system is based on the opposition between the two simple finite forms *fa’ala* and *yaf’alu*, commonly referred to as the ‘perfect’ and the ‘imperfect’. The present study reconsiders the validity of this premise, with regard to the syntactic distribution of *fa’ala* and *yaf’alu*, and its adequacy in explaining the structure of the verbal system of Classical Arabic.

## 2. A brief account of the ‘perfect’–‘imperfect’ opposition in the literature

Starting with de Sacy (1831: I, 156), Arabists have time and again stated the great difficulty of understanding the Classical Arabic *Tempora* (cf. Reckendorf 1895: I, 52; Brockelmann 1913: II, 144). This difficulty has to do with the extent to which the Arabic system deviates from the models familiar from European languages. Indeed, the relatively small number of verbal forms, covering all temporal and modal distinctions and thus translatable in numerous ways, has posed a real analytical challenge to these scholars. Concerning this matter, they sensed they could not rely on the Arab grammarians, who were obviously not intrigued by the same questions (cf. de Sacy 1831: I, 156).<sup>1</sup> In accordance with the common practice in the study of Semitic languages, Arabists have sought to define a key notion, a primary semantic category that would crack the logic underlying the structure of the verbal system and explain its peculiarity in comparison to the verbal systems of their own languages.

All the descriptions start out from stating the binary nature of the Classical Arabic verbal system, consisting of two main components: the ‘suffix conjugation’ and the ‘prefix conjugation’. The first is manifested in the pattern *fa’ala* (1SG *fa’al-tu*); the latter is manifested in a set of patterns, distinct from each other in the quality of their final short or long vowel, and/or in the presence of a final syllable *-na/-ni/-nna/-nni*, e.g.: 3MSG *yaf’alu* (1SG *’af’alu*), *yaf’ala*, *yaf’al*, *yaf’alanna*; 3MDU *yaf’alāni*, *yaf’alā*, *yaf’alānni*; 3MPL *yaf’alūna*, *yaf’alū*, *yaf’alunna*. The morphological distinction between the two main components is equated with a semantic distinction, either applied to the suffix conjugation and the prefix conjugation, considered as a whole, or restricted to the indicative sphere, where *fa’ala* and *yaf’alu* are considered to form a discrete pair. In any event, this semantic distinction is described in asymmetrical terms, as stemming from “the opposing aspects inherent in the perfect and the imperfect” (Fischer 2002: 102).

Whereas the binary structure of the Arabic verbal system has never been disputed, Arabists have been divided as to the semantic distinction marked by this structure. Since *fa’ala* and *yaf’alu* show only a typical (but not an exclusive) relation to a certain time-layer, scholars have widely rejected their interpretation as absolute tense forms, indicating the ‘past’ and the ‘non-past’ (or ‘present-future’),

1. The Arabic terminology, at least in its original phase (e.g., in Sibawayhi’s *Kitāb*), is not subjugated to the logical view of the time-line trichotomy, but rather reflects a formal-grammatical conception of the verbal system, articulating the fundamental distinction between the forms *fa’ala* – *al-māḍī* ‘the past’ – and *yaf’alu* – *al-muḍārī* ‘the [form] resembling [the agent noun]’.

respectively.<sup>2</sup> Tense, or the expression of the 'external time', if at all recognized, was considered in the main as secondary to the expression of the 'internal time' of the verbal event, i.e., of aspect. Since Ewald (1831: 112), who was the first to apply the opposing terms 'perfectum' and 'imperfectum' to the suffix conjugation and the prefix conjugation, the theory of aspect (even if only later referred to as such, e.g., by Cohen 1924) has become the prevailing one in grammars and specialized treatments of the verbal system of Arabic.<sup>3</sup> The category of aspect, as was generally defined in regard to Arabic (and Semitic in general), refers to the grammaticalized expression of the distinction between a completed (*vollendet*, *accompli*) and an incomplete (*unvollendet*, *inaccompli*) verbal event, signified by the 'perfect'–'imperfect' pair (cf. Fleischer 1885: 95–96; Reckendorf 1895: I, 52–53; Gaudefroy-Demombynes & Blachère 1952: I, 36–37; Cohen 1989). Other minor uses of the term 'aspect', as indicating the view of a situation as 'dynamic' vs. 'static' or as 'constative' vs. 'cursive', were also brought up (Beeston 1970: 76; Reuschel 1996: 24).

However, the theory of aspect did not prove to be the ultimate solution in itself either. The pure notion of verbal aspect, as outlined, e.g., by Fleischer (1885: 95–96), was generally not maintained. In fact, most scholars have admitted that a certain correlation exists between the form's aspect and the temporal and modal meanings conveyed by it. Thus Reckendorf (1895: I, 53–56) has stressed the association between the perfect and the expression of past time, so strong an association as to bring it close to becoming "a real preterit". Being the exponent of complete realization and certitude (*Gewissheit*), Reckendorf argues that the perfect is also the form suitable for the expression of such diverse meanings as the performative, the conditional, the generic, and the narrative time. The imperfect, on the other hand, is entirely dissociated from any relation to a certain time-layer. According to Fleisch (1979: II, 186–188), it is the aspectual property of incompleteness that allows the imperfect to indicate such modal meanings as possibility and aptitude.

Apart from such (ana)logical derivations, scholars have also resorted to the presumed historical situation of Semitic, in order to account for the various uses of the perfect and the imperfect, and especially the correlations between their primary and secondary meanings. Thus Brockelmann (1913: II, 145–146) outlines the common opinion that the present use of the perfect is to be traced back to its function in East Semitic, whereas the indifference of the imperfect to temporal distinctions is to be viewed as originating in a primitive state, where it served as the single, all-purpose, verbal form.

2. For a different analysis of the suffix and prefix conjugations in Arabic, arguing that the opposition is one of a temporal nature, see Aartun 1963.

3. For a detailed account of the development and use of the terms 'perfect' and 'imperfect' and the notion of aspect in grammars of Semitic languages, see Goldenberg (1966: 88–94).

More recent treatments have tried to resolve the multi-functionality of Arabic verbal forms within the frameworks of relevance or markedness. Thus, Götz (1980) finds the distinction between the forms to be one of *Relevanzakzent* that with *fa'ala* lies in the temporal perspective and with *yaf'alu* in the lexical content. According to Bahloul (2008: 140–141), the perfect is the verb form marked for the meanings of anteriority and dimensionalization, whereas the imperfect is its unmarked or neutral pair.

### 3. Methodological problems

However different from each other they may be, all the outlined suggestions share in common the same fundamental view that (a) the Arabic verbal system is a binary one, based on the opposition between the forms known as the 'perfect' and the 'imperfect', and (b) these forms have an invariable meaning (*Grundbedeutung*), present in all of their various uses. Deviations from this basic meaning – whatever its essence is taken to be – are always explained in logical terms as correlating with the basic meaning or as being derived from it. While these explanations cannot be totally discarded, they look more like post-factum rationalizations. As the situation in other 'aspect-languages' (as well as in other varieties of Arabic) appears to be, there is nothing inherent or universal in the category of aspect to entail the temporal and modal nuances that are often ascribed to the verbal forms in Classical Arabic.<sup>4</sup> The recourse to markedness is not satisfying either, inasmuch as it evades the task of identifying a contentful *signifié* to each form, but rather defines the form's positive or privative value relative to a limited selection of semantic features.

The idea that each element in language is linked with an invariable meaning is deeply rooted in linguistic thinking.<sup>5</sup> The pursuit of clear-cut dichotomies is also

---

4. Thus, the typical correlation between the perfect and past time reference in Arabic is not comparable to the situation in Russian, where there exists a perfective non-past with future time reference (see Comrie 1976: 66–71). The fact that in Post-Classical Arabic (cf. Fischer 2002: 103), as well as in Arabic dialects, optative expressions employ as a rule the imperfect rather than the perfect form (as opposed to Classical Arabic) is again evidence for the ad hoc validity of some aspectual-modal correlations suggested for Classical Arabic.

5. García (1991), e.g., strongly advocates the search for invariance, for cognitive as well as methodological reasons; in both, the principle of economy plays a major role and is favored over "ad hoc and arbitrary enumeration of particular facts" (p. 34). One may justly argue, however, that invariant meanings are not naturally given but are also the product of a certain interpretation that may (and even to a greater extent) fall into circularity and arbitrariness. An alternative approach, such as that implemented in this study, aims to identify not the invariant meaning but the contextual features that induce a certain interpretation of the grammatical form.

not new but structured in aesthetics and logic.<sup>6</sup> That these have both a pedagogical advantage and a methodological appeal, in simplifying the data and reducing their analysis to a conclusive bottom-line, is indisputable. However, it is also sure that this kind of thinking bears the risk of flattening complex paradigmatic relations and dependencies into a single axis of oppositions, regardless of the range of functions that the grammatical form fulfills in actual practice. The unsettled dispute over the 'correct' analysis of the Classical Arabic verbal system stems to a large extent from endorsing this kind of reductive, theory-driven approach. The postulation of a single and fixed opposition between the perfect and the imperfect cannot account for the complexity of the verbal system and the nuanced palette of semantic distinctions it serves to convey. Moreover, it ignores both the principle that a semantic opposition is only pertinent in a defined syntactic environment and the fact that the verbal system consists of other verbal forms as well. These two points should be brought to the fore in a detailed and extended analysis of the Classical Arabic verbal system.

#### 4. Reconsidering the *fa'ala*–*yafalu* opposition in Classical Arabic

In the following section, I will examine the contrast between *fa'ala* and *yaf'alu* in relation to a selection of parameters. These parameters consist of various aspects of the syntagmatic and paradigmatic properties of the verbal forms. The examination is based on data extracted from a large and varied corpus of Classical Arabic prose, composed or compiled in the early Islamic period, between the 8th and the 10th centuries CE.

##### 4.1 Syntactic environment

The notion of syntactic environment subsumes both the micro-syntactic structure of the clause and the macro-syntactic context of its use. The latter is more difficult to define, as it consists of diverse aspects of the textual structure and the situation of communication. However, one can start out from considering two significant aspects that transect micro-syntax and macro-syntax and define the syntactic environment; these are termed here REFERENCE and DEPENDENCY.

6. Cf., e.g., Ewald (1891: 2) on the logic underlying the binary structure of the verbal system in Biblical Hebrew: "[N]o language, when it introduces distinctions can start from anything three-fold; antithesis is almost always merely simple and thoroughgoing, because elicited by its [counter] thesis ... . Thus, both in thought and language, every distinction is at first drawn between no more than two things".



REFERENCE is the relation holding between an utterance and the deictic center of discourse. Particular reference is established with respect to the speaking person (1p) or the narrating person (1p/3p); generic reference, by contrast, precludes such an exclusive relation. DEPENDENCY refers to the degree of cohesion that exists between two or more linked clauses. The scale of cohesion, defined in view of several grammatical categories, ranges between the ends of (explicit) embedding and (implicit) coordination.<sup>7</sup> The coordinates of reference and dependency determine to a large extent the interpretation of the (finite) verbal form, being a PERSONAL form that is always ANCHORED IN the clausal or textual grid.<sup>8</sup>

Table 1. Reference and dependency

REFERENCE		
PARTICULAR		GENERIC
1st person (e.g., dialogue)	1st/3rd person (e.g., narrative)	3rd ‘non-personal’ (e.g., axiom)
DEPENDENCY		
embedding←		→coordination

In the following, I present a general mapping of the syntactic environments in which *fa’ala* and *yaf’alu* occur. The main distinction is drawn between environments that present opposition between *fa’ala* and *yaf’alu* and environments that present only a marginal opposition or no opposition at all between the forms. A brief account of other verbal forms that are found in each environment is also given.

Before discussing our main distinction, another preliminary distinction should be made between affirmative and negative clauses. The system of negation presents a different distribution of the verb forms: *fa’ala* is rather limited in negations, giving way to *lam yaf’al* (the negated apocopate) as the common form for past negation. Moreover, *fa’ala* is negated only through the particles *lā* and *mā*, whereas *yaf’alu* – like the participle – can also be negated through *laysa*. The negated form *lā fa’ala*, as opposed to *lā yaf’alu*, is restricted to emphatic expressions, such that the opposition between *fa’ala* and *yaf’alu* is fully at work only with the negation particle *mā*.

7. Dependency or ‘syntactic linkage’ is defined, according to Lehmann (1988), with respect to several semant syntactic continua, all extending from a pole of ‘maximal elaboration’ to a pole of ‘maximal compression’.

8. The generic verb (embodying a ‘non-personal’ person) has, with respect to the reference axis, the privative value of non-referentiality/non-anchoredness.

Requiring an intricate chapter unto itself, the system of verbal negation will not be further discussed in the present study.

Syntactic environments that present an opposition between *fa‘ala* and *yaf‘alu* consist of particular independent clauses in the dialogue, and dependent clauses, in both the dialogue and the narrative. Independent clauses are either unmarked or preceded by operators, such as introductory particles and interrogatives. Some substantial and adverbial clauses allow for the opposition between *fa‘ala* and *yaf‘alu*, as well as adjectival clauses, whether syndetically or asyndetically linked. Table 2 presents the types of clauses that show an opposition between *fa‘ala* and *yaf‘alu*, followed by one example for each type.<sup>9</sup>

Table 2. Syntactic environments presenting an opposition between *fa‘ala* and *yaf‘alu*

ENVIRONMENT	EX.	OPERATOR / LINKAGE DEVICE	EXPONENT <sup>10</sup>	<i>fāa</i>	<i>yafā</i>
A	1	#	–	+	+
1p reference independent cl.	2	introductory part.	<i>‘inna, lākinna, la‘alla, layta, ‘ammā [... fā-]</i>	+	+
	3	interrogative part./pron.	<i>‘a, hal, mā, man, ‘ayy, kayfa, kam, ‘ayna, matā</i>	+	+
B	4	conj. (subs.cl.)	<i>‘anna</i>	+	+
1p/3p reference dependent cl.	5	conj. (adv.cl.)	<i>ḥīna, ḥaytu, ‘id</i>	+	+
	6	rel.pron. (adj.cl.)	<i>llaḏī, mā, man</i>	+	+
	7	asyndesis(adj.cl.)	∅	+	+

- (1) *wayla-kum qataltum ‘abā-hu bi-l-‘amsi*  
 woe\_onto-you kill.SC.2MPL father-his in-ART-day\_before  
*wa-‘aqtulu-hū l-yawma* (Sīra 222)  
 CONN-1SG.kill.PC-him ART-day  
 “Woe unto you! You killed his father yesterday, and am I to kill him today?!”

9. The examples are glossed according to the Leipzig Glossing Rules, with the following additions: CONN=connective, CONS=construct state, EMPH=emphasis particle, INTRO=introductory particle, JUSS=jussive form, MOD=modifying particle, PC=prefix conjugation (*the yaf‘alu* pattern), SC=suffix conjugation (*the fa‘ala* pattern).

10. The exponents presented in Table 2 and in Table 4 below do not exhaust all possibilities; complex conjunctions (prep.+‘an/‘anna/mā) are not included in the present discussion.

- (2) *'allāhumma 'inna 'Ibrāhīma 'abda-ka wa-xalīla-ka*  
 O\_God INTRO 'Ibrāhīm servant-yours CONN-friend-yours  
*wa-nabiyya-ka da'ā-ka li-'ahli Makkata*  
 CONN-prophet-yours call.sc.3MSG-you to-people.CONS Makka  
*wa-'innī Muḥammadun 'abdu-ka wa-nabiyyu-ka*  
 CONN-INTRO.me Muḥammad servant-yours CONN-prophet-yours  
*'ad'ū-ka li-'ahli l-Madīnati* (Mağāzī 22)<sup>11</sup>  
 1SG.call.PC-you to-people.CONS al-Madīna  
 “O God, 'Ibrāhīm, your servant, your friend and your prophet, prayed to you  
 on behalf of the people of Mecca, and [here] I am Muḥammad, your servant  
 and your prophet, pray to you on behalf of the people of Medina”
- (3a) *mā ra'ayta* (Mağāzī 62)  
 Q see.SC.2MSG  
 “What have you seen?”
- (3b) *fa-mādā tarā yā rasūla llāhi* (Sīra 295)  
 CONN-Q 2MSG.see.PC voc Messenger.CONS God  
 “What do you think (lit. ‘see’), O Messenger of God?”
- (4a) *fa-'arafū 'anna šāhiba l-rāḥilati*  
 CONN-know.SC.3MPL COMP owner.CONS ART-riding\_camel  
*qatala-hū* ('Ağānī 2, 178)  
 kill.sc.3MSG-him  
 “And they knew that the owner of the riding camel had killed him”
- (4b) *wa-la-wadidtu 'annī 'uqtalu fī sabīli*  
 CONN-EMPH-love.SC.1SG COMP.me 1SG.kill.PC.PASS in cause.CONS  
*llāhi* (Ṣaḥīḥ 17)  
 God  
 “And I would have loved to be killed for the cause of God”
- (5a) *fa-qultu li-'aṣḥāb-ī hīna*  
 CONN-say.SC.1SG to-friends-my when.CONS  
*'uxriḡnā* (Ṣaḥīḥ 8)  
 take\_out.SC.PASS.1PL  
 “And I told my companions when we were turned out [of the court]”

11. The editor marks the first series of attributes, i.e., *'abd, xalīl, nabiyy*, with the nominative case, apparently in correspondence to the latter *Muḥammadun*. In the interpretation presented here, the accusative is preferred, since it makes little sense to analyze these attributes as predicates. The case vocalization of the second part should be understood as an indication of the presentative – rather than declarative – reading of the *'inna*-clause.

- (5b) *laytanī ’akūnu ḥayyan hīna yuxriḡu-ka*  
 if\_only.me 1SG.be.PC alive when 3MSG.take\_out.PC-you  
*qawmu-ka* (Ta’rīx 3, 1148)  
 people-your  
 “I wish I could be alive when your people drive you out”
- (6a) *wa-’in ’anā ra’aytu min-hu hādihi l-’alāmāti llati*  
 CONN-if I see.SC.1SG from-him DEM.FSG ART-signs REL.FSG  
*ḍakarta* (KalDim 95)  
 mention.SC.2MSG  
 “And if I notice from his part these signs that you have mentioned”
- (6b) *wa-hum muntadūna bi-ḥafiri l-makāni lladi*  
 CONN-they gather.PTCP.MPL in-dug.CONS ART-place REL.MSG  
*yadkuru-hū* ‘Adiyyu bnu Zaydin fī  
 3MSG.mention.PC-it ‘Adiyy son.CONS Zayd in  
*ši’ri-hī* (‘Aḡānī 2, 177–8)  
 poetry-his  
 “While they were gathered in the dug of the place which ‘Adiyy b. Zayd mentions in his poetry”
- (7a) *fa-baynamā hum ’alā ’amri-him ’utiya Hirqalu*  
 CONN-while they on thing-theirs come.SC.PASS.3MSG Heraclius  
*bi-raḡulin ’arsala bi-hī maliku Ġassāna* (Ṣaḥīḥ 9)  
 with-man send.SC.3MSG with-him king.CONS Ġassān  
 “And while they were discussing it, a messenger [who] the king of Ġassān had sent was brought before Heraclius”
- (7b) *’amā waḡada ’Utbatu ’aḥadan yursilu-hū*  
 Q.NEG find.SC.3MSG ‘Utba anyone 3MSG.send.PC-him  
*ḡayra-ka* (Maḡāzī 66)  
 other\_than-you  
 “Didn’t ‘Utba find anyone [that] he could send other than you?”

Independent clauses are anchored in the here-and-now of the dialogue situation, as determined by the first person and further projected on the second and third persons. The first person is the deictic center of discourse, the origo which defines spatial, temporal and modal relationships. The semantic opposition marked by *fa’ala* and *yaf’alu* in these clauses is a complex one: *fa’ala* depicts events that have been realized prior to the dialogue situation, and which may persist through the dialogue time (see (23)); *yaf’alu* depicts either a situation concurrent with that of the dialogue or an intended, desired, or expected event. The opposition between *fa’ala* and *yaf’alu* is not binary: the paradigm of the verbal forms that are found

in independent clauses consists also of the ‘modified perfect’ *qad fa’ala*, the ‘future’ *sa-yaf’alu*, and the participle, as well as the imperative, the jussive, and the energetic.

Dependent clauses are anchored in their matrix clause and thus refer only indirectly to the dialogue (or narrative) situation. Here, again, the semantic opposition between *fa’ala* and *yaf’alu* is not only one of relative tense (i.e., anteriority vs. simultaneity/posteriority), as the temporal meaning conveyed by *yaf’alu* is often fraught with the modal nuances of potential and possibility. The paradigm of the verbal forms in dependent clauses consists of *qad fa’ala*, *sa-yaf’alu* and the participle. It precludes the modal forms, i.e., the imperative, the jussive, and the energetic, as summarized in Table 3.

Table 3. The paradigms of simple and modified forms in environments A and B

A (INDEPENDENT)		B (DEPENDENT)
SEMANTIC OPPOSITION	VERBAL FORM	SEMANTIC OPPOSITION
temporal-aspectual, epistemic and deontic modality	<i>fa’ala</i> <i>yaf’alu</i> <i>fā’ilun</i> <i>qad fa’ala</i> <i>sa-yaf’alu</i>	temporal-aspectual, epistemic and dynamic modality
deontic modality	<i>if’al / li-yaf’al /</i> <i>(la)-yaf’alanna</i>	

There is one more clause type in the dialogue that should be considered in this section, namely, the optative clause. In most cases, optative expressions exhibit the structure *fa’ala* (+pron.) *llāhu/rabbu* (+pron.), e.g., *la’anahū llāh* (*Mağāzī* 31) ‘May God curse him’, *ğazāhu llāhu xayran* (*Sīra* 229) ‘May God bless him’, and *raḥimaka rabbuka* (*Ta’rīx* 1, 92) ‘May God have mercy upon you’. The form *yaf’alu* is only rarely found in optative clauses, alternating mostly with the last quoted example, e.g., *yarḥamuka rabbuka* (*Ta’rīx* 1, 156) ‘May God have mercy upon you’. In fact, it is quite hard to tell the difference between the optative use of *yaf’alu* and other cases where it bears the meanings of wish and hope.

Syntactic environments that present only a marginal opposition or no opposition at all between *fa’ala* and *yaf’alu* consist of verbal complexes, dependent clauses, and mutually-dependent constructions. These are presented in Table 4 and illustrated in Examples (8a)–(21).

**Table 4.** Syntactic environments presenting marginal or no opposition between *fa‘ala* and *yaf‘alu*

ENVIRONMENT	EX.	VERB / LINKAGE DEVICE	EXPONENT	<i>fa‘ala</i>	<i>yaf‘alu</i>
C verbal complex	8	auxiliary v.	<i>kāna</i>	(+)	+
	9	function v.	<i>ǧa‘ala, mā zāla,</i> <i>kāda,</i> <i>šāra,</i> <i>’ašbaḥa,</i> <i>’axaḍa</i>	– <sup>12</sup>	+
	10	perception/permission v.	<i>ra’ā, naẓara,</i> <i>sami‘a,</i> <i>ẓanna,</i> <i>waǧada,</i> <i>taraka</i>	(+)	+
	11	motion/caused motion/setting v.	<i>xaraǧa,</i> <i>’aqbala,</i> <i>inṭalaqa,</i> <i>ǧā’a, marra,</i> <i>’arsala,</i> <i>ba’ata,</i> <i>qāma,</i> <i>makaṭa, bāta</i>	–	+
	12	speech v.	<i>qāla, ḍakara,</i> <i>kataba</i>	–	+
	13	action v.	<i>’akala,</i> <i>iǧtama’a,</i> <i>labisa</i>	–	+
	14	conj. (subs.cl.)	<i>’an</i>	+	–
	15	conj. (adv.cl.)	<i>mā (l-daymūma),</i> <i>ḥattā</i>	+	–
	16	conj. (adv.cl.~comment)	<i>wa-</i>	–	+
	17	conj. conditional	<i>’in, law, mā,</i> <i>man,</i> <i>kullamā</i>	+	–
E mutually-dependent cls.	18	conj. temporal~ conditional	<i>’idā, lammā</i>	+	–
	19	conj. presentative (v.)	<i>’id</i>	+	–
	20	conj. presentative (n.)	<i>’idā</i>	–	+
	21	conj. setting	<i>baynamā</i>	–	+

12. Examples of function verbs with *fa‘ala*, such as *’ašbaḥat ‘aḍalatnī* quoted by Reckendorf (1921: 297), are not attested in prose, but all stem from poetry.

- (8a) *wa-kāna* *Ġabalun xaraġa* *laylan min mawđi‘in*  
 CONN-be.SC.3MSG Ġabal go\_out.SC.3MSG at\_night from place  
*kāna fī-hi* (Buxalā’ 65)  
 be.SC.3MSG in-it  
 “Ġabal had gone out at night from the place where he was”
- (8b) *fa-kuntu* *’axruġu* *kulla ġadātin* (Sīra 315)  
 CONN-be.SC.1SG 1SG.go\_out.PC every morning  
 “And I used to go out every morning”
- (9) *wa-ġa‘ala* *rasūlu* *llāhi yaşuffu*  
 CONN-begin.SC.3MSG Messenger.CONNS God 3MSG.line\_up.PC  
*’aşhāba-hū* (Maġāzī 219)  
 companion-his  
 “And the Messenger of God began to set up his companions in rows”
- (10a) *wa-’innī* *ra’aytu* *l-yawma şayyādayni* *’atayā*  
 CONN-INTRO.me see.SC.1SG ART-day two\_fishermen come.SC.3MDU  
*hādā l-mawđi‘a* (KalDim 84)  
 DET.MSG the-place  
 “And I saw today [that] two fishermen came to this place”
- (10b) *fa-ra’aytu* *’insānan* *ya’ti-hā* *min ġawfi*  
 CONN-see.SC.1SG man 3MSG.come.PC-to\_her at middle.CONNS  
*l-layli* (Sīra 335)  
 ART-night  
 “And I saw a man coming to her in the middle of the night”
- (11) *fa-xaraġnā* *nas’alu* *’an rasūli* *llāhi* (Sīra 294)  
 CONN-go\_out.SC.1PL 1PL.ask.PC about Messenger.CONNS God  
 “So we went to ask about the Messenger of God”
- (12) *qāla* *l-’A‘šā* *yamdaḥu* *l-Samaw’ala* (’Aġānī 2, 27)  
 say.SC.3MSG al-’A‘šā 3MSG.praise.PC al-Samaw’al  
 “Al-’A‘šā said while praising al-Samaw’al”
- (13) *fa-’akalnā* *min-hu nabtaġi* *bi-ḍālika* *l-barakata* (Sīra 338)  
 CONN-eat.SC.1PL of-it 1PL.seek.PC in-DEM.MSG the-blessing  
 “And we ate from it, looking [to gain] a blessing by that”
- (14) *fa-lam* *yalbaṭ* *Qaysun* *ba‘da ḍālika* *’an*  
 CONN-NEG 3MSG.abide.PC Qays after DEM.MSG COMP  
*māta* (’Aġānī 2, 25)  
 die.SC.3MSG  
 “And after this it did not take long before Qays died”

- (15) *'imḏū 'alā smi llāhi fa-la-kumu l-naṣru*  
 go\_away.IMP.MPL on name.CONS God CONN-to-you the-victory  
*mā ṣabartum* (Mağāzī 214)  
 as\_long\_as be\_patient.SC.2MPL  
 "Go, by the name of God, and victory will be yours as long as you are patient"
- (16) *'ataytu 'Ā'īša wa-hiya tuṣalli* (Ṣaḥīḥ 33)  
 come.SC.1SG 'Ā'īša CONN-she 3FSG.pray.PC  
 "I came to 'Ā'īša as she was praying"
- (17) *'innā 'in fararnā l-'āna ṭalaba-nā*  
 INTRO.us if run\_away.1SC.PL ART-time look\_for.SC.3MSG-us  
*qawmu-hū* ('Ağānī 2, 23)  
 people-his  
 "If we run away now – his people will look for us"
- (18) *kāna rasūlu llāhi 'idā 'amara-hum*  
 be.SC.3MSG Messenger.CONS God when order.SC.3MSG-them  
*'amara-hum min l-'a'māli bi-mā*  
 order.SC.3MSG-them from ART-deeds with-REL  
*yufīqūna* (Ṣaḥīḥ 13)  
 3MPL.be\_capable.PC  
 "Whenever the Messenger of God ordered them [i.e., the Muslims], he ordered them the deeds that they were capable of"
- (19) *fa-baynā l-qawmu 'alā dālika mina l-'amri ... 'id*  
 CONN-while ART-people on DEM.MSG of ART-thing ... when  
*xaraḡa rasūlu llāhi* (Mağāzī 214)  
 come.SC.3MSG Messenger.CONS God  
 "And while the people were [concerned] with that affair ... there came out the Messenger of God"
- (20) *fa-xaraḡa mina l-ḡadi yurīdu*  
 CONN-go\_out.SC.3MSG from ART-following\_day 3MSG.want.PC  
*l-Mirbada fa-'idā raḡulun yunšidu* ('Ağānī 1, 259)  
 al-Mirbad CONN-when man 3MSG.recite.PC  
 "And he went out the following day towards al-Mirbad and there [was] a man reciting"
- (21) *fa-baynā huwa yawman min 'ayyāmi-hī ya'kulu fī ba'ḍi*  
 CONN-while he day of days-his 3MSG.eat.PC in some  
*l-mawāḍi'i 'id marra bi-hī raḡulun* (Buxalā' 47)  
 ART-places when pass.SC.3MSG by-him man  
 "And while he was eating one day is some place, [suddenly] a man passed by him"



The term ‘verbal complex’ is applied here to a group of constructions that consist of a matrix verb and a dependent verb; the first is usually realized by *fa’ala*, the second by *yaf’alu*. Matrix verbs can be ordered on a scale extending from a pole of lexically depleted verbs to a pole of lexically full verbs. Dependent (predicative or adverbial) *yaf’alu* forms are compatible with all classes of matrix verbs: they may be combined with auxiliaries, function verbs, verbs of perception and permission, motion and setting verbs, and verbs of speech and action. *fa’ala* forms, by contrast, are mostly not found in the position of the predicative. They are only attested with the auxiliary *kāna* and with perception verbs, when these indicate ‘internal vision’ or the perceiving of a fact. However, also in the last two cases, *fa’ala* is not too common, giving way to *qad fa’ala* as the perfective (or resultative) member of the ‘predicative paradigm’ (see Table 6).

Certain dependent clauses do not allow for the opposition between *fa’ala* and *yaf’alu*: (1) substantival and adverbial clauses linked by *’an* and *ḥattā*, exhibiting an opposition between *fa’ala* and the subjunctive *yaf’ala*;<sup>13</sup> (2) adverbial clauses linked by *mā l-daymūma* ‘the *mā* of continuance’ wherein *fa’ala* is the sole option; (3) circumstantial clauses linked by *wa-* in which *yaf’alu* shares a paradigm with *qad fa’ala* and the participle, rather than with *fa’ala*.<sup>14</sup>

Mutually-dependent constructions consist of two units neither of which can be viewed as modifying the other, that is, neither can be reduced without drastically altering the meaning of the entire construction. Conditionals, setting, and presentative clauses form mutually-dependent constructions. They show a clear division between *fa’ala* and *yaf’alu*: conditionals and preposed temporal clauses referring to an anterior event, as well as ‘dynamic’ presentative clauses, depicting a peak-event, exhibit *fa’ala* forms; setting clauses or ‘static’ presentative clauses, depicting an unfolding scene, exhibit *yaf’alu* forms, as summarized in Table 5 below.

Table 5. Mutually-dependent constructions

CLAUSE TYPE		FORMAL STRUCTURE	
conditional		<i>’in fa’ala</i>	<i>fa’ala</i>
temporal~conditional		<i>’idā fa’ala</i>	<i>fa’ala</i>
preposed anterior		<i>lammā fa’ala</i>	<i>fa’ala</i>
setting	dynamic presentative	<i>baynamā huwa yaf’alu</i>	<i>’id fa’ala</i>
	static presentative	<i>fa’ala</i>	<i>’idā huwa yaf’alu</i>

13. Examples of *ḥattā* with *yaf’alu* are extremely rare and stem mostly from poetry, see Reckendorf (1921: 477–478).

14. For a detailed discussion of syndetic and asyndetic circumstantial clauses in Classical Arabic, see Waltisberg 2009.

Syntactic environments that preclude *fa’ala* forms, partially or totally, exhibit the same verbal paradigm. This paradigm, referred to above as the ‘predicative paradigm’, consists of the forms *yaf’alu*, *qad fa’ala* and the participle, indicating an ongoing situation, an outcome and a state, respectively. The distributional division between *fa’ala* and this paradigm is found not only at the clause-level, e.g., in verbal complexes, but also at the text level, particularly in narratives. In the narrative, *fa’ala* is the ‘evolution’ form, indicating the main stream of events, whereas *yaf’alu*, *qad fa’ala* and the participle are all ‘commentative’ forms, providing background information that completes the plot.<sup>15</sup> Table 6 summarizes the discussion of syntactic environments that present a marginal opposition or no opposition at all between *fa’ala* and *yaf’alu*; forms in round brackets are marginal or rare, forms in square brackets are restricted in their distribution.

Table 6. The paradigms of simple and modified forms in environment C, D, and E

C (VERBAL COMPLEX)		D (DEPENDENT)		E (MUTUALLY-DEPENDENT)	
<i>kāna</i> , <i>ra’ā</i>	( <i>fa’ala</i> )	<i>’an</i> ,	<i>fa’ala</i>	<i>’in’iḏā</i> , <i>lammā</i>	<i>fa’ala</i>
	( <i>sa-yaf’alu</i> ) <sup>16</sup>	<i>ḥattā</i> , <i>mā</i>	[ <i>yaf’ala</i> ]	<i>’iḏ</i>	[ <i>yaf’al</i> ]
all range	<i>yaf’alu</i>		<i>yaf’alu</i>		<i>yaf’alu</i>
	<i>fā’ilan</i>	<i>wa-</i>	<i>fā’ilun</i>	<i>baynamā</i> , <i>’iḏā</i>	<i>fā’ilun</i>
	<i>qad fa’ala</i>		<i>qad fa’ala</i> <sup>17</sup>		<i>qad fa’ala</i>

## 4.2 Compatibility with particles

The forms *fa’ala* and *yaf’alu* can be preceded by the modifying particles *qad*, *la-*, and *sawfa/sa-*. The modifier *qad* can precede both *fa’ala* and *yaf’alu*, whereas *la-* and *sawfa/sa-* are combined as a rule with *yaf’alu*.<sup>18</sup> The modifier *la-* is a focus particle, occurring mainly in clauses introduced by *’inna*. Clauses of the pattern

15. According to Shisha-Halevy (2007: 34–35), the narrative is structured in two modes: the Evolution Mode, “carrying the course and unfolding of the plot”, and the Comment Mode, “giving reasons for main-line acts and states and information on prior and anterior action”.

16. *sa-yaf’alu* is rarely attested as the predicative of *ra’ā*. In these cases *ra’ā* does not indicate concrete perception but rather cognitive, phantastic vision, see Waltisberg (2009: 185–186).

17. *qad fa’ala* is rarely preceded by its subject, as opposed to *yaf’alu* and *fā’ilun* that, within a syndetic circumstantial clause, are preceded as a rule by their subject.

18. The particle *la-* may precede *fa’ala* in the *ḡawāb*, i.e., ‘complement’ or ‘apodosis’, of an oath or a hypothetical conditional sentence; in these cases, *la-* is traditionally analyzed as a component of the sentential structure, rather than of the verbal form, cf. Wright (1896: I, 282–283).

'inna ... la-, as opposed to other clauses introduced by 'inna, do not show an opposition between *fa'ala* and *yaf'alu*. Besides *yaf'alu*, *la-* may also precede the participle and *qad fa'ala*, thus tying again these three forms together against *fa'ala*. The particle *sawfa/sa-*, which can only precede *yaf'alu*, provides an explicit mark for the posterior time reference of the event indicated by *yaf'alu*.

The modifier *qad* is the only one with which the opposition between *fa'ala* and *yaf'alu* can be examined. Here, as well, a lack of symmetry is shown by the forms, especially with regard to their frequency and distribution: *qad fa'ala* is attested far more frequently than *qad yaf'alu*, and, unlike the latter, it occurs not only in independent but also in dependent clauses. It seems, in fact, that the modified *qad fa'ala*, which prevails in syntactic environments where the simple *fa'ala* is not so often used, if at all (see Table 6), has become syntactically and functionally distinct from the latter. In contrast, *qad yaf'alu* is used more sporadically and does not seem to have acquired a dedicated function. As for the internal interaction between the modifier *qad* and the verbal forms, one may argue that *qad* serves the same function with both *fa'ala* and *yaf'alu*: in the first case, it stresses the perfective/certain meaning of *fa'ala*, while in the latter case it stresses the imperfective/uncertain meaning of *yaf'alu*. However, such a claim would be too reductive, considering the tight cohesion of *qad* and *fa'ala*, which has given rise to a further grammaticalization of this form, vis-à-vis the infrequent combination of *qad* and *yaf'alu*, which remains a marginal pattern in the verbal system of Classical Arabic.<sup>19</sup>

### 4.3 Clause types ('word-order')

Another point where *fa'ala* and *yaf'alu* present divergence concerns the clause type in which they are realized. The distinction between the two basic clause types in Classical Arabic – the so-called 'verbal clause' and 'nominal clause' – resides not only in the linear order of the subject and the predicate but also in their agreement relations. In a verbal clause, the first position is taken by the verb, which does not agree in number and possibly in gender with the following nominal subject. A nominal clause, by contrast, exhibits full agreement between the topicalized nominal element and the following nominal or verbal predicate. The two clause types

19. For a short summary on the various modal and temporal-aspectual functions indicated by *qad fa'ala* in Classical Arabic, see Kinberg (1989: 170–171). Bahloul (2008: 72–103) dedicates a whole chapter to the modal particle *qad* and its co-occurrence with *fa'ala*. He arrives at the conclusion that in Modern Standard Arabic *qad* is 'inherently assertorial' (101), and thus its various temporal, aspectual, and modal functions are all derived from this invariable meaning.

mark the functional distinction between a ‘block-predication’ or an ‘event-oriented’ clause, and a topicalization or an ‘entity-oriented’ clause.<sup>20</sup>

Independent clauses that are not introduced by a certain operator (e.g., *inna* and its ‘sisters’) can be structured as either verbal or nominal clauses. The verbal forms *fa’ala* and *yaf’alu* show different tendencies in independent clauses: *fa’ala* occurs nearly always in verbal clauses, whereas *yaf’alu* is realized quite often in nominal clauses (cf. Khan 1988: 30–31). Consequently, an extraposed noun or pronoun preceding *fa’ala* carries a higher informative prominence than one preceding *yaf’alu*. An extraposed entity occurring with *fa’ala* is found in contexts where a contrast obtains between the utterance and the (implicit or explicit) surrounding context, while the same entity preceding *yaf’alu* serves to indicate the (natural or given) topic entity, as shown in (22a)–(22b):

- (22a) *yā banī ‘Adiyyin kayfa raġa’tum lā fī l-‘iri*  
 VOC banū ‘Adiyy Q return.SC.2MPL NEG in ART-caravan  
*wa-lā fī l-naḡīri qālū ‘anta ‘arsalta ‘ilā*  
 CONN-NEG in ART-group say.SC.3MPL you send.SC.2MSG to  
*Qurayšin ‘an tarġī’a fa-raġa’a man*  
 Qurayš COMP 3FSG.return.SBJV CONN-return.SC.3MSG REL  
*raġa’a wa-maḡā man maḡā* (Maġāzī 45)  
 return.SC.3MSG CONN-go\_away.SC.3MSG REL go\_away.SC.3MSG  
 “O Banū ‘Adiyy, how is it that you returned neither in a caravan nor in a group? They said: *You* sent to Qurayš to return, so some went back and some went away”

- (22b) *qāla wa-mā radda ‘alay-ka bnu ‘Ubayyin*  
 say.SC.3MSG CONN-Q answer.SC.3MSG to-you son.CON.S ‘Ubayy  
*fa-qāla Ġudayyun lam ‘ara ‘inda-hū xayran*  
 CONN-say.SC.3MSG Ġudayy NEG 1SG.see.JUSS from-him good  
*qāla ‘anā ‘ursilu ‘ilā ḡulafā’-ī fa-yadḡulūna*  
 say.SC.3MSG I 1SG.send.PC to allies-my CONN-3MSG.join.PC  
*ma’a-kum* (Maġāzī 370)  
 with-you  
 “He said: And what did Ibn ‘Ubayy answer you? Ġudayy said: I saw no good from him. He [i.e., Ḥuyayy] said: I will send [a messenger] to my confederates and they will join you”

20. For an outline of the essential distinction between the two basic sentence patterns in Arabic, the ‘verb-initial’ and the ‘topicalized’ sentence, see Goldenberg (2006). The distinction between ‘event-oriented’ and ‘entity-oriented’ clauses is first outlined in Khan (1988: 30–31), and further elaborated in Holes (2004: 252–253).

4.4 Lexical classes

The overall meaning of the verb is affected not only by its grammatical pattern (i.e., suffix or prefix conjugation) but also by the specific verbal lexeme which it realizes. Various lexical classes bring about different interpretations of *fa'ala* and *yaf'alu* and therefore different oppositions between them. The distribution of the verbal forms, too, may be determined by the lexical class, since particular lexemes may show preference of one form over the other.

The distinction between dynamic (action) verbs and stative verbs is highly significant to the interpretation of *fa'ala* and *yaf'alu*. Generally speaking, dynamic *fa'ala* forms depict an 'over-and-done' event, whereas dynamic *yaf'alu* forms are interpreted by default as prospective or intentional (unless implied otherwise by the context).<sup>21</sup> Stative *fa'ala* forms, indicating a mental state, may refer to a realized state of affairs that still persists at the time of the dialogue, or at the time indicated in the matrix clause. Stative *yaf'alu* forms, indicating a mental state, comprehension, and physical perception, are interpreted by default as concomitant with the dialogue situation or with that indicated in the matrix clause. Table 7 summarizes this short description, and it is followed by one illustration of the different perspectives indicated by *fa'ala* and *yaf'alu* with mental verbs.

Table 7. Interaction of dynamic and stative lexemes with *fa'ala* and *yaf'alu*

LEXICAL CLASS	<i>fa'ala</i>	<i>yaf'alu</i>
dynamic	over-and-done <sup>22</sup> (anterior)	prospective, intentional (posterior)
stative (mental, perception)	realized and persisting (resultative) over-and-done (anterior)	concurrent (simultaneous)

21. Dynamic *yaf'alu* forms assuming a concomitant ('present') interpretation refer either to habitual or to individual concrete situations. The habitual reading is usually advanced by adverbs such as *kulla yawmin* 'everyday' or by topicalization structures such as '*ammā*... [*fa*-] "as to ..."; the concrete reading is usually induced by anaphoric reference, i.e., *yaf'alu* repeats an already (explicitly or implicitly) mentioned information, e.g., *fa-dakara-hā fa-'adraka-humā 'Abū Ġahlin fa-qāla mā tuḥaddiṭāni bi-hi qālā naḍkuru ru'yā 'Ātika* (Mağāzī 41) – "And he recollected it, and Abū Ġahl reached them and asked: What are you talking about? They said: We are recollecting the dream of 'Ātika."

22. The descriptions 'over-and-done' and 'concurrent' are preferred to the common terms 'past' and 'present' in the intention of departing from the time-line metaphor and representing what is seen here as more crucial, i.e., the subjective perspective of the speaker.

- (23) *mā yuḥzinu-ka*                      ‘ayyuhā l-maliku wa-qad  
 Q    3MSG.make\_sad.PC-you VOC        ART-king    CONN-MOD  
 ‘aẓfara                      llāhu yada-ka        wa-‘ahlaka  
 make\_triumph.SC.3MSG    God    hand-yours    CONN-destroy.SC.3MSG  
 ‘aduwwa-ka fa-qāla                      l-‘asadu ḥazintu        ‘alā ‘aqli  
 enemy-yours    CONN-say.SC.3MSG    ART-lion    be\_sad.SC.1SG    on    mind.CON  
 l-ṭawri wa-karami                      xulqi-hī                      (Kalīla wa-Dimna 121)  
 ART-ox    CONN-nobility.CON    nature-his  
 “What makes you sad, O king? God has made you triumph and destroyed  
 your enemy! The lion said: I am sad for the [clever] mind of the ox and his  
 noble nature”

A lexical class that determines different functional relations between *fa’ala* and *yaf’alu* is the category of performative verbs. In Classical Arabic, this functional category is divided between *fa’ala* (or *qad fa’ala*) and *yaf’alu*. Performatives of the *fa’ala* type occur mainly in contexts of transactions and agreements (or their break), e.g., *wafat ḍimmatuka* (Sira 243) “Your liability [to me] is hereby over”. This type of performative does not form a homogeneous lexical class. Performatives of the *yaf’alu* type – or better, the *’af’alu* type, as they all feature the first person – are declarative and include speech verbs such as: *’ašhadu* “I swear”, *’aḍkuru* “I state”, *’a’ūdū bi-llāhi* “I seek protection in God”, *’ahlifu* “I take an oath”, *’ad’ū* “I call”, *’as’alu* “I ask”. With *fa’ala*, this class of speech verbs is interpreted as all other classes of dynamic verbs, thus marking an opposition of a different order with the corresponding performative *yaf’alu* forms (see (2)).

Certain verbal lexemes show preference of one form over the other. Such, for instance, are the verbs *šadaqa* “to say the truth” and *kaḍaba* “to lie, say what is untrue”, which are usually realized as *fa’ala* forms, not only in narratives but also in dialogues. In dialogues, *šadaqa* and *kaḍaba* are often used to convey a judgment that has present relevance rather than to state a fact about the past, e.g., *šadaqta* “You said the truth” → “You are right”.

#### 4.5 Textual domains

The term ‘textual domain’ refers here to a defined macro-syntactic unit, such as a narrative or an expository unit. A textual domain is characterized, inter alia, by the clause types, and further down, by the verbal forms prevailing in it.<sup>23</sup> Classical

23. Cf. Cohen 2006, for the characterization of the dialogue and the narrative textemes in the Old Babylonian epic, in view of a cluster of syntactic features, e.g., personal sphere, modality, the information structure of the clause, and forms of verbal and non-verbal predication.

Arabic narratives, as already noted by Khan (1988: 30–31), contain far more verbal clauses with *fa'ala* forms than expository and descriptive texts, of which nominal clauses with *yaf'alu* forms are characteristic.

The verbal forms *fa'ala* and *yaf'alu* are not interchangeable at the macro-syntactic level: the narrative, specifically its plotline, is the domain of *fa'ala* forms, whereas expository texts, especially ones of generic nature, are the domain of *yaf'alu* forms. In few cases, however, one encounters a verbal form in the 'natural' domain of the other, a fact implying that, in a specific context within the domain, a subtle contrast between the forms can exist. Following are two examples of *fa'ala* and *yaf'alu* exceeding their domains. In the first, *fa'ala* (with a generic theme) assumes a generic sense, advanced by the context of a 'sweeping' negation (Marmorstein 2017). That is, rather than referring to the non-occurrence of an episode, *mā fa'ala* refers to the whole interval of time in which a certain occurrence did not take place. In the second example, *yaf'alu* occurs in a narrative sequence, as is occasionally the case with the verb *yağidu* 'to find'. Sequential *fa-yaf'alu* forms indicate an event that ensues or reacts to the previous event indicated by *fa'ala*.<sup>24</sup>

- (24a) *mā naqaša mālun qaṭṭu min zakātin* (Buxalā' 50)  
 NEG decrease.SC.3MSG money never of charity  
 "Money never grew less through charity"
- (24b) *fa-ndaḥa'ū tilqā'a l-Zuraybi fa-yağidūna 'alā*  
 CONN-proceed.SC.3MPL towards al-Zurayb CONN-3MPL.find.PC on  
*tilka l-qalibi llati qāla rasūlu llāhi*  
 DEM.FSG ART-well REL.FSG say.SC.3MSG messenger.CONS God  
*rawāyā Qurayšin fi-hā suqqā'u-hum* (Mağāzī 51)  
 water\_carriers.CONS Qurayš in-it water\_carriers-their  
 "They proceeded towards al-Zurayb and found at that well, which the Messenger of God mentioned, the watering camels of Qurayš, [and] in it their water carriers"

24. See also *Mağāzī* 27, *Mağāzī* 364, and the first of Nöldeke's examples (1963: 68) for a narrative perfect followed by an imperfect. The phenomenon of sequential *yaf'alu* forms in Classical Arabic narratives is discussed in detail in Marmorstein (Marmorstein 2016).



## 5. Conclusions

This study presented a re-examination of the opposition between the two simple finite forms *fa'ala* and *yaf'alu* in Classical Arabic, in view of an array of grammatical and lexical features. The examination has shown that the opposition between *fa'ala* and *yaf'alu* does not exist in any environment and, where it does, it serves to indicate a variety of semantic distinctions. Certain SYNTACTIC ENVIRONMENTS, at the clause-level (e.g., the predicative slot) and at the text level (e.g., the narrative unit), do not exhibit an opposition between the forms but rather a division between *fa'ala* on the one hand and *yaf'alu*, *qad fa'ala*, and the participle on the other. Moreover, *fa'ala* and *yaf'alu* show divergence in the CLAUSE-TYPES in which they mainly occur, and in their COMPATIBILITY with modifying particles. The two forms also present different types of interaction with various classes of VERBAL LEXEMES, which give rise to a nuanced range of interpretations of *fa'ala* and *yaf'alu*. Considering all these parameters, it is evident that the complexity of the system, which besides *fa'ala* and *yaf'alu* consists of other simple, modified, and compound forms, cannot be reduced to a single temporal or aspectual dichotomy. Rather, the synthesis must account for the entire list of diverse cases that was shown to exist in the present study. The ultimate goal of such a synthesis would not be to expose *fa'ala* and *yaf'alu*'s invariant meanings, but to identify the mechanism, the system of interrelations underlying the structure of the verbal system of Classical Arabic.

## References

### Primary sources

- 'Agānī=Rannāt al-maṭāliḥ wal-maṭānī fī riwāyāt al-'aḡānī. 2 vols. 1923 (vol I), 1946 (vol II). (Extracts from *Kitāb al-'Aḡānī*, 'Abū al-Faraḡ al-'Iṣbahānī [d. 967]). Edited by Anṭūn Ṣāliḥānī. Beirut: al-Maṭba'a al-Kāṭūlikiyya.
- Buxalā'*= 'Abū 'Uṭmān 'Amr b. Baḥr b. Maḥbūb al-Ġāḥiẓ [d. 869], *Al-Buxalā'*. 1993. Edited by 'Abbās 'Abd al-Sātir. Beirut: Dār wa-Maktabat al-Hilāl.
- Kalīla wa-Dimna*= 'Abdallāh 'Ibn al-Muqaffā' [d. 757], *Kalīla wa-Dimna*. 1926. Edited by Louis Cheikho. Beirut: Maṭba'at al-'Ābā' al-Yasū'iyyīn.
- Maḡāzī*= 'Abū 'Abdallāh Muḥammad b. 'Umar al-Wāqidi [d. 823], *Kitāb al-Maḡāzī*. vol 1. 1966. Edited by Marsden Jones. London: Oxford University Press.
- Ṣaḥīḥ*= 'Abū 'Abdallāh Muḥammad b. 'Isma'īl al-Buxārī [d. 870], *al-Ġāmi' al-Ṣaḥīḥ*. vol 1. 1862. Edited by Ludolf Krehl. Leiden: Brill.
- Sīra*= 'Abū Muḥammad 'Abd al-Malik b. Hišām [d. 834], *Kitāb Sīrat Rasūl Allāh*. vol 1. 1858. Edited by Ferdinand Wüstenfeld. Göttingen: Dieterichsche Universitäts-Buchhandlung.
- Ta'riḫ*= 'Abū Muḥammad b. Ġarīr al-Ṭabarī [d. 923], *Ta'riḫ al-rusul wa-l-mulūk*. 15 vols (in 3 series). 1879–1901 (reprint 1964). Edited by Michael J. de Goeje. Leiden: Brill.



## Secondary sources

- Aartun, Kjell. 1963. *Zur Frage altarabischer Tempora*. Oslo: Universitetsforlaget.
- Bahloul, Maher. 2008. *Structure and Function of the Arabic Verb*. London: Routledge.
- Beeston, Alfred Felix Langdon. 1970. *The Arabic Language Today*. London: Hutchinson University Library.
- Brockelmann, Carl. 1908–1913. *Grundriss der vergleichenden Grammatik der semitischen Sprachen*, 2 vols. Berlin: Reuther & Reichard.
- Cohen, David. 1989. *L'aspect verbal*. Paris: Presses Universitaires de France.
- Cohen, Eran. 2006. "The Tense-Aspect System of the Old Babylonian Epic". *Zeitschrift für Assyriologie* 96.31–68. doi:10.1515/ZA.2006.002
- Cohen, Marcel. 1924. *Le système verbal sémitique et l'expression du temps*. Paris: Imprimerie nationale.
- Comrie, Bernard. 1976. *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*. Cambridge: Cambridge University Press.
- Ewald, Heinrich. 1831–1833. *Grammatica critica linguae arabicae*. Leipzig: Hahn.
- Ewald, Heinrich. 1891. *Syntax of the Hebrew Language of the Old Testament*, translated by James Kennedy. Edinburgh: Clark.
- Fleisch, Henri. 1961–1979. *Traité de philologie arabe*. 2 vols. Beirut: Dar el-Machreq.
- Fleischer, Heinrich Leberecht. 1885. *Kleinere Schriften*, vol 1. Leipzig: S. Hirzel.
- Fischer, Wolfdietrich. 2002. *A Grammar of Classical Arabic*, 3rd edn. trans. by Jonathan Rodgers. New Haven: Yale University Press.
- García, Erica. 1991. "Grasping the Nettle: Variation as Proof of Invariance". *New Vistas in Grammar: Invariance and Variation*, ed. by Linda R. Waugh & Stephan Rudy, 31–59. Amsterdam: John Benjamins. doi:10.1075/cilt.49.05gar
- Gaudefroy-Demombynes, Maurice & Blachère, Régis. 1952. *Grammaire de l'arabe classique*, 3rd edn. 2 vols. Paris: G.P. Maisonneuve.
- Goldenberg, Gideon. 1966. *The Amharic Tense-System*. Jerusalem: Hebrew University dissertation.
- Goldenberg, Gideon. 2006. "On Grammatical Agreement and Verb-Initial Sentences". *Loquentes linguis: Studi linguistici e orientali in onore di Fabrizio A. Pennacchietti*, ed. by Giorgio Borbone, Alessandro Mengozzi & Mauro Tosco, 333–339. Wiesbaden: Otto Harrassowitz.
- Götz, Manfred. 1980. "Bemerkungen zu den Tempora des Althocharabischen". *Studien aus Arabistik und Semitistik: Anton Spitaler zum siebzigsten Geburtstag von seinen Schülern überreicht*, ed. by Werner Diem & Stefan Wild, 86–98. Wiesbaden: Otto Harrassowitz.
- Holes, Clive. 2004. *Modern Arabic: Structures, Functions, and Varieties*. Washington: Georgetown University Press. 2nd edition.
- Kinberg, Naphtali. 1989. "Some Modal, Aspectual and Syntactic Constraints on the Use of *qad* in the Verbal System of Classical Arabic". *Studia Linguistica et Orientalia Memoriae Haim Blanc Dedicata*, ed. by Paul Wexler, Alexander Borg & Sasson Somekh, 170–179. Wiesbaden: Otto Harrassowitz (=Kinberg, Naphtali. 2001. *Studies in the Linguistic Structure of Classical Arabic*. Leiden: Brill, 121–131).
- Khan, Geoffrey. 1988. *Studies in Semitic Syntax* (London Oriental Series 38). Oxford: Oxford University Press.
- Lehmann, Christian. 1988. "Towards a Typology of Clause Linkage". *Clause Combining in Grammar and Discourse*, ed. by John Haiman and Sandra A. Thompson, 181–225. Amsterdam: John Benjamins. doi:10.1075/tsl.18.09leh

- Marmorstein, Michal. 2016. "Subtleties of Narrative Syntax: Patterns of Chaining in Classical Arabic Narratives". *Jerusalem Studies in Arabic and Islam* 43, 219–249.
- Marmorstein, Michal. 2017. "On Verbal Generic Sentences in Classical Arabic". *Labor omnia uicit improbus: Miscellanea in Honorem Ariel Shisha-Halevy*, ed. by Nathalie Bosson, Anne Boud'hors & Sydney H. Aufrère, 493–518. Louvain: Peeters (Orientalia Lovaniensia Analecta).
- Nöldeke, Theodor. 1963. *Zur Grammatik des Classischen Arabisch. Im Anhang: Die handschriftlichen Ergänzungen in dem Handexemplar Theodor Nöldekes, bearbeitet und mit Zusätzen versehen von Anton Spitaler*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Reckendorf, Hermann. 1895–1898. *Die Syntaktischen Verhältnisse des Arabischen*, 2 vols. Leiden: Brill.
- Reckendorf, Hermann. 1921. *Arabische Syntax*. Heidelberg: Carl Winter.
- Reuschel, Wolfgang. 1996. *Aspekt und Tempus in der Sprache des Korans* (Leipziger Beiträge zur Orientforschung 6). Frankfurt am Main: Peter Lang.
- de Sacy, Antoine Isaac Silvestre. 1831. *Grammaire arabe à l'usage des élèves de l'École spéciale des langues orientales vivantes*, 2 vols. Paris: Imprimerie royale. 2nd edition.
- Shisha-Halevy, Ariel. 2007. *Topics in Coptic Syntax: Structural Studies in the Bohairic Dialect*. Leuven: Peeters.
- Waltisberg, Michael. 2009. *Satzkomplex und Funktion: Syndese und Asyndese im Althocharabischen*. Wiesbaden: Otto Harrasowitz.
- Wright, William. 1896–1898. *A Grammar of the Arabic Language*, 3rd edn. 2 vols. Cambridge: Cambridge University Press.

**PAGE PROOFS**  
© JOHN BENJAMINS PUBLISHING COMPANY

**2nd proofs**

# The imperfective in Berber

## Evidence of innovated forms and functions

Mena Lafkioui

Llacan-CNRS/Université Sorbonne Paris Cité

This article analyzes the morphological oppositions and semantic distinctions that originate from transformation processes regarding the imperfective in Berber. The synchronic and diachronic phenomena studied in this contribution offer examples of how language is continually modulated through innovations that emerge in structurally layered and causal formations dictated by system-based properties.

**Keywords:** imperfective, comparative Berber, morphological verbal innovation, semantic diversification, diachronic verbal development, system-internal language change

### 1. Introduction

This study addresses several morphological innovation processes of the Berber verbal system driven by linguistic (formal and functional) parameters. It focuses on how these transformations affect morphological oppositions and related semantic thematic distinctions, at both a synchronic and a diachronic level. In doing so, I offer evidence that confirms the importance of system-internal induced language variation and change.<sup>1</sup>

---

1. The following abbreviations are used in this article: A ‘aspect(ual)’, ACC ‘accusative’, AOR ‘aorist’, DIST ‘distal’, IMP ‘imperative’, NIMP ‘negative imperative’, IPFV ‘imperfective’, INT ‘intonation’, M ‘masculine’, MOD ‘modality’, NEG ‘negation’, NIPFV ‘negative imperfective’, NPFV ‘negative perfective’, NOM ‘nominative’, O ‘object’, PFV ‘perfective’, PL ‘plural’, PRSM ‘personal marker’, PTCP ‘participle’, PTCPM ‘participle marker’, S ‘subject’, SG ‘singular’, V ‘verbal syntagm’, + ‘marked’, – ‘unmarked’. The original transcription of the cited Berber examples is maintained. Examples from my own corpus are transcribed according to the general Berber phonological system and to IPA. All English glosses of the Berber examples are my own.

The current Berber verbal system is based on a fundamental morphological opposition of perfective versus imperfective for the positive aspects and perfective versus negative perfective for the negative aspects (Basset 1952; Chaker 1989; Galand 1977; Lafkioui 2007: 174–191). Tuareg, which is mostly spoken in the Sahara area, and Tarifit, mainly spoken in northern Morocco, differ considerably from this basic system in that they have developed a series of secondary morphological verbal oppositions which mark distinctive semantic values (Lafkioui 2007: 174–191; Lafkioui & Kossmann 2009; Leguil 1987, 2000; Prasse & ägg-Älböşan ägg-Sidiyän 1985; see Figure 1, dark grey zones). Although both languages show derivational mechanisms generating these new variants, there is a significant difference between them with respect to the precise transformation processes involved and the corresponding reorganization of the verbal paradigmatic structure. These phenomena will be discussed in detail in the ensuing sections.

From a typological perspective, the Berber languages, belonging to the Afroasiatic phylum, are predominantly synthetic (inflection, derivation and compounding) and inflecting. Common features are their VSO base word order, their pro-drop feature (i.e., verbal constructions obligatorily contain an incorporated subject marker), their preposition-noun sequence, their possessive suffixes and their mixed morphological plural formation (affixation and/or apophony). Aside from noun-verb oppositions, all other word class distinctions are unclear in Berber. The Berber languages also provide accounts of one of the indisputably typological linguistic features of Africa. This is the marked-nominative, a feature which is barely attested outside Africa but is present in most case languages within it, including those of East Africa (König 2006; Lafkioui forthcoming). In these systems, it is the nominative that is functionally marked with reference to the accusative, even if the languages in question may differ as regards the morphosyntactic marking procedures.

For example in Tamazight Berber (central Morocco), we find:

- (1) *i-ča*                      *urgaz*                      *ağrum*  
       eat-PFV-3MSG man-NOM bread-ACC  
       “The man ate bread.”

The nominative is encoded by the morphological marker *u-* (dependent state marker) and a continuous intonation unifying the noun with the preceding verbal syntagm, while the accusative is encoded by the absence of these morphemes and so matches the unmarked form; i.e., *a-* (accusative, independent state) versus *u-* (nominative, dependent state).

---

This work is part of the program “Investissements d’Avenir” overseen by the French National Research Agency, ANR-10-LABX-0083 (Labex EFL)

An examination of the Berber verbal system follows this introduction (§ 2). In § 3, the negative imperfective and its specific features are considered. § 4 analyzes the innovations of the Tuareg and Tarifit verbal system, with a focus on the morphology-semantics interface. The article ends with a number of conclusions about language variation and its connection with language evolution.

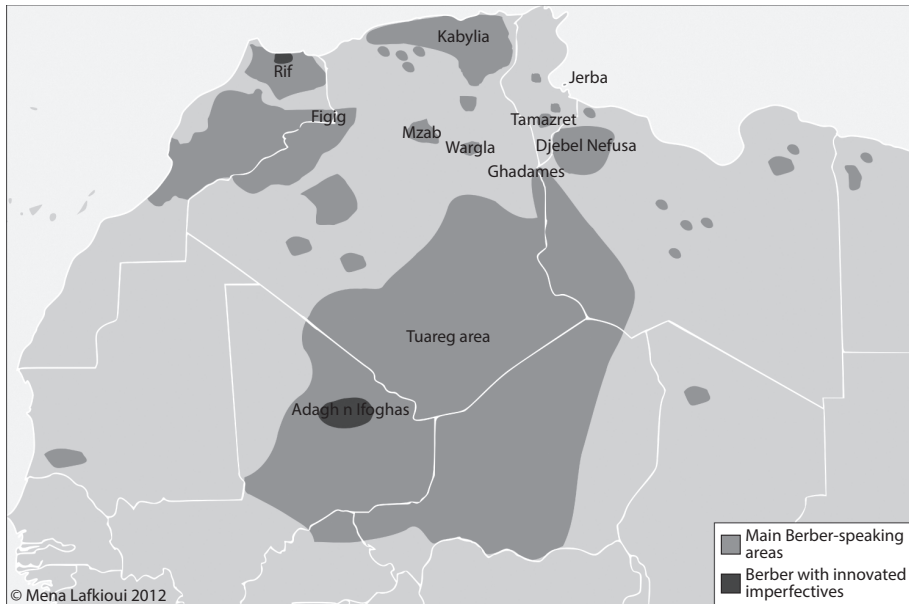


Figure 1. Berber areas with innovated imperfectives

## 2. The Berber verbal system

The Berber verbal system is of a Root-Pattern type. It is structured around a complex aspectual hierarchical configuration of three levels, which are structurally more complex and semantically more specific as one ascends the hierarchy. It displays a predominance of tri-radical roots and makes use of both consonant length and intra-radical vowel alternation (apophony) to indicate aspectual categories. From a synchronic perspective, the Berber verbal form is composed of a stem and a personal marker (PRSM) or participle marker (PTCPM). The stem itself consists of a consonantal root, which refers to a central semantic value, and a vowel pattern that orientates or specifies this value:

1. Verb = stem + person or participle marker (highest level)
2. Stem = root + vowel pattern (middle level)
3. Root = consonantal radicals (basic level)

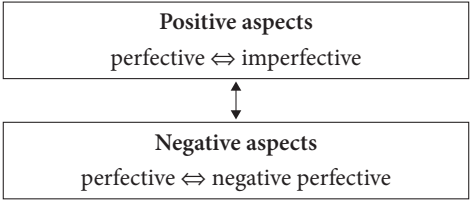
Examples of verbal forms:

- (2) *y-udf* [juðəf] “he has entered” (PFV-3MSG) = PRSM *y-* (3MSG) + stem *-udf*
- (3) *y-udf-n* [juðfən] “has/have entered” (PFV-PTCP) = PTCPPM *y-en* + stem *-udf*  
⇒ Stem: *udf* [uðəf] = root /d/ (action “to enter”) + pattern /u- ( ) ( )  
⇒ Bi-radical root: /cc/ with /d/ [ð] and /f/ [f] as radicals (action “to enter”)

Stem and person/participle markers are inevitably connected in the verbal syntagm; the former functions as predicate and the latter as subject. The subject marker is encoded on the verb as a prefix and/or a suffix and is made explicit by postverbal lexical complements with specific features, e.g., dependent state. The V(erb)-S(ubject) syntagm represents the minimal required structure of an assertion, which is expanded by an object (O) in the case of transitive verbs and various other complements insofar as the morphosyntactic adequacy and the semantic sufficiency of the verbal stem permit, with a VSO basic word order.

The contemporary Berber verbal system is based on a fundamental morphological opposition of perfective (PFV) versus imperfective (IPFV) for the positive aspects and perfective (PFV) versus negative perfective (NPFV) for the negative aspects (Basset 1952; Galand 1977; Cadi 1987: 59–65; Chaker 1989; Lafkioui 2007: 174–191). Consequently, Berber disconfirms the cross-linguistic claim that the perfective is less compatible with negation than the imperfective (Schmid 1980: 39; Matthews 1990: 84).

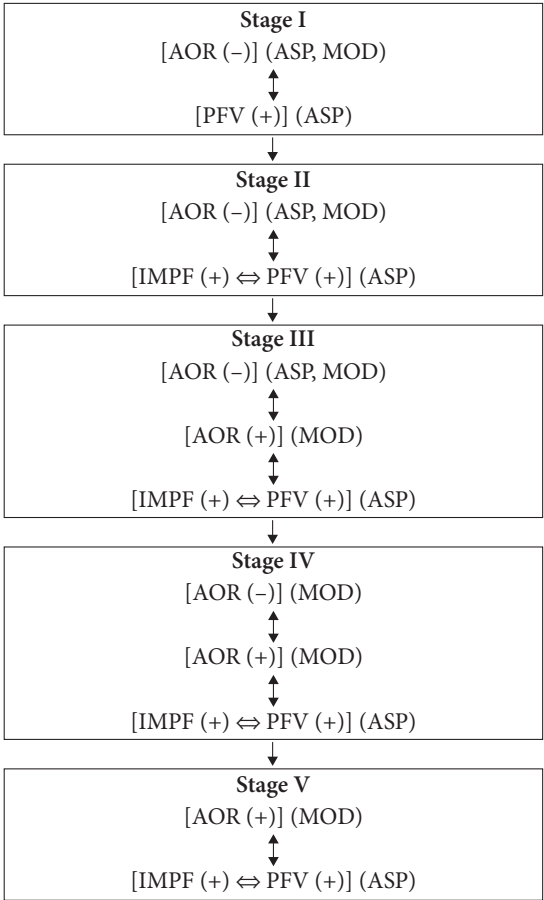
Table 1. Basic verb oppositions in Berber



Regarding the positive aspects, the following diachronic reconstruction, based mainly on Galand’s hypothesis (1977, 2002: 259–271), may well explain the various phenomena presently attested in the different Berber languages:

1. On a Proto-Berber level (stage I), the opposition [aorist (unmarked) ~ perfective (marked)] existed and initiated the evolution process that has led to the present system. The neutral aorist stood for modal values as well as for aspectual values, whereas the perfective only rendered aspectual values.
2. In stage II, still on a Proto-Berber level, derivational devices (e.g., *t(t)*-affixation, radical gemination and vowel insertion) – which later changed into aspectual markers – signaled the emergence of a new verbal theme, the imperfective, which took over the aspectual values of the aorist and, consequently,

**Table 2.** Diachronic reconstruction of the positive thematic oppositions in Berber



entered into opposition with the perfective: [aorist (unmarked) ~ {imperfective (marked) ~ perfective (marked)}]. The newly created imperfective mainly signified habituation and progression.

3. Stage III is presently attested in diverse Berber languages, such as Tashelhit (southern Morocco), Tamazight (central Morocco) and western Tarifit (also called Senhaja Berber, northern Morocco). It consists of the development of a marked aorist by means of a periphrastic marker, i.e., modality particle. The unmarked aorist is retained for specific functions, like, for instance, to express injunctions and to mark concatenation. The modality particle can also be combined with the imperfective, generating a double marking system (i.e., aspectual and modal marking) and, hence, signifying layered values. This stage could be summarized as follows: [aorist (unmarked) ~ {aorist (marked) ~ imperfective (marked) ~ perfective (marked)}].



4. In Stage IV, the aorist has lost some of its unmarked non-modal functions, although these are generally preserved in literary practices. This language phase is best represented by, for instance, Kabyle Berber (northern Algeria), Tuareg Berber, and eastern Tarifit (northern Morocco).
5. The last stage (stage V) is characterized by the loss of the 'bare' aorist so that only the marked thematic forms remain; that is, [aorist (marked) ~ imperfective (marked) ~ perfective (marked)]. This thematic evolution is principally observed in the more **eastern** Ber languages, such as the Berber language of Nefusa in north-western Libya.

As to the negative aspects, most scholars of Berber agree that these aspectual forms were derived from positive themes tracing back to Proto-Berber times, but there is no general consensus as to the exact triggers and processes. Nonetheless, two interesting tendencies regarding this phenomenon should be mentioned here. The first tendency considers the emergence of the negative aspects as a consequence of the development of an 'intensive' variant of the existing themes (e.g., intensive aorist  $\Rightarrow$  imperfective; intensive perfective  $\Rightarrow$  resultative), which subsequently would have been reanalyzed as a negative theme, particularly in the case of the negative perfective (Chaker 1996; Picard 1959; Prasse 1973). The second one regards this ancient innovation as the ultimate outcome of a phonetic transformation process triggered by the presence of a post-verbal negation marker (Brugnatelli 2002; Vycichl 1953–1955: 322). The enclitic negator would have provoked a shift in stress towards the final syllable position, which would have engendered the series of features that, at present, mark negative verbal forms in Berber, namely a shortened initial syllable or syllable group, a lengthened final syllable or syllable group and an umlaut of the vowel *i*.

### 3. The negative imperfective in Berber

A number of Berber languages have a morphologically marked negative imperfective (NIPFV) which is used in negation contexts. This is the case, for example, for Berber spoken in Figig, the Rif area, Ghadames, Jerba, Tamazret, Wargla, Mzab and the Tuareg areas (see Figure 1). However, most of the Berber languages make use of the positive imperfective in both positive and negative configurations. Given its similar marking and functional procedures in a wide range of Berber languages spread over the whole of North Africa, it is most likely that the negative imperfective is a remnant of a verbal theme in Proto-Berber (Brugnatelli 2002; Chaker 1996: 18; Kossmann 1989). Other evidence supporting this hypothesis is that those languages which lack the negative imperfective in their daily language

practices do, however, exhibit its use in ancient or literary texts, such as the medieval text in Tashelhit (southern Morocco) of Mhemmed Awzal (Van den Boogert 1997: 270) or the Kabyle expression of northern Algeria as in Example (4) (see also Brugnatelli 2002):

- (4) *ur i-ṭṭudeggir* (Mammeri 1980: 234)  
 NEG shake-NIPFV-3MSG  
 “He will not be shaken.”

The negative imperfective is generally marked by an apophonic modification, that is, the full (unreduced) vowel *a* is changed into the vowel *i*: [*a* > *i*] (5)–(7). However, the vowel *a* is maintained as such in the negative imperfective form when the corresponding *a* of the positive imperfective form is preceded by the vowels *i* or *u*, as in (8) and (9):

- |  |  |                              |
|--|--|------------------------------|
| (5) <i>i-ttaḍəf</i><br>enter-IPFV-3MSG<br>“He enters.”               | ⇔ <i>ul i-ttiḍəf</i><br>NEG enter-NIPFV-3MSG<br>“He does/will not enter.”    | <i>šay</i><br>NEG<br>(Figig) |
| (6) <i>yə-ččəṭ</i><br>beat-IPFV-3MSG<br>“He beats.”                  | ⇔ <i>u yə-ččit</i><br>NEG beat-NIPFV-3MSG<br>“He does/will not beat.”        | <i>š</i><br>NEG<br>(Jerba)   |
| (7) <i>yə-ssawar</i><br>speak-IPFV-3MSG<br>“He speaks.”              | ⇔ <i>wa yə-ssiwir</i><br>NEG speak-NIPFV-3MSG<br>“He does/will not speak.”   | <i>ša</i><br>NEG<br>(Rif)    |
| (8) <i>i-ttilaz</i><br>be hungry-IPFV-3MSG<br>“He is always hungry.” | ⇔ <i>ul i-ttilaz</i><br>NEG be hungry-IPFV-3MSG<br>“He will not be hungry.”  | <i>šay</i><br>NEG<br>(Figig) |
| (9) <i>yə-tzuzzar</i><br>winnow-IPFV-3MSG<br>“He winnows.”           | ⇔ <i>wa yə-tzuzzar</i><br>NEG winnow-IPFV-3MSG<br>“He does/will not winnow.” | <i>ša</i><br>NEG<br>(Rif)    |

Beside apophonic change, Tuareg Berber also marks the negative imperfective by shortening the first vowel of the stem (10) or by a lengthening of its last vowel (11):

- |  |   |                            |
|--|---|----------------------------|
| (10) <i>i-lâss</i><br>dress-IPFV-3MSG<br>“He dresses himself.” | ⇔ [ <i>wər</i> ] <i>i-less</i><br>NEG dress-NIPFV-3MSG<br>“He does/will not dress himself.” | (Tuareg; Prasse 1973: 109) |
| (11) <i>i-tākār</i><br>steal-IPFV-3MSG<br>“He steals.”         | ⇔ [ <i>wər</i> ] <i>i-tiker</i><br>NEG steal-NIPFV-3MSG<br>“He does/will not steal.”        | (Tuareg; Prasse 1973: 96)  |

In contrast to the positive imperfective, the negative imperfective necessarily occurs in a negation context. The only exception to this rule is the negative imperative, i.e., the interdiction, in which a ‘positive’ form is employed after a negation marker, such as in Example (12), which comes from Jerba but occurs in Tarifit as well:

- (12) *ruh*                     $\Leftrightarrow$  *yə-traḥ*                     $\Leftrightarrow$  *u*    *traḥ*                    *š*  
 go-IMP-2SG            go-IPFV-3MSG            NEG go-IMP-2SG NEG  
 “Go!”                    “He goes.”                    “Do not go!”

Nevertheless, some Berber varieties of the Rif area (northern Morocco) – especially of its eastern and western part – make use of both imperfective forms to render an interdiction (Lafkioui 2007: 176). For example, the Ayt Iznasen and the Ikebdanen (eastern Rif) employ the following expression:

- (13) *ur qqir*                    *šayt amənni*  
 NEG say-NIMP-2SG NEG SO  
 “Do not speak like that!”

The corresponding positive form of *qqir* (< *ini* “say”) in Example (13) is *qqar*. In the varieties of the Ayt Itteft (border western and central Rif), on the other hand, a free alternation between the positive form (14) and the negative form (15) is observed in interdictions:

- (14) *u tižža*                    *mmi=m*                    *ḍin*  
 NEG leave-IMP-2SG son=3MSG DIST  
 “Do not leave your son there!”
- (15) *u tižži*                    *mmi=m*                    *ḍin*  
 NEG leave-NIMP-2SG son=3MSG DIST  
 “Do not leave your son there!”

#### 4. Innovations in the Berber verbal system

Even though Tuareg and Tarifit share a comparable system-based innovating apparatus regarding the verb, they are clearly dissimilar with respect to the derivational transformations and the related formal and functional restructuring of the paradigms in question, as will be demonstrated in the next two sections. While the Tuareg case has already been examined to some extent – though further investigation is still required – the Tarifit innovations have never been the subject of a thorough study, especially from a comparative viewpoint. In what follows, both innovated systems will be addressed and compared so as to better understand the complex evolution process of the Berber verb. A starting basis for the Tarifit study

has been presented in Lafkioui (2007: 175–176) and Lafkioui & Kossmann (2009). As for Tuareg, I mainly refer to the studies of Leguil (1987, 2000) and Prasse & ägg-Älbos̄tan ägg-Sidiyän (1985).

#### 4.1 Innovations in the Tuareg verb

The Tuareg language of Mali displays secondary morphological oppositions for the positive perfective (PFV) as well as for the positive imperfective (IPFV), which were historically derived from original sets of variants. Most of the Tuareg languages have lost the original imperfective and just use the innovated one. But a number of varieties in Mali have retained the opposition between the two types of positive imperfectives, as shown in Table 3.

**Table 3.** Verb *əlməd* ‘to learn’ (Adagh of Ifoghas – Mali, Prasse & ägg-Älbos̄tan ägg-Sidiyän 1985)

	Original variants	Secondary variants
Aorist	<i>əlməd</i>	-
Positive perfective	<i>əlmād</i>	<i>əlmād</i>
Negative perfective	<i>əlmed</i>	-
Positive imperfective	<i>lämmād</i>	<i>lämmād</i>
Negative imperfective	<i>ləmməd</i>	-

The derivational mechanism of these developments is based on a modification of the stem vowel (called lengthening) and an insertion of a fixed stress accent (Heath 2005: 299 ff.; Leguil 2000):

- $\text{ə} > i$
- $\text{ə} > u$
- $\text{ä} > a$

The modification rules  $\text{ə} > i$  and  $\text{ə} > u$  are mutually exclusive and are mostly predictable on the basis of the preceding long vowel.

Regarding the perfective, the Tuareg languages belong to the few Berber languages that have developed a resultative from the positive perfective by means of a vowel lengthening and a stable accent placement (see Table 3; for details, see Galand 1974; Heath 2005; Prasse 1973).

As to the imperfective, its diachronic evolution could be explained through a six-step process, in which the Adagh varieties of Mali have reached Stage V, whereas the other Tuareg varieties have got to the final stage – that is, the stage in which the opposition between the original variants and their secondary developments has disappeared.

1. Based on Leguill's (2000) accounts, this diachronic analysis starts with a one-part structure with no axiological opposition between its habitual and progressive values. This stage (Stage I) is still observed in the Berber language of Ghadames (Libya).<sup>2</sup>

(16) [...] tən igəddälän  
 [... 3PL PTCP-IPFV]  
 “un qui les abrite (habituellement)” (< habitual)  
 [“one who habitually shelters them”]  
 “un qui était en train de les abriter” (< progressive)  
 [“one who was sheltering them”]

2. Stage II, attested in Tuareg, is characterized by a variable opposition in which a new imperfective marked by a long vowel represents the progressive, while the original unmarked imperfective stands for both progressive and habitual values:

(17) [...] *tən igʷaddälän* ⇔ [...] *tən igʷaddälän*  
 [...] 3PL PTCP-IPFV [...] 3PL PTCP-IPFV  
 “un qui les abrite (habituellement)” “un qui était en train de les abriter”  
 [“one who habitually shelters them”] [“one who was sheltering them”]  
 “un qui était en train de les abriter”  
 [“one who was sheltering them”]

3. In Stage III, the innovated imperfective in Tuareg has entirely claimed the progressive, which has left the original form with just the habitual. This development has resulted in a bipartite marked opposition, as demonstrated in (18).

(18) [...] *tən igʷəddälän* ⇔ [...] *tən igʷəddälän*  
 [...] 3PL PTCP-IPFV [...] 3PL PTCP-IPFV  
 “un qui les abrite (habituellement)” “un qui était en train de les abriter”  
 [“one who habitually shelters them”] [“one who was sheltering them”]

4. Again, a variable opposition between the two imperfectives is observed in Tuareg Berber (Stage IV), but this time their markedness is inverted. The original short form is marked and signifies the habitual. The new long form, on the other hand, is unmarked and expresses the habitual as well as the progressive:

2. Unfortunately, Leguil (2000) does not specify the exact Tuareg varieties that are subject to the changes explained in stages II, III and IV. Further investigation based on fieldwork in the area is needed in order to get a better understanding of the phenomena at hand.

- (19) [...] *tən igʷāddālān* ⇔ [...] *tən igʷāddālān*  
 [...] 3PL PTCP-IPFV] [...] 3PL PTCP-IPFV]  
 “un qui les abrite (habituellement)” “un qui était en train de les abriter”  
 [“one who habitually shelters them”] [“one who was sheltering them”]  
 “un qui les abrite (habituellement)”  
 [“one who habitually shelters them”]

5. Stage V is that of the Ifoghas Tuareg of Mali, which provides evidence of a new equipollent opposition with marked modal values instead of marked aspectual values, since both the habitual and the progressive are associated only with the innovated imperfective (i.e., non-modal values):

- (20) [...] *tən igʷāddālān* ⇔ [...] *tən igʷāddālān*  
 [...] 3PL PTCP-IPFV] [...] 3PL PTCP-IPFV]  
 “un qui est apte à les abriter” “un qui était en train de les abriter”  
 [“one who is capable of sheltering them”] [“one who was sheltering them”]  
 “un qui les abrite (habituellement)”  
 [“one who habitually shelters them”]

6. The last stage (Stage VI) has lost the modal opposition and is thus similar to the departure stage where one form – in this case, the innovated imperfective – conveys both aspectual values. This stage represents most of the present Tuareg languages:

- (21) [...] *tən igʷāddālān*  
 [...] 3PL PTCP-IPFV]  
 “un qui était en train de les abriter”  
 [“one who was sheltering them”]  
 “un qui les abrite (habituellement)”  
 [“one who habitually shelters them”]

#### 4.2 Innovations in the Tarifit verb

Compared to the other Berber languages, Tarifit (i.e., Central Tarifit) has a remarkable formation system for the imperfective. Two major characteristics are observed:

1. There are various innovated positive imperfectives associated with corresponding negative imperfectives, all standing for different semantic values which comprise axiological oppositions.
2. There is a morphological-semantic opposition between the positive imperfectives and the negative imperfectives within negation configurations.

These properties are at the basis of a considerable system enhancement that has brought its imperfectives to a total of six, compared to generally one or two imperfectives in the other Berber languages.

**Table 4.** The imperfectives of Central Tarifit (ex. *aɖəf* “to enter”, 3MSG)

Positive context	Negative context
<i>i-ttaɖəf</i>	<i>ur i-ttaɖəf</i>
–	<i>ur i-ttiɖəf</i>
<i>i-ttaɖaf</i>	<i>ur i-ttaɖaf</i>
–	<i>ur i-ttiɖif</i>

The thematic oppositions in question were developed from the progressive imperfective (imperfective type A, IPFVa; see Table 5) which, in Berber, is commonly constructed by means of the following morphological procedures:

- a gemination or tenseness of the first or second verbal radical;
- an affixation of the proclitic *tt-* (mainly before vowels) or *t-* (mainly before consonants);
- an insertion of a full vowel before or after the final consonant of the stem; i.e., vowel *a* or a repetition of the preceding vowel.

These morphological derivation devices are, however, subject to the following combinatorial restrictions:






- gemination/tenseness and prefixing of *(t)t-* are usually not combined;
- prefixing of *(t)t-* is mainly associated with vowel insertion;
- the combination of gemination/tenseness and vowel insertion is allowed in IPFVa for verbs of the /cc/ and /ss-icc/ type: e.g., *ǧəz* (AOR) ~ *qqaz* (IPFVa) “to dig” (see also Table 5).

Regarding the secondary imperfectives (IPFVb and IPFVc), these are derived from IPFVa and are developed by means of morphological combinations drawing on the same procedures used for IPFVa (see Table 5); that is:

- prefixing *(t)t-* to IPFVa (imperfective B, IPFVb);
- inserting the vowel *a* before the final consonant of the IPFVb stem (imperfective C, IPFVc).

As shown in Table 5, imperfectives b and c can only be formally derived from imperfectives which are not already supplied with the *(t)t-* prefix or the *a* vowel (in all stem positions). In such cases, the imperfective oppositions are neutralized. Most of the verbs concerned are transitive or labile (i.e., they show interchangeability of arguments; also called ‘symmetric’ verbs in Berberology). In the case of

Table 5. Positive imperfective oppositions in different verb types

Type	AOR	IPFVa	IPFVb	IPFVc	Translation
cc:c	<i>məlləg</i>	<i>tməlləg</i>	 IPFVa	= IPFVa	“to joke”
ac	<i>af</i>	<i>ttaf</i>	= IPFVa	= IPFVa	“to find”
cac:	<i>ḏarr</i>	<i>tḏarra</i>	= IPFVa	= IPFVa	“to harm”
cc:	<i>bədd</i>	<i>tbədda</i>	= IPFVa	= IPFVa	“to stand up”
cc (1)	<i>nəg</i>	<i>nəqq</i>	<i>tnəqq</i>	= IPFVb	“to kill”
cc (2)	<i>gəz</i>	<i>qqaz</i>	<i>tqaz</i>	= IPFVb	“to dig”
cca	<i>fna</i>	<i>fənna</i>	<i>tfənna</i>	= IPFVb	“to die”
ac:c	<i>azzər</i>	<i>tazzər</i>	= IPFVa	<i>tazzər</i> 	“to run”
acc	<i>aḏəf</i>	<i>ttəḏəf</i>	= IPFVa	<i>ttəḏəf</i>	“to enter”
ss-icc	<i>ssiwəḏ</i>	<i>ssakk<sup>w</sup>aḏ</i>	<i>tsakk<sup>w</sup>aḏ</i> 	= IPFVb	“to bring”
ccc	<i>ḏfəš</i>	<i>ḏəffəš</i>	<i>ḏəffəš</i> 	<i>ḏəffəš</i> 	“to fold”

a labile verb, it is its transitive dimension that is subject to the innovation process; e.g., *šfa* (AOR) (“to be pure”) ~ *šaffa* (IPFVa)/*tšaffa* (IPFVb)/*tšaffa* (IPFVc) (“to be pure”/“to purify”).

Furthermore, a number of verbs in Central Tarifit allow for the existence of even four positive imperfectives, such as, for instance, the verb *fṭər* (“to tress”, “to roll couscous”, “to wreath”).

Table 6. The verb “to roll couscous” and its four positive imperfective forms

“to roll couscous”	
Aorist	<i>fṭər</i>
Imperfective A	<i>fəttər</i>
Imperfective B	<i>tfəttər</i>
Imperfective C	<i>tfəttar</i>
Imperfective D	<i>fəttar</i>

However, these morphological oppositions do not usually occur in verbal forms of the /c:c/ type, where they might be most expected; e.g., the form *\*təkkas* is not attested as a variant of the regular forms *kkəs* (AOR) ~ *təkkəs* (IPFV) “to take away”. They are mainly found in verbs of the /ccc/ type, as exemplified in Table 6, which is the most generalized verbal template in Berber. In addition, the semantic values of the imperfective D are variable and hence difficult to pinpoint, although they generally convey notions similar to those expressed by means of the imperfective C (see Table 8 and corresponding examples). As regards the negative imperfective, these are the oppositions that correspond to the positive imperfectives presented in Table 5.



Table 7. Negative imperfective oppositions in different verb types

Type	AOR	NIPFVa	NIPFVb	NIPFVC	Translation
cc:c	<i>məlləg</i>	<i>təllig</i>	= NIPFVa	= NIPFVa	“to joke”
ac	<i>af</i>	<i>ttif</i>	= NIPFVa	= NIPFVa	“to find”
cac:	<i>ḏarr</i>	<i>tḏarri</i>	= NIPFVa	= NIPFVa	“to harm”
cc:	<i>bədd</i>	<i>tbəddi</i>	= NIPFVa	= NIPFVa	“to stand up”
cc (1)	<i>nəg</i>	<i>nəqq</i>	<i>tnəqq</i>	= NIPFVb	“to kill”
cc (2)	<i>gəz</i>	<i>qqiz</i>	<i>tqiz</i>	= NIPFVb	“to dig”
cca	<i>fna</i>	<i>fənni</i>	<i>tfənni</i>	= NIPFVb	“to die”
ac:c	<i>azzər</i>	<i>tizzər</i>	= NIPFVa	<i>tizzir</i>	“to run”
acc	<i>aḏəf</i>	<i>ttidəf</i>	= NIPFVa	<i>ttidif</i>	“to enter”
ss-icc	<i>ssiwəḏ</i>	<i>ssikk<sup>w</sup>iḏ</i>	<i>tsikk<sup>w</sup>iḏ</i>	= NIPFVb	“to bring”
ccc	<i>ḏfəṣ</i>	<i>ḏəffəṣ</i>	<i>tḏəffəṣ</i>	<i>tḏəffis</i>	“to fold”

Every full vowel *a* of the IPFV is replaced by *i* in the corresponding negative form (except /cac:/ verb type). Verbs ending with *\*ər*, e.g., *\*akər* “steal”, have no distinctive forms for the IPFV, because the opposition between [*ā* < *\*ər*] and [*ā* < *\*ar*] is neutralized here due to the vocalization of *r* (Lafkioui, 2006, 2007, 2009, 2011). In negation contexts, however, this opposition is marked:

- (22) *ur i-ttikā* (< *\*i-ttikər*) ~ *ur i-ttikī* (< *\*i-ttikir*)

From a functional perspective, both the original and the innovated imperfectives in Central Tarift express semantic distinctions which have not usually become possible as a result of thematic change in the other Berber languages, as is demonstrated in the examples (23)–(25).

- (23) *i-kənnəf* *izumbiyyən*.  
grill-IPFVa-3MSG corn-MPL  
“He is grilling corn cobs.”
- (24) *kur tṡəddiṡ(,)* *i-tkənnəf* *izumbiyyən*.  
every evening + INT grill-IPFVb-3MSG corn-MPL  
“Every evening, he grills corn cobs.”
- (25) *i-tkənnəf* *izumbiyyən*.  
grill-IPFVC-3MSG corn-MPL  
“He is used to grilling corn cobs.”

Table 8. Semantic distinctions of the imperfective in central Tarift

Imperfective A ⇒ progression, concomitance (example 23)

Imperfective B ⇒ iteration (example 24)

Imperfective C ⇒ habituation (example 25)

These morphological oppositions of the imperfective not only have a high functional productivity in Tarifit, but also indicate distinctive semantic values. (26) and (27) corroborate this and show how a morphological neutralization implies an axiological neutralization:




- (26) *i-ttaḡəm* *ssin.*  
 draw up water-IPFVa/IPFVb-3MSG from there  
 “He draws up water from there.”  
 “He is drawing up water from there.”
- (27) *i-ttaḡam* *ssin.*  
 draw up water-IPFVc-3MSG from there  
 “He is used to drawing up water from there.”  
 “He always draws up water from there.”  
 “He is always drawing up water from there.”


Another remarkable feature of central Tarifit is the use of positive perfectives (A, B, C) in negation structures after a negation marker such as *ur*, in order to create new oppositions. The precise semantic values of these innovated oppositions are difficult to define for the time being because of their variable nature, but their principal properties cover the values described in Table 9.

Table 9. Semantic distinctions of positive imperfectives in negation

Punctual verbs	
NEG+IMPF	⇒ habitual
NEG+NIMPF	⇒ non-habitual
Non-punctual verbs	
NEG+IMPF	⇒ durative (intensity)
NEG+NIMPF	⇒ non-durative

In (28)–(31), I present a series of examples, based on the punctual verb *aḡər* “to hang up”, that account for these formal-functional oppositions:

- (28) *ur i-ttaḡər*  *azḡab=inās.*  
 NEG hang up-IPFVa/IPFVb-3MSG djellaba=3MSG  
 “He is not used to hanging up his djellaba.”
- (29) *ur i-ttaḡar*  *zḡab=inās.*  
 NEG hang up-IPFVc-3MSG djellaba=3MSG  
 “He is not used to always hang up his djellaba.”
- (30) *ur i-ttiḡər*  *zḡab=inās.*  
 NEG hang up-NIPFVa/NIPFVb-3MSG djellaba=3MSG  
 “He does not hang up his djellaba.”  
 “He will not hang up his djellaba.”

- (31) *ur i-ttiḡir*  *zḡaḇ=inəs.*  
 NEG hang up-NIPFVC-3MSG djellaba=3MSG  
 “He does not hang up his djellaba every time.”

With non-punctual verbs, the combination [NEG + IPFV (positive imperfective)] commonly expresses duration and intensity, as in (32) and (33). Given the semantic nature of the non-punctual verbs, configurations conveying non-durative values, i.e., [NEG + NIPFV], are rather uncommon (34) or grammatically unacceptable (35). A verb like *asəm* “be jealous”, for instance, is mainly associated with positive perfectives, even when occurring in negation structures.

- (32) *u ttasm-ən* *ši zaḡ=s.*  
 NEG be jealous-IPFVa/IPFVb-3MPL NEG of=3SG  
 “They are not jealous of him.”  
 “They will not be jealous of him.”
- (33) *u ttasam-ən* *ši zaḡ=s.*  
 NEG be jealous-IPFVC-3MPL NEG of=3SG  
 “They are not really jealous of him/her.”
- (34) *u ttisim-ən* *ši zaḡ=s.*  
 NEG be jealous-NIPFVC-3MPL NEG of=3SG  
 “They are not always jealous of him/her.”
- (35) \**u ttism-ən* *ši zaḡ=s.*  
 NEG be jealous-NIPFVa/ NIPFVb-3MPL NEG of=3SG  
 “They are not jealous of him.”  
 “They will not be jealous of him/her.”

## 5. Conclusion

Given the geographical distance between the two areas in which the innovations discussed have occurred – mainly the Sahara for the Tuareg languages and northern Morocco for Tarifit (see Figure 1) – one can confidently infer that they were not engendered by contact between these languages. Nor do the examined data reveal any possible contact with other non-related languages spoken in the area, such as Arabic, Spanish, French or Hausa. Moreover, although Tuareg and Tarifit show a language-driven derivational mechanism engendering new variants and oppositions in the Berber verbal system, the precise patterns of change and corresponding reorganization of the paradigmatic structure of the verb are significantly divergent. The table below provides an overview of the major differences.

**Table 10.** Comparison between Tuareg and Tarifit verbal innovations

Tuareg (Mali)	Central Tarifit
Original imperfectives	Original imperfectives
↕	↕
Innovated imperfectives	Innovated imperfectives
Original perfectives	Original perfectives
↕	
Innovated perfectives	
Innovated imperfectives = Positive imperfectives	Innovated imperfectives = Positive imperfectives
	↕
	Negative imperfectives
1 new external derivation procedure = restricted to this use	Various new reanalyzed internal derivation procedures = new morphological combinations mark distinctive semantic values

This brings us to the following question: What motivated these system-internal changes in Berber? The data and analysis presented in §4 indicate that the innovations in question, and especially those related to the imperfective, were driven mainly by functional parameters.

Tarifit, in particular, displays a hyper-diversification of the morphological marking procedures through new derivations and combinations so as to express fine-tuned semantic distinctions of the imperfective. This language also allows for the innovated pattern [NEG + positive imperfective], which has provoked a number of reanalysis phenomena causing a restructuring of the verbal system. The new morphosyntactic pattern represents a twofold semantic distinction; it expresses habitual values in the case of punctual verbs and durative (intensity) values with non-punctual verbs.

Semantic and pragmatic motivations are also central to the creation of the new morphological variants and their development in Tuareg. Yet, most of the Tuareg languages have been the subject of a simplification of their verbal system (last stage) after several diversification stages, the penultimate of which has introduced certain modality values (Stage V, Adhagh Tuareg). The Tuareg imperfective seems to have followed a cyclical pattern, departing from a one-part aspectual configuration, evolving into a bipartite aspectual configuration followed by a bipartite aspectual/modal configuration, so as to end in a one-part aspectual configuration again. With the exception of the last stage, which may be the result of formal (i.e., grammatical) cohesion and generalization pressure, all developments have for the most part been functionally determined.

## References

- Basset, André. 1952. *La langue berbère*. London: Oxford University Press.
- Brugnatelli, Vermondo. 2002. "Les thèmes verbaux négatifs du berbère: quelques reflexions". *Articles de linguistique berbère. Mémoial Werner Vycichl*, ed. by Kamal Naït-Zerrad, 165–180. Paris: L'Harmattan.
- Cadi, Kaddour. 1987. *Système verbal rifain. Forme et sens*. Paris: SELAF.
- Chaker, Salem. 1989. "Aspect (verbe)". *Encyclopédie berbère*, 7.971–977.
- Chaker, Salem. 1996. "Quelques remarques préliminaires sur la négation en berbère". *La négation en berbère et en arabe marocain*, ed. by Dominique Caubet and Salem Chaker, 9–22. Paris: L'Harmattan.
- Galand, Lionel. 1974. "Introduction grammaticale". *Contes touaregs de l'Aïr*, ed. by Petites sœurs de Jésus, 13–41. Paris: SELAF.
- Galand, Lionel. 1977. "Continuité et renouvellement d'un système verbal: le cas du berbère". *Bulletin de la Société de Linguistique de Paris* 72.1.275–303.
- Galand, Lionel. 2002. *Etudes de linguistique berbère*. Louvain & Paris: Peeters.
- Heath, Jeffrey. 2005. *A Grammar of Tamashek (Tuareg of Mali)*. Berlin: Mouton de Gruyter. doi:10.1515/9783110909586
- Kossmann, Maarten. 1989. "L'inaccompli négatif en berbère". *Études et Documents berbères* 6.19–29.
- König, Crista. 2006. "Marked Nominative in Africa". *Studies in Language* 30.4.705–782. doi:10.1075/sl.30.4.02kon
- Lafkioui Mena. 2006. "La vocalisation des alvéolaires /r/ et /rr/ dans les variétés berbères du Rif". *Studien zur Berberologie/Etudes Berbères* 3, ed. by Dymitr Ibriszimow, Rainer Vossen & Harry Stroemer, 175–184. Köln: Köppe.
- Lafkioui, Mena. 2007. *Atlas linguistique des variétés berbères du Rif*. Köln: Köppe.
- Lafkioui, Mena. 2009. "Synchronic and Diachronic Linguistic Variation as an Indicator of Language Change and Diffusion in Tarifit". *Studi Magribini* (Nuova Serie) 7.65–84.
- Lafkioui, Mena. 2011. "How System-Internal Linguistic Factors Indicate Language Change and Diffusion: A Geolinguistic Analysis of Berber Data". *Dialectologia et Geolinguistica* 19.62–80. doi:10.1515/dig.2011.003
- Lafkioui, Mena. Forthcoming. "Dialectology and Linguistic Geography". *Handbook of African languages*, ed. by Rainer Vossen. Oxford: Oxford University Press.
- Lafkioui, Mena & Maarten Kossmann. 2009. "Les inaccomplis du Tarifit". *L'Amazighe dans l'Orient et le Nord du Maroc: variations et convergences*, ed. by Aïcha Bouhjar & Hamid Souifi, 101–108. Rabat: Editions IRCAM.
- Leguil, Alphonse. 1987. "Les quatre inaccomplis du touareg de l'Adagh". *Proceedings of the Fourth Hamito-Semitic Congress (Marburg, 20–22 September 1983)*, ed. by Herrmann Jungraithmayr & Walter W. Mueller, 395–406. Amsterdam: John Benjamins. doi:10.1075/cilt.44.22leq
- Leguil, Alphonse. 2000. "Une opposition fluctuante en touareg". *Etudes berbères et chamito-sémitiques. Mélanges offerts à Karl-G. Prasse*, ed. by Salem Chaker and Andrzej Zaborski, 257–262. Paris & Louvain: Peeters.
- Mammeri, Mouloud. 1980. *Poèmes kabyles anciens*. Paris: Maspero.
- Matthews, Stephen J. 1990. *A Cognitive Approach to the Typology of Verbal Aspect*. Los Angeles: University of Southern California dissertation.

- Picard, André. 1959. "Du preterit intensif en berbère". *Memorial André Basset (1895–1956)*, 107–120. Paris: Maisonneuve.
- Prasse, Karl-G. 1973. *Manuel de grammaire touarègue. Volume: Verbe*. Copenhagen: Akademisk Forlag.
- Prasse, Karl-G. & ägg-Älbostan ägg-Sidiyän, Ekhya. 1985. *Tableaux morphologiques. Dialecte touareg de l'Adrar du Mali (berbère)*. Copenhagen: Akademisk Forlag.
- Schmid, Maureen A. 1980. *Co-occurrence Restrictions in Negative, Interrogative, and Conditional Clauses: A Cross-Linguistic Study*. Buffalo, NY: SUNY Buffalo dissertation.
- Van den Boogert, Nico. 1997. *The Berber Literary Tradition of the Sous – With an Edition and Translation of the 'Ocean of Tears' by Muḥammad Awzal (d. 1749)*. Leiden: de Goeie Fund.
- Vycichl, Werner. 1953–1955. "Der Umlaut in den Berbersprachen Nordafrikas. Eine Einführung in die berberische Sprachgeschichte". *Wiener Zeitschrift für die Kunde des Morgenlandes* 52.304–325.

**PAGE PROOFS**  
© JOHN BENJAMINS PUBLISHING COMPANY

**2nd proofs**

# Condition, interrogation and exception

## Remarks on particles in Berber

Catherine Taine-Cheikh

LACITO (CNRS – Universités Paris III Sorbonne Nouvelle – INALCO)

Comparison of the different Berber dialects shows that there are many shared characteristics, particularly typological ones. The degree of similarity varies, however, depending on the element. Conjunctions have been recognized as a class with very little similarity among the varieties.

This diversity is indeed found in the inventory of those Berber particles which introduce the protases of hypotheses. However, the number of markers is much lower if one takes into account the probable origin of the basic units these markers are built upon. From this perspective, we examine all of these particles' uses and meanings and find true convergences.

Convergences have a spatial dimension. They are also to be found in the meanings of the particles. Particles which introduce the protasis of a conditional are often closely linked to, on the one hand, those which introduce interrogative clauses and, on the other, those used to express exceptions.

**Keywords:** conditional particle, Berber, grammaticalization path, question, exceptionality

Comparison of the different Berber dialects (whether they be considered languages, dialects or local varieties) shows that there are a great many shared characteristics, particularly from a typological perspective. The degree of similarity varies, however, depending on the element studied. Conjunctions have been recognized as one of the least unified of elements:

Le système des conjonctions varie notablement d'un parler ou d'un dialecte (ou langue) à l'autre, ce qui permet de les considérer comme un outillage relativement récent, que chacun des groupes berbérophones a élaboré pour son propre compte. À côté d'éléments berbères, on y reconnaît souvent des emprunts à l'arabe.<sup>1</sup>

(Galand 1988: 225)

---

1. "The conjunction system varies noticeably from one local variant or dialect (or language) to the next. This indicates that they are a relatively recent tool, which each of the Berber-speaking



This diversity is indeed found in the inventory of those Berber particles which introduce the protases of hypotheses. However, the number of markers is much lower if one takes into account the probable origin of the basic units these markers are built upon. From this perspective, it is necessary to examine all of these particles' uses and meanings. This examination will focus on observed convergence in order to shed light on grammaticalization pathways.

### 1. *ad* in Zenaga

Zenaga is an endangered Mauritanian language. It is generally classified as a specific subgroup within Berber. In this peripheral variety (located in the south-westernmost part of the Berber area) the uses of *ad* vary widely, as in other Berber varieties, and they show some specificities. For example, Zenaga is the only Berber variety where *ad*<sup>2</sup> has a fully grammaticalized use as a conditional particle ("if").

The various occurrences of Zenaga *ad* are examined in Taine-Cheikh 2010a where several possible grammaticalization pathways are laid out, based on uses of *ad* as a deictic, for instance as a demonstrative and copula.

The following examples provide an overview of the various uses of *ad* in Zenaga.

- a. *ad* is a proximal demonstrative pronoun (MSG *äd* vs. FSG *täd* vs. MPL *ädniä*), a demonstrative clitic (SG *-äd* vs. PL *-id*) and an invariable copula (with non-adjectival nominal predicates). These three possibilities are illustrated in (1).

- (1) *äd* / *äräbiy=äd*                      *äd*    *änähtəf*  
          this.one / M.child.SG=this.SG COP    M.wary.SG  
          "This one / this child is wary."

- b. *ad* is an invariable particle used to introduce independent clauses expressing orders, prohibitions or injunctions
- c. *ad* is an invariable particle used to introduce dependent clauses in second position. The latter may be:
  - i. completive clauses (verb in the aorist) of governing verbs expressing orders, requests or wishes (expressed to others or to oneself);
  - ii. indirect speech (verb in the aorist), e.g., indirect orders, following verbs such as "say" or "ask";
  - iii. quotatives (verb in the perfective or imperfective) following the verb "say";<sup>3</sup>

---

groups created independently. Alongside Berber elements, one often finds borrowings from Arabic" (translation by Margaret Dunham).

2. /ad/ is generally pronounced [æd], with an interdental consonant, hence its transcription *äd*. Furthermore, for simplicity's sake, assimilations are generally not transcribed (only marked by ^).

3. This use of *ad* is only found in Zenaga (as is the case for the two following uses).

- iv. indirect polar questions following verbs such as *yäžgäD'äh* “ask, question, interrogate”, *yuzza'ɾ* “look”, *wär yəssən* “not know”, *yäznäzgām* “think, seek to understand”.

In the case of indirect questions, *äd* must be followed by a verbal auxiliary frozen in the 3MSG person: either *yu(u)gä* “become, unfold (time)”, or (more rarely) *yumrä* “already be”.

- (2) *wär Sən-äg äd y-u(u)gä y-ässhäh^däh*  
 NEG know.PFV-1SG if 3MSG-become.PFV 3MSG-come.PFV=PROX  
 “I do not know if he came.”

When the verb in the subordinate clause is in the imperfective – and not in the perfective as in (2) – it is usually preceded by the future auxiliary *yänhäyā* (... *äd^y-u(u)gä y-änhäyā y-ättässhäh^däh* “... if he will come”).

- d. *ad* is an invariable particle used to introduce protases of conditional clauses.  
 i. When *äd* is used on its own (+ verb in the aorist, with or without negation), it is used as a conditional (the ‘simple’ condition, which often shows a tendency to conflate with the temporal in “when, every time that”):

- (3) *äd wär^ y-äšbi y-uffud*  
 if NEG 3M.SG-drink.AOR 3M.SG-be.thirsty.AOR  
 “If (when) he does not drink, he is thirsty.”

- ii. When *äd* is followed by the frozen form *yu(u)gä*, the sentence takes on the meaning of a factual, referential hypothesis, used to express an alternative:

- (4) *äd^ y-u(u)gä y-äsbä*  
 if 3M.SG-become.PFV 3M.SG-drink.PFV  
*wär^ y-uffud*  
 NEG 3M.SG-be.thirsty.NEG.PFV  
 “If he drank, he is not thirsty.” [but If he did not drink, he is thirsty]

Within the irrealis system, one thus sees that *äd* covers a wide array of uses, where only counterfactual hypotheses are impossible (the particle (*h*)*äm* is used instead).

Similar to the *äd* used to introduce injunctions, prohibitions, wishes, requests, or even simply speech, the *äd* used in conditionals probably stems from a deictic (in a broad sense). Moreover, the *äd* used in protases can be more specifically used as an actualizer, where the clause is used to express fictional actualization (setting the clause in the referential sphere of all possible worlds).<sup>4</sup> From that perspective, the conditional *äd* may be more closely linked to injunctive *äd* and polar question *äd*.

4. The set protasis-apodosis then functions as a pairing of the type topic-commentary (cf. Taine-Cheikh 2011: 390–393).

This possibility of a tighter link between conditions and questions has only one visible manifestation in Zenaga, which is the widespread – and highly specific – use of the frozen verb form *y-u(u)gä*. In other Berber varieties certain factors seem to indicate that the systems are evolving in similar ways.

Table 1. The particle *ad* in Zenaga

	conditional	indirect question	deictic/ presentative	complementizer	injonction/ wish
<i>ad</i>	+	+	+	+	+
<i>ad</i> <sup>^</sup> <i>y-u(u)gä</i>	+	+	–	–	–

2. *is* in Tamazight and Tashlhit

Tamazight<sup>5</sup> and Tashlhit refer to two large dialect areas in central and southern Morocco, corresponding, respectively, to the Middle-Atlas and to Soûs. The particle *ad* is found there, but the invariable morpheme *is* is used in a great variety of contexts, for example in questions and, to a more limited extent, in conditionals.

- a. *is* is regularly described as being used to introduce independent or content questions. In this use, *is* is never mandatory (“Une intonation montante suffit à donner à l’énoncé chleuh une valeur interrogative” Galand 1988: 222),<sup>6</sup> but its presence does trigger affix displacement (such as the pronominal or adverbial *-t* in (6), used for spatial orientation).<sup>7</sup>

Consider this example from Tashlhit:

(5) *is i-ša                      u-srdun?* (Galand 1988: 222)<sup>8</sup>  
Q 3M.SG-eat.PFV M.AS-mule.SG  
“Did the mule eat?”

5. The term Tamazight is used here in its narrow sense.

6. “Rising intonation is sufficient to indicate a question in Tashlhit” (translation by Margaret Dunham).

7. Displacement is usually triggered by the presence of verbal modality (e.g., negation) or of a particle in a subordinating function. It can also happen, however, with sentence modalities such as injunctive *ad*.

8. For most of the examples given, I am responsible for the transliteration and, with Margaret Dunham, for the translation into English – and, in some cases, for the morphological segmentation. I have also attempted to standardize the transcriptions (thus the palatal fricatives are always transcribed by *š* and *ž*, the voiced pharyngeal fricative by *ʕ* and the post-velar fricatives by *g* and *h*).

- (6) *is t t-ut-t?* (Destaing 1920: 261)  
 Q PR.OBJ.3M.SG 2-hit.PFV-SG  
 “Did you hit him?”

In Tamazight:

- (7) *is t-kerz-d assa?* (Taifi 1991: 607)  
 Q 2-plow.PFV-SG today  
 “Did you plow [the field] today?”

- (8) *iz=d i-dda ass-anna?* (Penchoen 1973a: 82)  
 Q=PROX 3M.SG-come.PFV yesterday  
 “Did he come yesterday?”

- b. In contrast, *is* is obligatory when it introduces an indirect question. Example (11) shows that these two uses of *is* (as a direct and an indirect question marker) can coexist in a single sentence.

Tashlhit (Destaing 1920: 261):

- (9) *i-seqsa t is i-herš*  
 3M.SG-ask.PRIV PR.OBJ.3M.SG if 3M.SG-be.sick.PFV  
 “He asked him if he was sick.”

Tamazight (Taifi 1991: 607):

- (10) *ʔlu is da y-snufus*  
 see.IMP.SG if LOC 3M.SG-rain.IPFV  
 “See if it is raining.”
- (11) *is te-ssen-d is i-zzenza' a-gmar=n=s?*  
 Q 2-know.PFV-SG if 3M.SG-sell.PFV M.FS-horse.SG=of=PR.POSS.3SG  
 “Do you know if he sold his horse?”
- (12) *sal=t is i-lla gma=s*  
 ask.IMP=PR.OBJ.3M.SG if 3M.SG-be.PFV M.AS.brother.SG=PR.POSS.3SG  
*g taddar-t*  
 in F.AS.house-SG  
 “Ask him if his brother is in the house.”

- c. *is* is also found following certain verbs and serves as a complementizer, introducing the governed clause. The following examples come from Tamazight (Taifi 1991: 607):

- (13) *ǧal-ḥ is=k y-aǧ ša*  
 think.PFV-1SG COMP=PR.OBJ.2SG 3M.SG-reach.PFV something  
 “I thought you were sick.”
- (14) *i-ssen is i-ga ta-zennir-t*  
 3M.SG-know.PFV COMP 3M.SG-do.PFV F.FS-serious.offense-SG  
 “He knows he has committed a serious offense.”

The list of verbs which take *is* shows some variation, including within a single local variety. Thus Bentolila states (1981: 303 et seq.), on the subject of Aït Seghrouchen Tamazight, that *is* is frequently, but not obligatorily, found after operant verbs such as *ẓr* “see”, *af* “find”, *ẓRb* “experience”, *duNa* “believe”, *sL* “hear”, and *ašy* “smell” but that it can be practically obligatory, especially in sentences containing the governing verb *ini* “say” (as in (15)) or *isin* “know” (as in (16)), particularly when the subordinate clause contains a non-verbal predicate (as in (17)).

- (15) *is t-Ni-d is i-La ša*  
 Q 2-say.PFV-SG COMP 3M.SG-be.PFV something  
*Lbas?* (Bentolila 1981: 306)  
 bad

“Do you think that there is something bad?”

- (16) *Tu-ḥ is i-rah*  
 forget.PFV-1SG COMP 3M.SG-leave.PFV  
*u-rba* (Bentolila 1981: 307)  
 M.AS-child.SG

“I forgot that the child has left.”

- (17) *Sn-ḥ is ur dis ša* (Bentolila 1981: 332)  
 know.PFV-1SG COMP NEG there something  
 “I know that there is nothing.”

In Ayt Ndhir Tamazight, verbs built on *is* are of the type *isin* “know”, *ini* “say, tell”, *annay* “see”, *səl* “hear”, *iẓir* “be cognizant” (Penchoen 1973a: 73):

- (18) *yun=w-ass i-ttuna=yas is i-lla*  
 one=M.AS.day 3M.SG-hear.PRV=PR.OBL.3SG COMP 3M.SG-be.PRV  
*yul=lkanz...* (Penchoen 1973a: 85)  
 one=treasure

“Now one day he heard that there was a treasure...”

Galand (1988: 225), who usually draws his examples from Igchan Tashlhit, highlights the differences in meaning induced by the use or not of *is* following operant verbs such as *ini* “say” and *isin* “know”.

- (19) *Ni-ḡ / is i-Qn imi*  
 say.PRV-1SG COMP 3M.SG-be.closed.PRV M.AS.door.SG  
 “I thought the door was closed.”

as opposed to

- (19') *Ni-ḡ i-Qn imi*  
 say.PRV-1SG 3M.SG-be.closed.PRV M.AS.door.SG  
 “I said the door was closed.”

- (20) *Sn-ġ is y-ara*  
 know.PRV-1SG COMP 3M.SG-write.PRV  
 “I know he wrote.” [... it is that ...]

as opposed to

- (20') *Sn-ġ ad ara-ġ*  
 know.PRV-1SG POT write.AOR-1SG  
 “I can write.”

These examples show that *ini* “say” takes on the meaning of “think” when it is followed by *is*, whereas *isin* “know” retains its full verb meaning with *is* but takes on modal ability meaning when the construction contains the particle *ad*. For Galand, the *is* in (19) and (20) is not a true subordinator but is rather a particle made up of the determiner prop *i* and the preposition *s*.<sup>9</sup> Thus the original construction would be better rendered by the explanatory phrase “it is that”.

- d. Presentative particle usage, found e.g., in Tamazight, can be illustrated by Examples (21)–(23), taken from Taifi’s *Dictionnaire* (1991: 607–608).

- i. It usually introduces clauses with explanatory or causal meaning (“it is that”, “because”):

- (21) *meš ur t^ t-ufi-d asekka*  
 if NEG PR.OBJ.3M.SG=2-find.NEG.PRV-SG tomorrow  
*ha=t is i-safer*  
 here.is=PR.OBJ.3M.SG because 3M.SG-travel.PRV  
 “If you do not find him tomorrow, it means that he traveled.”

- (22) *ur i-ri ad^ d=i-ddu*  
 NEG 3M.SG.will.NEG.PRV POT=PROX=3M.SG-come.AOR  
*is i-gg^ed*  
 because 3M.SG-be.afraid.PRV  
 “He does not want to come because he is afraid.”

- ii. At times it also introduces clauses, having the meaning “as for”. In this case, it is usually accompanied by the predicative particle *d* and has the form *idd/id* (< *iz-d* < *is-d*).

- (23) *id nekk, ša wr tannay-h*  
 as.for me something NEG see.IPRV-1SG  
 “As for me, I have not seen anything.”

- e. Lastly, the particle *is* is also found in hypotheses.

- i. *is* is used relatively rarely as a conditional particle, nonetheless the Tamazight *Dictionnaire* (Taifi 1991: 607) provides several examples.

9. For further information on the origins and explanatory values of *is*, see the article by Galand (2002 [1987]: 241–256), and more particularly pages 249–253.

- (24) *is t-ri-d*                      *ad^ t-eddu-d*                      *ger ssuq,*  
 if 2-want.PRIV-SG POT=2-go.AOR-SG to M.souk.SG  
*kker*                      *zikk*  
 get.up.IMP.SG early  
 “If you want to go to the souk, get up early.”

In contrast, Bentolila (1981: 310) maintains that in Aït Seghrouchen Tamazight, hypothetical meaning is rendered through inversion (like in the two verses of La Fontaine’s fable: “Se mire-t-on près d’un rivage / Ce n’est pas soi qu’on voit”).<sup>10</sup> In (25), the initial *is* indeed appears to function in the same way as the interrogative adverb *is*:

- (25) *is i-La*                      *u-zǧal,*                      *La TKs-n-t*  
 Q 3M.SG-be.PRIV M.AS-good.weather.SG HAB take.away.IPRV-3PL-F  
*t-<sup>1</sup>Yalin*                      *timŠin*                      *i u-ḥam*  
 F.AS-woman.PL F.FS.timshin.PL to M.AS-tent.SG  
 “Is the weather beautiful? Women are taking the *timshins*<sup>11</sup> from the tent.”

Therefore the grammaticalization of *is* as a conditional marker appears limited, even within Tamazight.<sup>12</sup> However, given that the difference in analysis for (24) and (25) is mostly due to the difference in intonation, it is easy to see how the language may have gone from free association between two independent clauses to a single sentence with two linked clauses.

- ii. In Tuareg, the particle *as*, for which the etymology and usage in relatives are comparable to that of *is*, has undergone specific changes in the southern varieties. In these varieties it is used both as a complementizer (following verbs such as *issân* “know”) and as an adverbial subordinator. The particle, which literally means “the situation entailing (such and such a thing)”, can in certain contexts take on temporal meaning (“when”), causal meaning (“given that”), or hypothetical meaning (“if”). The various stages in this grammaticalization process are well documented in Galand (2002

10. “Do we gaze in a crystal stream? This not our selves we survey” (translation by Margaret Dunham).

11. *TimŠin* was not translated by Bentolila, but this lexeme is comparable to *timšušin*, the plural of *tamšušt* which Taifi (1991: 438) translates as “old threadbare mat”.

12. In Ighchan Tashlhit, the conditional particle is *iǧ* “if, when” (virtual). The origins of this particle are similar to those of *is*: see *i s* “the one with” and *i ḡ* “the one in” (cf. Galand 1988: 226). There is also *mtaD* is “if” (irrealis), but in this case the presence of *is* probably denotes a lesser degree of specificity than *mta(D)*.

[1987]). Although these uses are quite different from what one finds for *is*, it is interesting to note that in Tamasheq of North-East Burkina Faso, “[t]he conjunction *as* is frequently used to express ‘if’ or ‘when’” (Sudlow 2011: 95). For example as in (26):

- (26) *as* *āmtäll-ăğ* *sasseğədăd=ahi*  
 if make.a.mistake.PR.V-1SG correct.IMP.SG=PR.OBJ.1SG  
 “If I make a mistake, correct me.”

- f. In Tamazight and Tashlhit, exceptions are often expressed by the prepositions *ğār* and/or *(a)bla* (cf. Destaing 1920: 256; Bentolila 1981: 217; Taifi 1991: 17). These originally Arabic particles clearly show no relation to *is/as*. On the contrary, the Zenaga preposition *ās* “with the exception of, except” (Taine-Cheikh 2008: 470; 2011: 546) is akin to *is/as*. The difference in frication is simply the regular reflex of the change *s>š* (which is why in Zenaga the complementizer for example is *äyš* and not *is/as*).

In support of this posited kinship is Bentolila’s study of oaths (1988). He shows that the particles which appear in the positive real verb phrases of oaths often had the original meaning “only”. This is true not only of *abla* and *(a)ğār/ğir*, but also of *has* and *hs*, particles ending in *-s* found in central Morocco, amongst, respectively, the Zemmours and the Ait Sadden (Bentolila 1988: 52–53, 64).

In Zenaga, one finds a particle *ās* (which is different from but similar to the exception particle *ās*) used in a positive oath concerning a future event (Taine-Cheikh 2010b: 202–203). Its close correspondence to central Moroccan *has* strengthens the kinship hypothesis between the various *s/š* particles. In this context, a relation between the above mentioned particles *is/as* and the restrictive particles becomes entirely plausible.

Table 2. Particles with *-s/-š* in Berber

	conditional	direct & indirect question	exception	deictic/ presentative	comple- mentizeur	temporal subordinator
<i>is</i> Morocco	(+)	+	–	+	+	–
<i>as</i> touareg	+	–	+	–	+	+
<i>ās</i> zénaga	–	–	+	–	–	–



### 3. *m(a)* in the northern varieties

Chamito-Semitic shows a varied use of particles built upon the basic element *m*-<sup>13</sup> – e.g., in interrogatives and indefinites. This phenomenon is also found in Berber, particularly in interrogative pronouns (“who?”, “what?”) and adverbial ones (“where?”, “when?”, “how?”, “why?”), with a tendency to spread to conditional particles in certain varieties.

#### 3.1 The polyfunctional *ma*

It is in the northern varieties, especially in Morocco and Algeria, that the uses of the invariable particle *ma* have diversified the most. It is not uncommon that it be used to introduce both interrogatives (direct and indirect) and the protasis of conditionals.

- a. *ma* is an interrogative adverb, it is sometimes in competition with *is*, but is more widely found outside of Tashlhit and Tamazight.

- i. Above (in 2.a), we did not provide any examples of *is* as an interrogative adverb in Aït Seghrouchen Tamazight. This is not because *is* never has the meaning of an interrogative in this variety but rather because, strictly speaking, the interrogative adverb is not *is*, but *ma*.<sup>14</sup> Only *ma* can introduce polar questions, to which a yes/no answer is possible, as in (28):

(28) *ma d ʔli?*  
 Q PRED Ali  
 “Is it Ali (or not)?”

- ii. Map 292 of the *Atlas des variétés berbères du Rif* (Lafkioui 2007: 240) shows that the interrogative *ma* is found in a great many varieties (both eastern and central). In eastern Rifain, Kossmann (2000: 179) notes that it is more common to use *ma* than to use only intonation:

(29) *ma t-əlli-d mliḥ?*  
 Q 2-be.IPRV-SG well  
 “Are you well?”

13. For Semitic, see Faber (1991) and Lipiński (2001: 336–368, 467, 480, 546).

14. Bentolila (1981: 188–189) specifies on the one hand that one does find the combination *ma-is* (only in that order) and, on the other hand, that with *is*, it is a simple request for confirmation, as in (27), where the primary meaning of “is it that ...” can still be clearly felt.

(27) *iz d ʔli?*  
 “is it really Ali?”

iii. The interrogative *ma* is also used in Aures Chaouia:

- (30) *ma i-lla*                      *ša*                      *n<sup>w</sup>=wa*  
 Q    3-be.PRIV.SG something of=PR.DEF.M.SG  
*t-ssn-d* (Penchoen 1973b: 55)  
 2-know.PRIV-SG  
 “Do you know something about it?”

- (31) *ma i-lla*                      *qli*                      *m b<sup>o</sup>eqqi?* (Lafkioui & Merolla 2002: 52–53)  
 Q    3-be.PRIV.SG a.little of love.SG  
 “Is there a little love?”

iv. Lastly, *ma* is found in Kabyle (where it does not trigger clitic displacement):

- (32) *ma t-efra*                                      *dde<sup>s</sup>wa?* (Dallet 1982: 475)  
 Q    3F.SG-be.solved.PRIV matter.F.SG  
 “Is the matter solved?”

b. The use of *ma* for indirect polar questions is found in the same varieties (with the apparent exception of Tamazight).

i. In eastern Rifain (Kossmann 2000: 180):

- (33) *ʕabbər*                      *ma a*                      *lam=d=y-as*  
 measure.IMP.SG if    POT    PR.OBL.2F.SG=PROX=3M.SG-go.AOR  
 “See if it fits you!”

ii. In Aures Chaouia (Penchoen 1973b: 54):

- (34) *t-ra<sup>ʕ</sup>a*                      *ma dag=š*                      *ša*                      *n<sup>y</sup>=h.brar*  
 3F.SG-look.PRIV if    in=PR.3SG something of=lumps  
 “She checks if it (the milk) contains lumps.”

iii. In the Algerian Chenoua dialect (Cherchell region)

- (35) *k'abel*                      *ma tsag-en*<sup>15</sup> (Laoust 1912: 77)  
 look.IMP.SG if    IPRV.rain-3PL  
 “Check if it rains.”<sup>15</sup>

iv. In Kabyle:

- (36) *Ur zri-ğ*                                      *ara*                      *ma ad*  
 NEG know.NEG.PRIV-1SG NEG.2 if    POT  
*y-eddu* (Nait-Zerrad 2001: 147)  
 3M.SG-come.AOR  
 “I do not know if he will come.”

15. *tsagen* [implied *ouaman*] “it is raining” – literally “get caught [waters]”, cf. Laoust 1912: 138.

- i. In Aures Chaouia, where *ma* “if” and *ma* “what” are homonymous with the negative particle *ma*:

(37) *ma=u=g̣ar=sn=š labas*  
 if=NEG=tō=PR.3PL=NEG2 much  
*nʸ=surɖggn* (Penchoen 1973b: 55)  
 of=M.AS.money.PL  
 ‘If they do not have much money [...]’

- ii. In Kabyle where, on its own, *ma* can take on the meaning “if” when it is followed by the perfective, as in (38):

(38) *ma y-ehwa=yak* *has*  
 if 3M.SG-like.PRV=PR.OBJ.2SG only  
*at=t-egred* (Dallet 1982: 476)  
 POT=2-read.AOR.SG  
 “If you like it, you can read it.”

Furthermore, it must be noted that one also finds longer forms in Kabyle: *ma d ay* and *ma d ara* (Naït-Zerrad 2001: 145).

### 3.2 $m(a)$ derivatives

Chaouia very clearly shows the three cumulative uses of *ma* (direct and indirect questions + condition), as does Kabyle, to a lesser degree. This is less the case in other varieties where the particle *ma* often needs additional material to be used as a conditional. Moreover, it is important to note that many varieties only use the element *m(a)* as a constitutive part of larger pronominal, adverbial or conjunctive units. In the case of conditional particles, it can be very difficult to identify the origin of all of the elements, but the presence of *m(a)* is as recurrent in conditionals as it is in units with interrogative meaning.

- a. *mer* and its variants. In Berber, for irrealis hypotheses, specific particles are used, e.g., *mer/mur/mr* in Tamazight and *mer/mmer/lemmer* in Kabyle. The fact that they are usually followed by the negative perfective strengthens Taifi's proposed etymology (1991: 426, 1993: 218–219) which posits that in *m(e)r/mur* one can detect the presence of the negation *ur/wr*.<sup>16</sup>

16. Its use for the potential, as in Kabyle (cf. Dallet 1982: 511), thus appears to be a secondary expansion.

- b. *mata, mala, mara* and their variants. Generally speaking, variation consists in the presence or not of a vowel following the nasal *m-*, and possible gemination of the second consonant. In some dialect areas, the second consonant varies but the overall pattern appears invariable (this is said to be the case in Soûs, where one finds *m̥la, m̥ra, m̥ta* “if”). In other areas, e.g., in the Rif, the pattern shows variation, but the second consonant is usually a liquid: *mala, malla, may, mara* (Lafkioui 2007: 230).<sup>17</sup> Lastly, in certain cases, the dental variant is dominant: *mta/metta* in Figuig,<sup>18</sup> *matta* (or *batta*) in Ouargla.

Furthermore, it is probably necessary to consider that there are at least two distinct formations. Even though in Ouargla the conditional particle *matta* does not require verb satellite displacement, in contrast to the indefinite *matta* (Delheure 1987: 200), etymological kinship is not impossible. As for the liquid consonant particles, their origin could be “la particule *ma* suivie de la forme verbale *yalla* ‘il est’”,<sup>19</sup> as suggested by Kossmann (2000: 199) for eastern Rifain, where the variant *mayalla* coexists alongside *malla*.

- c. *maka* and its variants will be studied in the next section of this paper.

Table 3. Particles with *m-* in Berber

	conditional	direct question	indirect question
<i>ma</i>	+	+	+
<i>m(e)r, m(a)ta, m(a)la...</i>	+	–	–

#### 4. Variants to the element *k(a)*

In many ways, the element *k(a)* appears to be a variant of *m(a)* (and even sometimes of *is*), even though, within the conditional system, it mostly appears as a second element added to *m(a)*, as in *maka*.

Like *m(a)*, *k(a)* is an element found in Semitic with negative and interrogative meaning (see Faber 1991: 414).<sup>20</sup>

17. See also *mala* in the Beni-Salah dialect, spoken on the slopes of the Blida Atlas, not far from Algiers (Laoust 1912: 77).

18. It is important to note that in Figuig *m(et)ta* is also used to link an operant verb to the following clause if that clause expresses doubt or uncertainty (Kossmann 1997: 322).

19. The particle *ma* followed by the verb form *yalla* “it is”.

20. For Lipiński (2001: 483), *ka-* is first and foremost a “deictic and asseverative particle”.

In Berber, *k(a)* may be used on its own, or it may constitute part of various particles, often with interrogative meaning.

- a. The use of *ka* in polar questions is rare but not unknown: it is found in western Rifain even though most of that dialect area uses *ma* (Lafkioui 2007: 237; 240, map 292).

In Figuig, the particle *waš* serves to introduce direct questions. It is also used for indirect questions governed by certain verbs: *t-eqqel waš...* “she was looking to see whether...” (Kossmann 1997: 322). If one allows that *waš* can be segmented into two elements, *wa*-<sup>21</sup> and *-š*, the second element could be analyzed as a representative of *k(a)*. This would be further proof of this particle’s wide array of uses.

In Zenaga, the element *k* appears in the global formulaic question:<sup>22</sup> *k-äyd? taʔK-äyd?* “what is/what did?”, where one recognizes the interrogative *taʔK* “what?” followed by the demonstrative *äyd* (Taine-Cheikh 2008: 293).

Also in Zenaga, the element *k* appears in the reinforcing discourse marker *äk/äK* (one must note in this case however the presence of the initial vowel *ä*-):

- (39) *nɪʔK äk aɖmā-g äyš ...*  
 I indeed think.PRIV-1SG COMP  
 “As for me, I think that...”

In Ouargla one finds uses comparable to those of Zenaga *äk(k)* and more particularly to those of Tamazight *is*. Delheure (1987: 137, 182) indicates that in Ouargla there are two particles which can be used with the presentative meaning “as for X”: both *akk* and *ammʷa*.<sup>23</sup>

- b. Parallel to these uses which tend to serve to structure information and for modality, the element *k*- also serves as a component in indefinites having the meaning “each” (e.g., Tuareg *ak* and Zenaga *äkki*) or “every” (e.g., Kabyle *akʷ*). When coupled with the particle *m(a)*-,<sup>24</sup> it provides various expressions. They can be interrogative as in the Tahaggart pronoun *manëkk* “what [is]?” or manner

21. One should note that Figuig *wa*, which is a component element of *wala* “even ..., ... as well”, is also a shortened form, as is *la* (Kossmann 1997: 344–345).

22. In this case it is often preceded by the element *taʔ* which raises the question of whether this dental consonant element is akin to Figuig and Ouargla (*ma*)*ta*.

23. *ammʷa* is considered a borrowing from Arabic, but this does not exclude parallel developments. Cf. the presence of *amma* in western Rifain (Kossmann 2000: 170) and more importantly of *uma* in Tamazight (Taifi 1991: 398).

24. This can however also be the case for the indefinite, cf. *makk* (*ma akk* “that which all”) “each” in Ouargla (Delheure 1987: 137).

phrases as in the Tawəlləmət lexeme *əmmək* “manner of being, nature ...”. They can also be both at once, as in the adverb “how?” which is expressed by *amek* in Kabyle, *mək/amək/mam(ə)k* in Ouargla, *maka/maša* in Tamazight and *maneš* in Figuig.<sup>25</sup>

- c. It is another coupling between *m(a) + k(a)* which serves as the conditional particle in the various Berber varieties. Although the apodosis particles seem to be akin to the others,<sup>26</sup> the forms remain distinct: voicing and/or a more marked tendency to palatalize *k* (>š) locally create this difference.

In Tamazight for example, the potential and realizable hypothetical particle is *mek* and more crucially *meš* (Bentolila 1981: 318–319; Taifi 1991: 414) – a form which is both similar to and distinct from *ma* “is it that?” and *maka/maša* “how?”<sup>27</sup> Inversely, in Ouargla, the reduced vowel form *mək* means “how?” while the form containing the vocalism *-a-a* has the meaning “if, in the case that” (Delheure 1987: 187):

- (40) *maka d nətta, uhu*  
       if     PRED him no  
       “If it is him, I refuse (no).”

We will return shortly to a possible etymology for *makan*, the second variant found in Ouargla, even if the origin of *kagella* “if”, to which we shall now turn, appears highly similar to that of *makan*. *kagella* is a variant of *kan* noted (along with *ouilla*) by Laoust (1912: 77) in the Algerian Chenoua variety. It could be composed along the same lines as *mayalla* and essentially only differ in its replacement (before the existence verb *yalla*) of *ma* by *ka*.

Based on this hypothesis, let us now consider the case of the preposition *gayr/gir* “except”. This preposition is a borrowing from Arabic, and it can be used as a preposition, as in Ouargli (Delheure 1987: 249). In Chenoua, in contrast, the exception particle is *gerka* “only if” (Laoust 1912: 78):

25. Cf. Foucauld 1951–1952: 1163; Prasse et al. 2003: 533; Dallet 1982: 494; Delheure 1987: 187; Taifi 1991: 414; Kossmann 1997: 204.

26. See Lipiński (2001: 548), *contra* Taifi (1993: 219). For the latter, the š in Tamazight *meš* is the truncated form of *ka / ša* which means “thing, something” and which stems from *kra*, the form used with the same meaning in Tashlhit and Kabyle. The hypothesis is interesting, but it could be due to coincidence (the evolution from *kra* to *ka/ša* being triggered through influence from the original particle *k-*). Indeed, in Ouargla, for example, “if” is expressed by *maka*, and the lexeme ‘thing’ *šra* has indeed retained its *r* (Delheure 1987: 324–325).

27. *maka/maša* is also used in Tamazight with the meaning “but, however” – as is *maša* in Rifain (Kossmann 2000: 194).

- (41) *hezda                      ġerka sen-aṭ n tsebbat*  
 3F.SG.crush.PRIV only two-F of F.AS.grain.PL  
 “She crushed only two grains.”

In Chenoua, the recurrence of the element *ka* thus seems to establish a particular link between conditional and exception particles.<sup>28</sup>

Table 4. Particles with *k* (or *k>š*) in Berber

	conditional	direct question	indirect question	exception	deictic/ presentative	indefinite/ interrogative
(a) <i>k/ka</i>	–	+	–	+	+	+
<i>waš</i>	–	+	+	–	–	–
<i>maka...</i>	+	–	–	+	–	+
<i>kagella</i>	+	–	–	–	–	–

## 5. *kan* in the eastern varieties


In Berber, and more particularly in the eastern varieties, *kan/kān* is often used as an introductory particle – for, among other uses, the conditional. Originally it was the Arabic existence verb *kān(a)* which, in the various Arabic dialects, has been grammaticalized with different uses, and became an invariable form (cf. Taine-Cheikh 2014).

- a. In Arabic, *ka(a)n* is very commonly found introducing the protasis of conditionals, either alone or in more complex units (*lukān, in kān ...*). The same is true in Berber.
  - i. The phrase *in kān* is close to the canonic Arabic form where the particle (')*in*, of deictic origin, is followed by the verb *kān(a)* in the perfective (always frozen in Berber). The use of *in* before *kān* is rare however, and the only other trace of it we were able to find was in the Libyan Sokna variety (Sarnelli 1925).<sup>29</sup>

28. This same element appears to be present (in the form *ka>ša*) in *haša* “except” in Kabyle (Dallet 1982: 302). One also finds *mak d* (*d* being the predicative) “nothing but, only that” in Ouargla (Delheure 1987: 187).

29. For example p. 33, § III, ln 12.

- ii. The phrase *lu(u)ka(a)n* (composed of *kān* preceded by a presentative particle) is also borrowed, more or less as such, from the Arabic.<sup>30</sup> It is found in Kabyle (where it is in competition with *limmer*), in the Algerian Chenoua variety (for past irrealis) and as far away as in Rifain, in the abbreviated form *luk*. It most often occurs however in Libyan varieties. In the Nefousa variety, the variant *lūkān* coexists alongside others, with closely related origins and etymologies: *liākān*, *liā*, *ídā kān* (Beguilot 1931: 94, 127).<sup>31</sup> In contrast, in the Augila variety, *luka(a)n* is the only conditional particle of Arabic origins – the others being *endú/úndu* and *amúr* (Paradisi 1960: 174). In Zuara, the *l* in *lukan* can be omitted (Mitchell 2009: 112):

- (42) *(l)ukán wəṭšá fli-ġ* *dídə=s*,  
 if NEG GO.NEG.PRIV-1SG with=PR.3SG  
*w=atthəṣṣlə-ġ=tí* *sgər=s*   
 NEG=got.NEG.IPRV.1SG=PR.OBJ.3M.SG from=PR.3SG  
 “If I hadn’t gone with him, I wouldn’t have got it from him.”

- iii. In Ouargla, the originally Arabic element *kan* is preceded by the Berber particle *ma*. It is thus a mixed form, half Berber, half Arabic, and is semantically equivalent to both *inkan* or *lukan* and *mayəlla*, and perhaps even to *kagəlla* (Delheure 1987: 187):

- (43) *makan d iʕzam*, *i-ʕəzzəm mʕa=s*  
 if PRED M.FS.study.SG 3M.SG-study.IPRV with=PR.3M.SG  
 “If it is about studying, he studies with him.”

- iv. Lastly, as in a certain number of Arabic dialects, one finds uses of *ka(a)n* on its own. This is the case in El-Foqāha, in the Fezzan region (Paradisi 1963: 121). It is also the case in Siwa Oasis Egyptian (Laoust 1931: 137):

- (44) *kan la hsiṭ g us=əd did=i*  
 if NEG want.PRIV.2SG POT [2SG].come.PRIV=PROX with=PR.1SG  
*ga h-ağ iman=ənn<sup>a</sup>=o*  
 POT GO.PRIV-1SG M.soul.SG=of=PR.POSS.1SG  
 “If you do not want to come with me, I will go alone.”

- b. *ka(a)n* (or one of its variants: *la-kān*, *ləkkān*, *(yā)kūn...*) is used at times in Arabic, to express exceptions or restrictions. It is quite frequently found in the Maghreb, particularly in the eastern varieties – from western Algeria to Libya.

30. On this particle in Semitic, see Lipiński (2001: 482–483). Furthermore, this raises the question whether the particle *l-*, which one finds for example in Kabyle in *limmer*, is not its Berber equivalent.

31. An older document provided a more Berber form: *ma ‘si’* (Motylinski 1898: 36).



In Berber, a comparable use of *kan* “except, only” has sometimes been noted in this same eastern region.<sup>32</sup> For the Tunisian Sened variety, Provotelle (1911: 114) considers *ġir* and *kän* equivalents: ... *ġir idjet/kän idjet* “... except only one”. In Zuara, *kan* on its own is noted by Mitchell (2009: 108):

- (45) *ádđwa w=əttitš-ən=ti kan*  
 medicines NEG=give.NEG.IPRV-3PL=PR.OBJ.3M.SG except  
*i=ṁamṁu d=ṁáḍun*  
 to=someone PRED=sick.man  
 “They only give medicines (lit. medicines they give them only) to someone who is ill”, i.e., “only the sick man takes medicines”.

- c. In Arabic, the particle which introduces direct questions is often also used to introduce conditionals: *iḍa*, (‘)in, (‘)in *ka(a)n*, more rarely *kān* alone, contrary to (a)*kān* in the Sudanese Šukriyya variety, and to (yā)*kān* in Mauritanian ḥassāniyya.

In Berber, it would seem that Siwi – a variety under strong Arabic influence – is alone in using the particle *kan* in this same context (Laoust 1931: 137):

- (46) *la ssen-ağ kan g us=əd*  
 NEG know.PRV-1SG if POT [3M.SG].COME.PRV=PROX  
 “I do not know if he will come.”

Table 5. The particle *kan* in Berber

	conditional	indirect question	exception	existence verb
<i>kan</i>	+	+	+	+

## 6. *kud* and its variants in the southern central area

The different varieties in the Tuareg area are characterized, like Libyan Ghadamès Berber, by the use of a particle based on *ku-*. This particle, which shows a certain degree of regional variation, is partially used in free variants (*kud/kūd/nkūd/kud-ənta/kunta*). In substance, it has comparable uses to those discussed above.

- a. In its most frequent use, *kud* introduces the first clause in complex sentences. *kud* can be used as a temporal subordinator, possibly with causal connotations, as in the Ghadamsi example in (47) (Lanfry 1973: 242):

32. It has also been noted in Kabyle (Amina Mettouchi, personal communication).

- (47) *nkūd ferren-n-et*                      *ane<sup>f</sup> əzzed-n-et*,  
 if    riddle.IPRV-3PL-F or    grind.IPRV-3PL-F  
 “While they riddled and ground (grain)...”

This use is also found in Tamazight, where *kud* (or *kkud*, cf. Taifi 1991: 322), has also been noted with the meaning “at the same time as, as long as” (Bentolila 1981: 335):

- (48) *kud i-Tḥasab*,                      *i-TrTab*  
 while 3M.SG-count.IPRV 3M.SG-set.the.amount.of.tax.IPRV  
 “While counting (cattle livestock), he sets the amount of tax.”
- (49) *kud i-Tḡar*,                      *Tḡar-n-t*  
 as 3M.SG-dry.IPRV dry.IPRV-3PL-F  
 “As (the snail) dried, (the warts) dried (also).”

However, it is only in the southern central area that *ku-* and its variants have specialized as conditional particles. For Prasse (2005: 169), the particle *kud*, *kudet* “if” which introduces the potential conditional stems from Arabic:

*kud* provient très probablement de la conjonction maghrébine *kūn*, *kū* ‘vu que, parce que, puisque; car’ qui provient du classique *kawn* ‘existence; événement, incident’ (*li-kawni-hi* ‘parce qu’il’ etc.). *kūn* s’abrège souvent en *kū* qui a pu être emprunté par le touareg et élargi de *əd* (‘dans l’incident que’).<sup>33</sup>

According to this hypothesis, the etymology of *ku(d)* would be akin to that of *kān*: the same Arabic root KWN, but a form derived from a noun phrase rather than from a verb form.<sup>34</sup> In Ghadamès, the initial nasal consonant in *nkūd*, a variant of *kūd*, does not *a priori* bear any relation to the *n* in (*i*)*inkan*, unless one imagines a mixed form reflecting dual origins (*li-kawni-hi* on one hand, (ʿ)*in-kān(a)* on the other). Be that as it may, the particle regularly introduces conditional protases,<sup>35</sup> as can be seen in the following example (Lanfry 1973: 242):

33. “*Kud* very probably derives from the Maghrebin *kūn*, *kū* “given that, since, because” from the Classical Arabic *kawn* “existence, event, incident” (*li-kawni-hi* “because he” etc.). *kūn* is often shortened to *kū* which may have been borrowed by Tuareg and expanded into *əd* (“in the incident that”)” (translation by Margaret Dunham).

34. In Maghreb Arabic, *kwn* means “action of being, existing, such or such a state”; “because, given that, because”; *lkwnh kbīr* “because he was big” (Beaussier 1958: 886).

35. Lanfry (1973: 180–181, n° 0893) also translates the particle (*i*)*lām* by “if”; however this particle, the etymology of which is highly mysterious (could there be some relation to the two particles *l-* and *m-*?), appears in a specific structure, of a correlative type: (*i*)*lām*... (*i*)*lām* ... .

- (50) *nkūd* *y-ūtef* *wī*  
 if 3M.SG-enter.PRIV PR.DEM.M.SG  
*meqqūr-en* *i=tali=yī*  
 be.older.PRIV-PTCP.SG to=F.room.SG=DEM.SG  
 “if (when) the elder entered the room...”

In Tuareg, *kud* is the most frequently occurring form. However, in Tahaggart, one also finds *ku* and *kudit* (Foucauld 1951–1952: 742) and, further to the southwest, a suffixed variant *-nta*: *kunta* the Tadraq of Kidal and *kud-ənta* in the Tamaghit of Oudalan (Sudlow 2011: 336).

- b. The above-studied element *k(a)* is not entirely unknown in southern Berber. In Ghadamsi, *ak* has supplanted *wel* as a negation particle in independent or main verb clauses (Lanfry 1973: 143, 388). In Tuareg, it is less a negation particle, except perhaps in *kala* “no” (Motylinski 1908: 51) than one of interrogation, as well as, more originally, one of insistence.<sup>36</sup> In Mali Tamashek for example, *ák* is one of the “[c]ause-initial particles for polar (i.e., ‘yes-no’) interrogatives” (Heath 2005: 649).<sup>37</sup>

It is possible that the existence of this element *k(a)* or *ak* played a role in the morphogenesis of the indirect question particle. Whatever the possible effect, this particle’s form is currently identical to that of the conditional, namely *kud* or one of its variants.

In Ghadamès, one finds examples of dependent interrogatives following verbs such as *ellām* “look” or *essən* “know”, illustrated in (51) (Lanfry 1973: 345, number 1483):

- (51) *essən* *nkūd* *əllān* *āman* *ğeşşuf=i*  
 know.IMP.SG if be.PRIV.3PL M.water.PL Rassouf=in  
 “Go see if there is water at Rassouf.”

In Tuareg also, *kud* “a développé des sens atténués [et] est devenue la marque de l’interrogative dépendante” (Prasse 2005: 147).<sup>38</sup> This is illustrated in the following example in Tahaggart (Prasse 2005: 335):

36. This is true for the particle *ak* “there, then” which, in Niger, is frequently suffixed to the personal pronouns in the 2nd person (Prasse et al. 2003: 351).

37. *Ák* also has wider uses as an “appetizer” (Heath 2005: 649) “in clause-initial position before a topicalized NP, followed by a WH-interrogative” but “it can be glossed contextually as ‘or rather’, introducing a self-correction”.

38. “Has developed bleached meanings and has become the dependent question marker” (translation by Margaret Dunham).

- (52) *ed=saggäd-än*      *däg*    *tädägdäqq=ən=et*      *kud*  
 POT=look.IPRV-3PL at    F.AS.arpit.SG=of=PR.POSS.3F.SG if  
*hän=tät*      *imžad-än*    *meğ*    *käla*  
 be.in.PR.V.RES.3PL=PR.OBJ.3F.SG M.AS.hair-PL or not  
 “They begin to look at his armpit [to see] if there is hair or not.”

- c. In Ghadamès, exceptions are expressed using *halef* “except” (Lanfry 1973: 137 number 0661). In Tuareg, they are often expressed using *(a)sel/(a)selid* or using *ar*, a particle which probably shares its origins with the preposition “until” (pronounced *ar* or *har* depending on the dialect; cf. Heath 2005: 618).<sup>39</sup> However, with the meaning “except if, unless, only if”, Tuareg also uses expressions where one recognizes the particle *kud*: “Une forme plus pleine se trouve dans la préposition/conjonction *kundäba* “si ce n’est, excepté; à moins que” = *kudäba* = *kud-ba-t*” (Prasse 2005: 169).<sup>40</sup> (53) is an example of *kundäba* in Aïr Tuareg (Kossmann 2011: 172):

- (53) *kundäba*    *t-əg-äm=i*      *tasəgbəs-t*      *a*  
 unless    2-do-PRV.PL=PR..    F.FS.underskirt-SG    PR.N.SG  
*he=din=əzzäbbe*  
 POT=ABL=descend.AOR.[1SG]  
 “It is only if you give me an underskirt that I will descend.”

The fact that the Tuareg exception particle also shows a certain amount of convergence with the conditional and indirect question particles indicates that the previously observed similarities are not due simply to homonymy.

Table 6. The particle *kud/nkü*d in Berber

	conditional	indirect question	exception	temporal subordinator
<i>kud/nkü</i> d...	+	+	+	+

## 7. Conclusion

The study of the particles used in Berber to express conditions shows that there are true convergences, albeit with some variation.

39. About *mäššär* “except” in Zenaga and its possible link with *är*, cf. Taine-Cheikh 2011: 546–547.

40. “A fuller form can be found in the preposition/conjunction *kundäba* “if not, excepted, unless” = *kudäba* = *kud-ba-t*” (translation by Margaret Dunham).

These convergences have a spatial dimension: five regions each show a preference for one of the variants (or group of variants): *ad* in Mauritania, *is* in southern Morocco, *ma*, *(a)k/ka* and *maka* in the northern dialects (in Morocco and Algeria), *kan* in the eastern dialects and lastly *kud* in the southern dialects. This highlights how innovations tend to spread from one group of speakers to another, in keeping with the wave propagation model.

Convergences are also to be found in the meanings of the particles. Particles which serve to introduce the protasis of a conditional are often more or less closely linked to, on one hand, those which (directly or indirectly) introduce interrogative clauses and, on the other hand, those used to express exceptions.<sup>41</sup>

Table 7 summarizes these results in a simplified manner (all variants are not noted, e.g., complex particles such as *in kan* or *lukan*). The other uses of these same particles are listed on lines 6 to 9.

Table 7. Uses of all particles

	<i>ad</i>	<i>is</i>	<i>as</i>	<i>ma</i>	<i>(a)k/ka</i>	<i>maka</i>	<i>kan</i>	<i>kud</i>
conditional	X	X	X	X		X	X	X
direct question		X		X	X			
indirect question	X	X		X			X	X
exception			X		X	X	X	X
deictic/presentative	X	X			X			
existence verb							X	
complementizer	X	X	X					
temporal subordinator			X					X

In the *World Lexicon of Grammaticalization*, the conditional is only listed as a source of grammaticalization for the concessive. However, Heine & Kuteva note four possible sources for the conditional (2002: 329): “conditional < (1) copula (2) s-question [marker of polar (yes-no) questions] (3) say (4) temporal”.

If one adopts these hypotheses, Berber shows three of the four sources: (1) *ad* and *is*; (2) *(a)k* and *ma*; and (4) *kud*. The case of *kan* does not fit this framework, but could be explained by the deletion, preceding the existence verb, of the particle *(ʿ)in*, of deictic origins.

41. These polygrammaticalizations are comparable to those of the French *si* (“if”).

Looking at the table, one sees another trait shared by *ad* and *is/as*: that of being used both to introduce the protasis of conditionals and as a complementizer.<sup>42</sup> However, *ad* and *is/as* are not radically different from the other particles: they all appear to have originally been used more for discursive purposes than for morphosyntactic ones. This is well illustrated by the marker *is*, which continues to be optional in direct questions, although it is not alone in this.<sup>43</sup>

Originally, the primary role of particles was to introduce an element or clause, to question or on the contrary reaffirm its reality (even when only in a fictive mode). Thus grammaticalization affected specific linguistic tools such as demonstratives, presentatives, topicalizing and focalizing particles, and existence verbs.<sup>44</sup> There is no single grammaticalization path, but rather several partially parallel paths (leading, in several cases, to morphemes with mixed origins). Among the chosen grammaticalization paths, some are highly frequent cross-linguistically. Others, more rare, found in some Arabic varieties, belong at least in part to the domain of borrowings.

### Specific abbreviations

AS	annexation state
FS	free state
HAB	habitual particle
POT	potential particle
PRED	predicative particle

### References

- Beaussier, Marcelin. 1958. *Dictionnaire pratique arabe-français*. Alger: La Maison des Livres.  
 Beguinot, Francesco. 1931. *Il Berbero Nefûsi di Fassâto*. Roma: Istituto per l'Oriente.  
 Bentolila, Fernand. 1981. *Grammaire fonctionnelle d'un parler berbère: Aït Seghrouchen d'Oum Jeniba (Maroc)*. Paris: SELAF.

42. According to Frank Lichtenberk (personal communication), this two-fold grammaticalization happens very frequently in the Oceanic languages: such is the case for *bu* in Bukawa, of *naka* in Tawala, of *ta* in Kokota, of *we* in Sakao, of *be* in Raga, of *ma* in Ponapean and of *pē* in Samoan.

43. Provotelle (1911: 77) notes that in Sened, the conditional “if” is not translated.

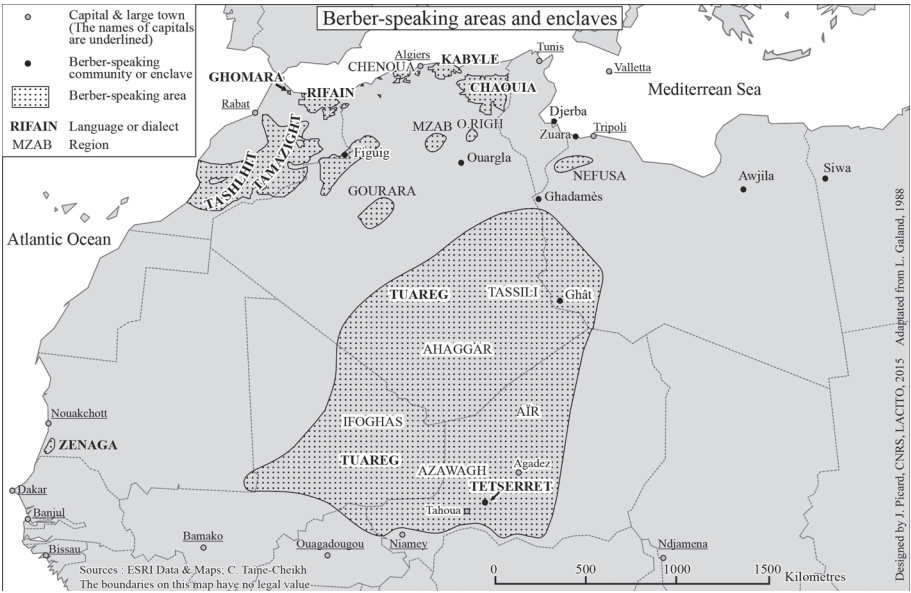
44. On these morphemes’ polysemy and the plurality of factors entering into their grammaticalizations, see, e.g., the articles published in Robert (2003) and Bril (2010).

- Bentolila, Fernand. 1988. "Les syntagmes verbaux des serments dans différents parlers berbères". *Awal* 4.43–72.
- Bril, Isabelle (ed.). 2010. *Clause Linking and Clause Hierarchy: Syntax and Pragmatics*. Amsterdam: John Benjamins. doi:10.1075/slcs.121
- Dallet, Jean-Marie. 1982. *Dictionnaire kabyle-français, parler des At Mangellat, Algérie*. Paris: SELAF.
- Delheure, Jean. 1987. *Dictionnaire ouargli-français*. Paris: SELAF.
- Destaing, Edmond. 1920. *Etude sur la Tachelhît du Soûs. I Vocabulaire français-berbère*. Paris: Leroux.
- Faber, Alice. 1991. "The Diachronic Relationship between Negative and Interrogative Markers in Semitic". *Semitic Studies (Hommage à Leslau)*, ed. by Alan S. Kaye, 411–429. Wiesbaden: Harrassowitz.
- Foucauld, Charles de. 1951–1952. *Dictionnaire Touareg-Français (Ahaggar)*, 4 vols. Paris: Imprimerie Nationale de France.
- Galand, Lionel. 1988. "Le berbère". *Les langues dans le monde ancien et moderne. III Les langues chamito-sémitiques*, ed. by Jean Perrot, 207–242. Paris: CNRS.
- Galand, Lionel. 2002 [1987]. "Subordination résultant de la relation: à propos de la relative berbère". *Etudes de linguistique berbère*, ed. by Lionel Galand 241–256. Paris & Louvain: Peeters.
- Heath, Jeffrey. 2005. *A Grammar of Tamashek (Tuareg of Mali)*. Berlin: Mouton de Gruyter. doi:10.1515/9783110909586
- Heine, Bernd & Tania Kuteva. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511613463
- Kossmann, Maarten. 1997. *Grammaire du parler berbère de Figuig (Maroc oriental)*. Paris & Louvain: Peeters.
- Kossmann, Maarten. 2000. *Esquisse grammaticale du rifain oriental*. Paris & Louvain: Peeters.
- Kossmann, Maarten. 2011. *A Grammar of Ayer Tuareg (Niger)*. Köln: Köppe.
- Lafkioui, Mena. 2007. *Atlas linguistique des variétés berbères du Rif*. Köln: Köppe.
- Lafkioui, Mena & Daniela Merolla. 2002. *Contes berbères chaouis de l'Aurès d'après Gustave Mercier*. Köln: Köppe.
- Lanfry, Jacques. 1973. *Ghadamès. II Glossaire*. Alger: Le Fichier Périodique.
- Laoust, Émile. 1912. *Etude sur le dialecte berbère du Chenoua comparé avec ceux des Beni-Menacer et des Beni-Salah*. Paris: Leroux.
- Laoust, Émile. 1931. *Siwa. Son Parler. Publ. de l'Institut des Hautes-Etudes Marocaines (tome XXII)*. Paris: Leroux.
- Lipiński, Edward. 2001. *Semitic Languages: Outline of a Comparative Grammar*. Leuven: Peeters. 2nd edition.
- Mitchell, Terence Frederick. 2009. *Zu'ar Berber (Libya): Grammar and Texts*. Köln: Köppe.
- Motylinski, Adolphe de Calassanti. 1898. *Le djebel Nefousa. Transcription, traduction française et notes avec une étude grammaticale*. Paris: Leroux.
- Motylinski, Adolphe de Calassanti. 1908. *Grammaire, dialogues et dictionnaire touaregs. I Grammaire et dictionnaire touaregs*. Alger: Imprimerie orientale P. Fontana.
- Naït-Zerrad, Kamal. 2001. *Grammaire moderne du kabyle*. Paris: Karthala.
- Paradisi, Umberto. 1960. "Il berbero di Augila, materiale lessicale". *Rivista degli studi orientali* 35.157–177.
- Paradisi, Umberto. 1963. "Il linguaggio berbero di El-Fogâha (Fezzan)". *Annales de l'Institut des études orientales* NS 13.93–126.
- Penchoen, Thomas G. 1973a. *Tamazight of the Ayt Ndhir*. Los Angeles: Undena.

- Penchoen, Thomas G. 1973b. *Etude syntaxique d'un parler berbère (Ait Fraḥ de l'Aurès)*. Napoli: Istituto Universitario Orientale.
- Prasse, Karl-G. 2005 [2008]. *Manuel de grammaire touarègue. Syntaxe*. Schwülper: Cargo.
- Prasse, Karl-G., Ghoubéïd Alojaly & Ghabdouane Mohamed. 2003. *Dictionnaire Touareg – Français (Niger)*. Copenhagen: Museum Tusculanum Press – Université de Copenhague.
- Provotelle, Paul. 1911. *Etude de la Tamazir't ou zenatia de Qalaât Es-Sened (Tunisie)*. Paris: Leroux.
- Robert, Stéphane (ed.). 2003. *Perspectives synchroniques sur la grammaticalisation*. Louvain & Paris: Peeters.
- Sarnelli, Tommaso. 1925. *Il dialetto berbero di Sokna. Estrait de Africa Italiana 1924–1925*, 46 p., Napoli.
- Sudlow, David. 2001 [2011]. *The Tamasheq of North-East Burkina Faso: Notes on Grammar and Syntax Including a Key Vocabulary*. Köln: Köppe.
- Taifi, Miloud. 1991. *Dictionnaire Tamazight–Français (Parlers du Maroc central)*. Paris: L'Harmattan-Awal.
- Taifi, Miloud. 1993. "L'expression de l'hypothèse en berbère". *A la croisée des études libyco-berbères. Mélanges offerts à Paulette Galand-Pernet et Lionel Galand*, ed. by Jeannine Drouin and Arlette Roth, 215–228. Paris: Geuthner.
- Taine-Cheikh, Catherine. 2008. *Dictionnaire Zénaga – Français. Le berbère de Mauritanie par racines dans une perspective comparative*. Köln: Köppe.
- Taine-Cheikh, Catherine. 2010a. "The Role of the Berber Deictic and TAM Markers in Dependent Clauses in Zenaga". *Clause Linking and Clause Hierarchy: Syntax and Pragmatics*, ed. by Isabelle Bril, 355–398. Amsterdam: John Benjamins. doi:10.1075/slcs.121.11tai
- Taine-Cheikh, Catherine. 2010b. "Ordre, injonction, souhait et serment en zénaga (étude comparative)". *Etudes berbères V – Essais sur des variations dialectales et autres articles*, ed. by Harry Stroomer, Maarten Kossmann, Dymitr Ibrizsimow & Rainer Vossen, 191–212. Köln: Köppe.
- Taine-Cheikh, Catherine. 2011. "L'énoncé négatif en berbère zénaga". *«Parcours berbères». Mélanges offerts à Paulette Galand-Pernet et Lionel Galand pour leur 90<sup>e</sup> anniversaire*, ed. by Amina Mettouchi, 533–553. Köln: Köppe.
- Taine-Cheikh, Catherine. 2014. "Du verbe à la conjonction. Quelques cas de grammaticalisation de *kân* dans les dialectes arabes". *Alf lahğa wa lahğa: Proceedings of the 9th Aida Conference*, ed. by Angela Daiana Langone, Olivier Durand and Giuliano Mion, 423–438. Münster-Wien: LIT.



Appendix



Map of Berber varieties

# The semantics of modals in Kordofanian Baggara Arabic

Stefano Manfredi

SeDyL UMR8202, CRNS, INALCO, IRD

This paper aims at describing the forms and the meanings of modal items in Kordofanian Baggara Arabic, a Western Sudanic Arabic dialect spoken in Southwestern Sudan. It presents a polysemic analysis of modal items in light of the participant-oriented approach, and it shows how modality can be described in terms of gradable scales rather than discrete meanings. Beyond that, the paper deals with the hypothesis of the unidirectional development of modal meanings according to which deontic meanings precede epistemic ones and it eventually argues that Kordofanian Baggara Arabic presents an atypical path of grammaticalization from a participant-external possibility to a deontic possibility.

**Keywords:** modality, grammaticalization, Baggara Arabic

## 1. Introduction

Modality undoubtedly represents one of the most complex aspects of both descriptive and comparative linguistics. This is mainly because modal meanings as well as their formal expressions vary a great deal across languages. As a further matter, there is as yet no consensus on a proper terminology for describing and comparing modal meanings.<sup>1</sup> Traditional approaches to modality make a broad distinction between ‘epistemic’ and ‘deontic’ modalities. Epistemic modality generally refers back to the degree of certainty the speaker has about what he or she is saying, whereas deontic modality codes the speaker wishes with respect to the uttered proposition. This classification of modal meanings implies a third kind of modality, that is, ‘dynamic’ modality, which typically includes ability and volition

---

1. Given the descriptive purposes of this paper, the following theoretical introduction is far from being exhaustive. See Nuyts (2005, 2006) and de Haan (2006) for different comprehensive accounts of typological approaches to modality.

(de Haan 2006: 31). Despite the fact that this terminology has been thoroughly used in functional-typological studies (Palmer 2001; Nuyts 2006; Frajzyngier 2002), during the last decades new classifications of modal meanings have been proposed.

One of the most influential approaches to modality is that of Bybee & Fleischman (1995) who distinguish between epistemic, speaker-oriented and agent-oriented modalities. According to this classification, speaker-oriented and agent-oriented modalities share the area of dynamic/deontic modality. Nevertheless, if speaker-oriented modality exclusively refers back to modal meanings in which the speaker is the enabling condition, agent-oriented (or 'intersubjective') modality typically describes modal meanings that predicate conditions on an agent with regard to the completion of an action (Bybee & Fleischman 1995: 6).

Van der Auwera & Plugian (1998), for their part, proposed a participant-oriented approach to modality. In this perspective, there is basic opposition between epistemic and situational modality (which is also referred to as non-epistemic modality). Epistemic modality concerns the speaker's degree of certainty about his or her assertion, and it therefore has scope over the whole proposition, whereas situational modality relates with aspects internal to the state of affairs that the proposition reflects. Central to the participant-oriented approach is that 'situational' modality is again divided in two types: 'participant-internal' and 'participant-external' modality.

The first term refers to a kind of possibility and necessity internal to a participant engaged in the state of affairs. ... The [second] term refers to circumstances that are external to the participant, if any, engaged in the state of affairs and that make this state of affairs either possible or necessary.

(Van der Auwera & Plugian 1998: 80)

Deontic modality is thus seen as a sub-type of participant-external modality that encompasses permission, advice, and obligation.

As far as the study of modal meanings in modern Arabic dialects is concerned, it reflects the different theoretical approaches to modality. On the one side, Ingham (1994) and Mitchell & al-Hassan (1994) adopt the traditional standpoint of epistemic, deontic, and dynamic modality. On the other side, Vanhove et al. (2009) prefer to analyze the grammaticalization of modal items in a number of Arabic varieties according to the intersubjective approach of Bybee & Fleischman. This paper presents a polysemic account of modal expressions in Kordofanian Baggara Arabic in light of the participant-oriented approach proposed by van der Auwera & Plugian. Since modal meanings develop gradually, I stick to the view that both situational and epistemic modalities can be described most adequately in terms of gradable scales rather than discrete meanings. That being so, the synchronic description of modal meanings in the Baggara dialect of Kordofan will be coupled

with a semantic map that will allow us to make certain assumptions about their diachronic development.

The paper is organized as follows. § 2 briefly introduces the main phonological and morphological isoglosses characterizing Kordofanian Baggara Arabic. § 3 presents some preliminary remarks about the expressions of tense, aspect, and mood in the dialect in question. The core of the paper in § 4 describes the semantics of different modal items. These include modal auxiliaries such as *gídīr/b=i-gdar* ‘can, be able’; pseudo-verbs such as *dāyir* ‘want, need’, *mimkin/imkin* ‘it’s possible’; adverbs such as *ille* ‘except’ and *lāzim* ‘it’s necessary’; complex adverbial constructions such as *la buddi* ‘inevitably’, *axēr lē* ‘had better to, ought to’, *min la buddi* ‘it’s likely’; and grammaticalized invariable markers such as *bukūn* (epistemic) ‘must’. Finally, § 5 summarizes the semantics of modal items and explains the diachronic relationships between their meanings.

## 2. The classification of Kordofanian Baggara Arabic

Kordofanian Baggara Arabic (hereafter KBA) is a Bedouin Arabic dialect spoken in the Southern Kordofan state of the Republic of Sudan. Adopting the label ‘Baggara’ for this dialect, I point to its inclusion into a dialect sub-type characteristic of Arab semi-nomadic cattle-herders living scattered through a vast area running from Lake Chad to the White Nile, the so-called Baggara Belt. Broadly speaking, KBA presents a number of features that clearly testify to its affiliation to Sudanic Arabic. Among the most important pan-Sudanic isoglosses found in KBA, we can enumerate the diachronic developments  $*q > g$ ,  $*ḡ > dʒ$  ( $j$  in my phonological transcription),  $*ḍ > ḍ$ ; the presence of  $tʃ$  ( $c$  in my phonological transcription) as a phonemic consonant; the presence of a single morphological set of demonstrative pronouns and determiners lacking the etymological morpheme *ha-*; the generalized use of a pre-formative element *a-* in the formation of imperatives; the presence of the auxiliary *gāʿid* for expressing the progressive aspect of action verbs (see § 3).

Beside, there is unmistakable evidence of the historical link of KBA with Western Sudanic varieties of Arabic and more in particular with other Baggara dialects spoken in Chad and Nigeria. Like western Baggara dialects, KBA is characterized by the historical development  $*ʔ > ʕ$ ; the presence of the non-etymological consonants  $n$  and  $ɲ$ , the presence of vowel backness harmony, and the alternation *-e/-a* in singular feminine marking. From a morphological point of view, KBA also aligns with Western Sudanic dialects with regard to the forms of the bound pronouns  $=ki$  2SG.F and  $=ku$  2PL.M, the singulative marker *-ay*, the elision of *-i* in final weak verbs, and the presence of the prefix *al-* for reciprocal derived verbs



Figure 1. Kordofanian Baggara Arabic and the Baggara Belt (Manfredi 2012: 142)

Furthermore, the contacts between Baggara Arabs and both sedentary populations of Kordofan and Arab camel-herders coming from Eastern Sudan made KBA more similar to Eastern Sudanic dialects (Manfredi 2012). On the one hand, the process of dialect contact resulted in the affirmation of new mixed features, as in the case of 1SG and 1PL indexes in the imperfective conjugation (1SG *a/-n-* ... 1PL *n-*... (*o/u*)). On the other hand, dialect contact lead to concurrence of historically unrelated forms, as in the case of the possessive particles *hān* (found in Nigeria and Chad; see Owens 1993a: 64–66; Owens 1993b: 111) and *hūl* (characteristic of the Bedouin dialects of eastern Sudan; see Reichmuth 1983: 111–112).

As a final remark, KBA displays a number of phonological and morphological features that characterize it as a Bedouin Arabic dialect. These include the development  $*\dot{g} > q$ , the form  $=a$  for the 3SG.M bound pronoun, and the presence of feminine plural as a morphological category (Rosenhouse 2006: 259–261).

### 3. Tense, aspect and mood in KBA

KBA presents an aspectual bipartite system with perfective and imperfective values respectively expressed by suffixed and prefixed conjugations. When the prefixed conjugation occurs in non-modal contexts, it is typically marked by the proclitic  $b(i)=$ . As is well known, the preverbal marker  $b(i)=$  is a wide-range morphological isogloss whose semantic values vary a great deal. In Levantine, Egyptian, and Northern Sudanese dialects,  $b(i)=$  can mark cursive, progressive and habitual aspects (see Eksell 2006 for Syrian Arabic). Northern Yemeni dialects use  $b(i)=$  to

indicate future tense (Cohen 1984: 281; Behnstedt 1985: 132). Likewise, in Gulf dialects *b(i)=* is related to the expression of an irrealis mood (Persson 2008: 48).<sup>2</sup> Looking at the Western Sudanic area, rural Chadian and Nigerian dialects use *b(i)-* as a person marker for distinguishing 1SG and 3SG.M/PL from 1PL and 2SG/PL that are respectively marked by *n-* and *t-* in the imperfective (Owens 1993b: 105). Thus, it seems that in this dialectal area the marker *b(i)-* has reached a higher degree of grammaticalization as compared to Middle-Eastern dialects.<sup>3</sup> In contrast to this, KBA aligns to Middle Eastern Bedouin dialects (Rosenhouse 2006: 266; Palva 2008: 58) in the use of the proclitic *b(i)=* as an indicative marker for introducing ordinary objective statements. In more detail, KBA *b(i)=* is associated with two temporal references the most common of which is generic present as we can see in the following examples.

- (1) *ana b=a-ʿarf=a zēn*  
 1SG IND=1SG-known=3SG.M well  
 “I know him well.”
- (2) *al=kalb ḥiss=a b=i-nbaḥ*  
 DEF=dog noise=3SG.M IND=3SG.M-bark  
 “The dog barks.”

Given the absence of a specific marker for future tense, *b(i)=* also refers to factual events that have not (yet) taken place.

- (3) *ambākir ana b=a-gūm xalāṣ*  
 tomorrow 1SG IND=1SG-get\_up definitively  
 “I will definitively leave tomorrow.”
- (4) *hu ma b=i-lgá be=suhūla*  
 3SG.M NEG IND=3SG.M-find\3SG.M by=simplicity  
 “He won’t find him effortless.”

From an aspectual point of view instead, *b(i)=* normally expresses a habitual action as showed in Example (5).

- (5) *ana yōt b=a-gabl=a*  
 1SG daily IND=1SG-meet=3SG.M  
 “I meet him daily.”

2. Note that *b(i)=* also expresses an irrealis mood in Juba Arabic, the Arabic-based expanded-pidgin spoken in South Sudan (Tosco 1995: 444).

3. In Abbéché (Eastern Chad), *b(i)-* occurs as a prefix only with the 3SG.M person of the imperfective (Roth-Laly 1979: 45–47).

Concerning the expression of the progressive aspect in KBA, it has different syntactical representations determined by the lexical aspect of the verbs (Manfredi 2010). Stative verbs such as *nām* “sleep”, *wágaf* “stand”, *jálas* “seat”, *rágad* “lay down”, *gá’ad* “stay” regularly express the progressive aspect by means of active participles. Motion verbs such *maša* “go”, *ja* “come”, *ráwwah* “leave”, *tála* “go up”, *ríji* “come back”, *sār* “move for the transhumance” express a progressive meaning by means of both *b(i)*= marking and active participles. In this case, the choice of *b(i)*= marking depends on the presence of a lexically expressed argument after the verb. If the motion verb does not entail any argument (6a), then its progressive value is expressed by means of an active participle. In contrast, if the motion verb is followed by a lexically expressed argument, then its progressive value is expressed by means of a *b(i)*= marked prefixed conjugation.

- (6a) *itte māši wēn ?*  
 2SG.M go\ACT.PTCP.SG.M where  
 “Where are you going to?”
- (6b) *b=a-mši s=sūk*  
 IND=1SG-go DEF=market  
 “I am going to the market.”

In addition, *b(i)*= regularly marks prefixed stems in the apodosis of factual conditional sentences, as we can see in Example (7).

- (7) *aṛ=ṛuwāba kin sāyr-īn*  
 DEF=curled\_milk if move\ACT.PTCP-PL.M  
*bi=na-dírr=ah lē=l=kulāb*  
 IND=1PL-pour=3SG.F to=DEF=dog\PL  
 “As far as the curled milk is concerned, if we are moving for the transhumance, we give it to the dogs.”

Apart from their occurrence after modal auxiliaries (see § 4), non *b(i)*= marked prefixed stems appear in the following dependency contexts. First of all, like Sudanese Standard Arabic (Ali & Miller 1986: 173), the progressive aspect of action verbs implying a direct object such *ákal* “eat”, *kátab* “write”, *sárag* “steel”, *šírīb* “drink”, *kátal* “hit” is expressed by the auxiliary *gā’id* (which is the active participle of the verb *ga’ad* “sit”) followed by an unmarked prefixed stem.

- (8) *ana gā’id a-ktib juwāb*  
 1SG sit\ACT.PTCP.SG.M 1SG-write letter  
 “I am writing a letter.”



In purposive clauses presenting two non-auxiliary verbs, the second finite verb of the sequence is always unmarked as we can see in (9)–(10).

- (9) *al=yōm ji-na na-kul ma'=ku*  
 DEF=day come-1PL 1PL-eat with=2PL.M  
 “Today we came to eat with you.”
- (10) *jī-t ašān a-llāgi ma'=ali*  
 come-1SG in\_order\_to 1SG-meet with=Ali  
 “I came in order to meet Ali.”

As a further matter, unmarked prefixed stems in non-dependent contexts can also introduce modal meanings such a prohibitive (11) or a jussive (12).

- (11) *ma ta-kul*  
 NEG 2SG.M-eat  
 “Don’t eat!”
- (12) *aḷḷa i-sallim=ak*  
 God 3SG.M-preserve=2SG.M  
 “May God preserve you.”

Finally, we can note that, apart from its ordinary perfective meaning, the suffixed conjugation of motion verbs is limitedly related to the expression of a performative modality, as we can see in (13).

- (13) *mašē-na xalāš ?*  
 go-1PL definitively  
 “Shall we leave?”

#### 4. The forms and the semantics of modals in KBA

In this section I will draw a synchronic analysis of the semantics of modals in KBA. Like other Arabic dialects, in KBA modal meanings can be expressed by fully inflected lexical verbs (e.g., *gidir/b=i-gdar*, see § 4.1), pseudo-verbs (e.g., *dāyir*, see § 4.2), particles (e.g., *ille*, see § 4.4), adverbs (e.g., *lāzim*, see § 4.5), and complex adverbial constructions (e.g., *min la buddi*, see § 4.8). In any case, the main verb of the modal clause is a finite verbal form inflected for gender and number and, in the imperfective, it is never marked by the indicative *b(i)=*. Following Cohen (1984), Simeone-Senelle & Vanhove (2003: 616) and Vanhove et al. (2009: 326–327), I identify auxiliary constructions on the basis of both syntactic and semantic criteria. First of all, an auxiliary item and the following main verb can be defined as a syntactic and semantic unit that cannot be separated by coordinating and subordinating



elements and whose subject is that of the main verb. Likewise, if the modal clause entails a complement, it is always that of the main verb. From a semantic point of view, only the main verb is necessary for communication, since the auxiliary item utters complementary information about modality.

#### 4.1 *gídir*, *b=i-gdar* “can, be able”

The verb *gídir* “can, be able” is etymologically related to the root √QDR expressing the meaning of “to possess strength, power, ability, be strong”. When it is used as a situational modal auxiliary, *gídir* can be inflected both in the perfective and in the *b(i)*= marked imperfective (i.e., *b=i-gdar*) depending on the temporal/aspectual contexts in which it occurs. In any case, it agrees in gender and number with the subject of the main verb. Differently from Egyptian and Moroccan dialects (Vanhove et al. 2009: 338, 344), *gídir* cannot be used as modal auxiliary either in its active participle form (i.e., *gādir*) or in its unmarked imperfective form (i.e., *i-gdar*). Furthermore, *gídir* cannot trigger any epistemic meaning. Most commonly, *gídir* expresses a participant-internal possibility when it is related to a mental and physical ability.<sup>4</sup>

- (14) *b=a-gdar a-rṭun ar=ruṭāna*  
 IND=1SG-can 1SG-speak\_local\_language DEF=local\_language  
*hint logorí*  
 POSS.SG.F Logorí  
 “I can speak the Logorí language.”
- (15) *bi=ta-gdar ti-gri ‘arabi ?*  
 IND=2SG.M-can 2SG.M-read Arabic  
 “Can you read Arabic?”
- (16) *‘ammanawwal ma gidir-na ni-sīr*  
 last\_year NEG can-1PL 1PL-move\_for\_trashumance  
 “Last year we couldn’t move for the transhumance.”

Besides, *gídir* also introduces a participant-external modality. In this case, *gídir* occurs both in non-deontic contexts when it expresses possibility (17) and in deontic contexts modality when it encodes permission (18).

4. Unlike other Arabic dialects that use the verb ‘*irif*’ “know” for conveying a mental ability, KBA does not make a distinction between mental and physical ability – both are expressed by *gídir*. In point of fact, in KBA the verb ‘*irif*’ expresses the meaning of “understand, know”, whereas the verb *dīra* means “know” (Manfredi 2010: 210).

- (17) *bi=na-gdar na-g'od hini 'ašān ni-lqadda*  
 IND-1PL-can 1PL-sit\_down here in\_order\_to 1PL-have\_dinner  
 “(If you want) we can sit down here to have our dinner.”
- (18) *bi=ta-gdar ta-rgud hināk 'ašān ta-nūm*  
 IND=2SG.M-can 2SG.M-lay\_down there in\_order\_to 1SG.M-sleep  
 “You can lay down there to sleep (but not here).”

Despite this participant-external function of *gídír*, *mimkik* remains the default modal marker of possibility and permission (see § 4.3).

#### 4.2 *dāyir* “want, need”

Situational modality is expressed by the active participle *dāyir* “want, need”. In the same manner of *‘āwiz* ~ *‘āyiz* in Egyptian Arabic, *dāyir* developed into a pseudo-verbal conjugation with nominal marks of gender and number (i.e., *dāyir-e* SG.F, *dāyir-īn* PL.M, *dāyir-āt* PL.F). In KBA, *dāyir* commonly expresses volition or intention.

- (19) *ya axā=i dāyir a-gūl lē=k šoqol*  
 VOC brother=1SG want\ACT.PTCP.SG.M 1SG-say to=2SG.M thing  
 “Hey brother, I want to tell you something.”
- (20) *al=yōm kan dāyir-e a-ji lakadīn ma gidír*  
 DEF=day ANT want\ACT.PTCP-SG.F 1SG-come but NEG can\1SG  
 “Today I would have liked to come, but I couldn’t.”

Besides, *dāyir* is the only modal item expressing a participant-internal necessity.

- (21) *az=zōl da mā=dāyir anihna*  
 DEF=man PROX.SG.M FOC=want\ACT.PTCP.SG.M 1PL  
*na-klif kulfa*  
 1PL-train training  
 “This man needs us to organize a good training (for the next wrestling match).”
- (22) *al=ḥabil dāk dāyir i-nḡašar*  
 DEF=rope DIST.SG.M want\ACT.PTCP.SG.M 3SG.M-be\_shortened  
 “That rope needs to be shortened.”

It is worth noting that, unlike Egyptian Arabic, which uses *‘āwiz* ~ *‘āyiz* for expressing a participant-internal necessity only when the subject of the modal clause is inanimate (e.g., *il=‘arabiyya ‘āwz-a ti-tḡisil* “the car needs to be washed”; see Vanhove et al. 2009: 348), in KBA *dāyir* utters necessity also with animate subjects as we can see in (21).

### 4.3 *mimkin*, *imkin* “it’s possible”

The modal adverb *mimkin* “it’s possible” is a frozen passive participial form, whereas *imkin* resulted from the grammaticalization of the 3SG.M unmarked imperfective \*i-mkin. Differently from Egyptian Arabic in which *mumkin* rates higher than *yumkin* on the probability scale (Mitchell & al-Hassan 1994: 47), in KBA *mimkin* and *imkin* do not present any semantic difference and they can be used for expressing both situational and epistemic meaning”. As already showed (see § 4.1), *mimkin*, *imkin* can overlap with *gídír* in marking a participant-external non-deontic possibility.

- (23) *imkin*            *ti-sawwi*    *šunú~šunú*    *ana*  
 it’s\_possible 2SG.M-do what~what 1SG  
*ma b=a-nṭi=k*                      *lē=ha*  
 NEG IND-1SG-give=2SG.M to=3SG.F  
 “You can do whatever you want, I won’t give it to you.”
- (24) *mimkin*        *ti-rkab*            *lōri*    *‘ašān*            *ta-rja’*            *kudūgli*  
 it’s\_possible 2SG.M-get\_on lorry in\_order\_to 2SG-come\_back Kadugli  
 “You can get on a lorry in order to come back to Kadugli.”

Furthermore, *mimkin* and *imkin* also convey a participant-external deontic possibility linked to a permission as we can see in Example (25a, b).

- (25a) *imkin*            *a-s’al* *lē=k*                      *su’āl ?*  
 it’s\_possible 1SG-ask to=2SG.M question  
 “May I ask you a question?” (am I allowed to ask you a question?)
- (25b) *mimkin* ! (yes, you are allowed to ask me a question)  
 it’s\_possible  
 “Yes, you can!”

As a final remark, a consequence of dialect leveling towards Sudanese Standard Arabic, *mimkin* and *imkin* limitedly overlap with *min la buddi* (see § 4.8) in conveying an epistemic possibility related to the degree of feasibility of a predication.

- (26) *imkin*            *fāt*                      *al=ḥille*  
 it’s\_possible pass.3SG.M DEF=village  
 “He probably went to the village.”

#### 4.4 *ille* “except”

*Ille* is related to the restrictive particle \*illa meaning “unless, if not, except”. Even if *ille* is not semantically negative, it introduces several negative meanings. In first instance, *ille* typically occurs in non-negative predications with the meaning of “anything except, everyone except, nothing except, etc.” as we can see in Example (27).

- (27) *zaman dāk at=turuk ille fi=l=obeyyid*  
 time DIST.SG.M DEF=authority except in=al-Obeyyid  
 “At that time, (there was no) authority except in al-Obeyyid.”

Furthermore, it can occur before the subject of negative clauses with the meaning of “only” (28) or in negative existential phrases with the meaning of “except” (29).

- (28) *ille itte ma akāl*  
 except 2SG.M NEG eat\2SG.M  
 “Only you did not eat.”
- (29) *māfi eyy=zōl ille anīḥna weḥēd=na*  
 NEG.EXS each=man except 1PL alone=1PL  
 “There is no one except us.”

Similarly to the Bedouin dialect of Najd (Ingham 1994: 194), in KBA *ille* competes with *lāzim* (§ 4.5) and *la buddi* (§ 4.6) in conveying a participant-external non-deontic necessity. When used as a modal particle, *ille* can precede both unmarked prefixed verbal stems as in (30) and (31), and nominal predicates as in (32).

- (30) *ille ti-sawwi mitil da*  
 except 2SG.M-do like PROX.SG.M  
 “You have to do like this”
- (31) *ille ti-ndass be=na’āl=ak*  
 except 2SG.M-get\_in by=shoes=2SG.M  
 “You have to get in wearing your shoes”
- (32) *al=ḥabil da ille gudur ragab-it al=kalb*  
 DEF=rope PROX.SG.M except size neck-F.CS DEF=dog  
 “This rope has to be the size of the dog’s neck”

It should be remarked that, differently from *lāzim* and *la buddi*, the modal value of *ille* is limited to a participant-external necessity.

4.5 *lāzim* “it’s necessary”

In all probability *lāzim* is the most common item for expressing necessity and obligation in Eastern Arabic dialects. Differently from *dāyir* (see § 4.2), *lāzim* can be described as an invariable modal adverb. That being so, it does not agree in gender and number with the subject of the main verb. Very often, *lāzim* overlaps with both *ille* (see § 4.4) and *la buddi* (see § 4.6) in the expression of a participant-external non-deontic necessity, as we can see in the following examples.

- (33) *lāzim ta-hfaḍ-i ‘iyāl=ki damman*  
 it’s\_necessary 2SG-guard-F children=2SG.F until  
*ti-šl-i kurdufān*  
 2SG-arrive-F Kordofan  
 “You have to protect your children until you reach Kordofan.”
- (34) *al ma ‘and=a gurus lāzim i-xdim*  
 REL NEG at=3SG.M money it’s\_necessary 3SG.M-work  
 “Who does not have money has to work.”

*lāzim* also overlaps with *la buddi* in conveying a participant-external deontic necessity related to expression of obligation (35) and strong obligation (36). In the latter case, *lāzim* is followed by an independent pronoun which is co-referential with the subject of the main verb.

- (35) *lāzim ti-‘arris as=sane di*  
 it’s\_necessary 2SG.M-marry DEF=year PROX.SG.F  
 “You must marry this year.”
- (36) *lāzim aniḥna na-xuṭṭ fi=hajar al=kūk*  
 it’s\_necessary 1PL 1PL-put in=Hajar\_al\_Kuk  
 “We must encamp in Hajar al-Kuk.”

*Lāzim* also competes with *la buddi* in conveying an epistemic necessity linked to an expression of certainty. However, the epistemic value of *lāzim* is restricted to existential clauses expressed by an unmarked imperfective of the verb “be” (i.e., *i-kūn*).

- (37) *lāzim ‘ali i-kūn ma‘=bābikir*  
 it’s\_necessary Ali 3SG.M-be with=Babikir  
 “Ali must be with Babikir.”

In the same manner as *mimkin* and *imkin*, the limited occurrence of *lāzim* as epistemic auxiliary gives evidence of a recent functional expansion due to the nowadays influence of Sudanese Standard Arabic on the Baggara dialect of Kordofan.

#### 4.6 *la buddi* “inevitably”

The complex adverb *la buddi* is diachronically related to \**lā budda* meaning “inevitably, without fail, by all means”. In KBA *la buddi* expresses a participant-external non-deontic necessity (38) or a participant-external deontic necessity related to obligation (39) and strong obligation (40).

- (38) *la buddi ad=daggāga i-g’a fi=turāb alla da*  
 inevitably al-Daggaga 3SG.M-fall in=ground God PROX.SG.M  
 “Al-Daggaga<sup>5</sup> has to fall down (in the course of the next wrestling match).”
- (39) *hāla la buddi ti-zgi al=bār*  
 now inevitably 2SG.M-water\_cattle DEF=cows  
 “Now you must water the cows.”
- (40) *al=’irse di la buddi na-ḥḍar=ah*  
 DEF=marriage PROX.SG.F inevitably 1PL-attend=3SG.F  
 “As for this marriage, we must attend it.”

*La buddi* is also the default adverb for expressing certainty. When it conveys an epistemic necessity, *la buddi* can combine either with a perfective verb (41) for expressing certainty in the past or with the unmarked imperfective of the verb “be” (i.e., *i-kūn*) followed by an unmarked imperfective (42) or by a participle (43) for expressing certainty in the present.

- (41) *la buddi raja’ min=as=sūk*  
 inevitably come\_back.3SG.M from=DEF=market  
 “He must have come back from the market.”
- (42) *la buddi fāṭme ta-kūn ta-sōṭ al=’asīde*  
 inevitably Fatima 3SG.F-be 3SG.F-stirr DEF=porridge  
 “Fatima must be stirring the porridge.”
- (43) *la buddi i-kūn gā’id fi=l=bēt*  
 inevitably 3SG.M-be sit\ACT.PTCP.SG.M in=DEF=house  
 “He must be at home.”

5. *Ad=daggāga* literally means “the hummer”. In this example the lexeme is used as a nickname of a Baggara wrestler and it is thus not translated.

4.7 *axēr lē* “had better, ought to”

In KBA *axēr* “better” is the suppletive elative form of the adjective *zēn* “good”. When *axēr* is combined with the preposition *lē* “to” followed by a bound personal pronoun, it conveys a participant-external deontic necessity related to a piece of advice. *axēr lē* typically combines with unmarked imperfective stems.

- (44) *axēr lē=k ta-miš ta-nūm henāk*  
 better to=2SG.M 2SG.M-go 2SG.M-sleep there  
 “You had better go sleep there.”
- (45) *axēr lē=ha ta-skut sākit*  
 better to=3SG.F 3SG.F=shut\_up shut\_up\ACT.PTCP.SG.M  
 “She ought to shut up.”

4.8 *min la buddi* “it’s likely”

The complex adverb *min la buddi* is built on the juxtaposition of the partitive preposition *min* and the complex adverb *la buddi* (see § 4.6), and it represents the default epistemic auxiliary for expressing feasibility (46) and near-certainty (47). Different from *mimkin* and *imkin* (§ 4.4), *min la buddi* does not convey any non-epistemic meaning. As a modal auxiliary, *min la buddi* can combine with both perfective and unmarked imperfective stems.

- (46) *min la buddi jāb as=sidge*  
 likely bring.3SG.M DEF=cattel\_dietary\_supplement  
 “He may have brought the *sigde*.”
- (47) *ambākīr da min la buddi a-kūn fi=l=bēt*  
 tomorrow PROX.SG.M likely 1SG-be in=DEF=house  
 “Tomorrow I might be at home.”

4.9 *bukūn* (epistemic) “must”

The invariable epistemic marker *bukūn* “must” resulted from the grammaticalization of the 3SG.M *b(i)*= marked imperfective of the verb “be” (\**b=i-kūn* “he is, he will be”). *bukūn* typically conveys an epistemic necessity related to a logical deduction. It can combine with both perfective (48), and imperfective (49) stems depending on the reference to an event that is supposed to have already happened or that is happening. *bukūn* can occur in predicative noun phrases (50).

- (48) *bukūn māš-at be=qādi*  
 must go-3SG.F by=there  
 “She must have gone this way.”
- (49) *bukūn ta-ji min=ba’īd*  
 must 3SG.F-come from=far  
 “She is certainly coming from far.”
- (50) *z=zōl dāk bukūn mayyit*  
 DEF=man DIST.SG.M must die\PASS.PTCP.SG.M  
 “That man must be dead.”

## 5. Conclusions

The following chart summarizes the semantic values expressed by modal items in KBA.

Table 1. The semantics of modals in KBA

Item	Semantics	Modal value
<i>gīdir, b=i-gdar</i>	participant-internal possibility	ability, capacity
	participant-external non-deontic possibility	possibility
	participant-external deontic possibility	permission
<i>dāyir</i> <i>mimkin, imkin</i>	participant-internal necessity	need
	participant-external non-deontic possibility	possibility
	participant-external deontic possibility	permission
<i>ille</i> <i>lāzim</i>	epistemic possibility	feasibility
	participant-external non-deontic necessity	necessity
	participant-external non-deontic necessity	necessity
<i>la buddi</i>	participant external deontic necessity	obligation
	epistemic necessity	certainty
	participant-external non-deontic necessity	necessity
<i>axēr lē</i> <i>min la buddi</i>	participant external deontic necessity	obligation
	epistemic necessity	certainty
	participant-external deontic necessity	advice
<i>bukūn</i>	epistemic possibility	feasibility
		near-certainty
	epistemic necessity	logical deduction

This summary gives rise to a number of questions concerning the relationships between semantic values and their formal encoding as well as about the functional overlapping of several modal items. The question could be raised whether



the semantic development of modal meanings in KBA adheres to the ‘unidirectionality hypothesis’ (Bybee et al. 1994: 12–14) according to which deontic meanings precede epistemic ones. The best tool for describing such functional developments is a semantic map that can visualize regular relationships between two or more meanings as expressed by different linguistic forms. The following semantic map of modality in KBA is largely based on the model proposed by van der Auwera & Plungian (1998: 98–100). It shows connections between lexical and modal values as well as the connections between modal values themselves.

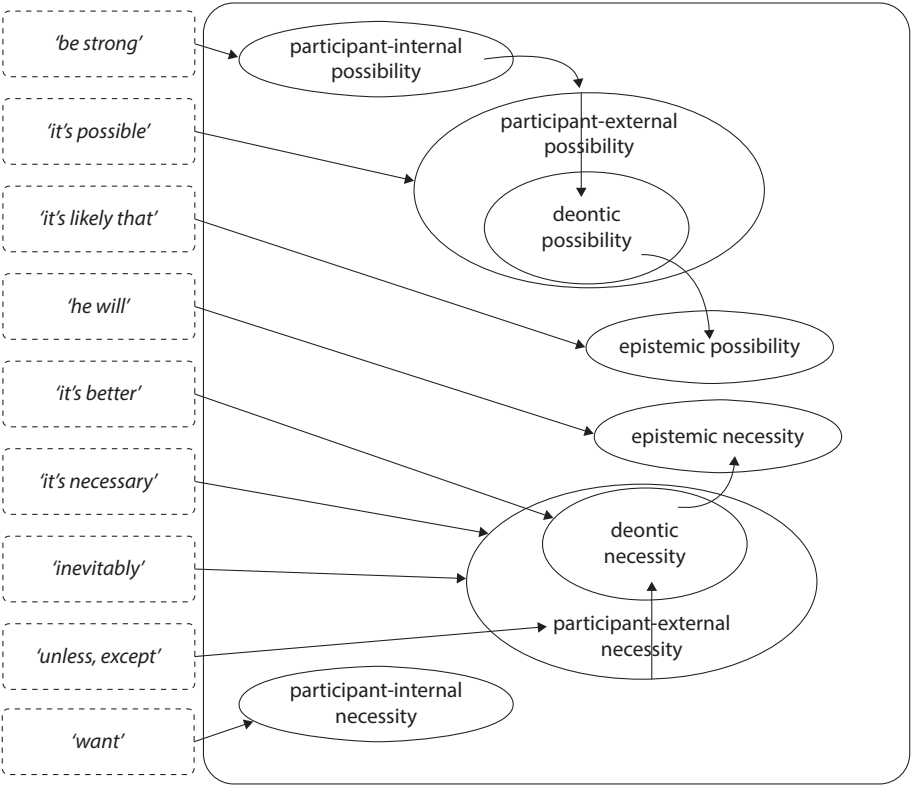


Figure 2. Semantic map of modality in KBA

Lexical source categories are placed to the left of the map, whereas grammatical categories are to the right. Grammaticalization proceeds from left to right, with the most grammaticalized categories at the right end. The map highlights the relationship between lexical items and meanings/functions by means of a line, while spatial adjacency gives evidence of similarity between meanings (Narrog & van der Auwera 2011: 323). Furthermore, ovals allow the representation of semantic

specialization/generalization. On that account, we can note that, in line with van der Auwera & Plungian's maps of modality, KBA presents the following path of semantic development in the area of possibility: 'participant-internal possibility' > 'participant-external (non-deontic) possibility' > 'participant-external deontic possibility' > 'epistemic possibility', where *gidir*, *b=i-gdar* "be strong" cannot introduce epistemic values, *mimkin* and *imkin* "be possible" cannot introduce a participant-internal modality, and *min la buddi* "it's likely that" is specialized in the expression of an epistemic possibility. In the same manner, the semantic map shows a direct relationship between the invariable *bukūn* "he will" and the expression of an epistemic necessity (i.e., "must") and this indeed conforms to the path of development 'future' > 'epistemic necessity' proposed by Van der Auwera & Plungian. On the other side, KBA diverges from van der Auwera & Plungian's representation of modality with respect to the fact that it displays a semantic redetermination from general participant-external possibility (as expressed by *lāzim* "it's necessary" and *la buddi* 'inevitably') to its subtype of deontic possibility. This can be a consequence of the abundance of modal items expressing a participant-external necessity in KBA. Furthermore, dialect levelling towards Sudanese Standard Arabic had a significant impact on the expression of modality in KBA as showed by the semantic expansion of *mimkin*, *imkin* "it's possible" and *lāzim* "it's necessary" into the domain of epistemic modality. As a final remark, in line with other eastern Arabic dialects (Vanhove et al. 2009), KBA shows a high degree of specialization of the auxiliary *dāyir* "want" that exclusively expresses a participant-internal necessity.

### List of symbols and glosses

-	affix boundary	IND	indicative
=	clitic boundary	INDF	indefinite
\	ablaut	M	masculine
1,2,3	1st, 2nd, 3rd person	NEG	negation
ACT	active	PASS	passive
ANT	anterior	PL	plural
CS	construct state	POSS	possessive
DEF	definite article	PROX	proximal
DIST	distal	PTCP	participle
EXS	existential	REL	relative
F	feminine	SG	singular
FOC	focus marker	VOC	vocative

## References

- Ali, Mustapha & Catherine Miller. 1986. "Aperçu sur le système verbal du dialecte du centre-nord soudanais: l'expression de la concomitance". *Matériaux Arabes et Sudarabiques* 2.167–187.
- Behnstedt, Peter. 1985. *Die nord-jemenitischen Dialekte*. Wiesbaden: Dr. Ludwig Reichert.
- Bybee, Joan & Suzanne Fleischman. 1995. "Modality in Grammar and Discourse: An Introductory Essay." *Modality in Grammar and Discourse*, ed. by Joan Bybee and Suzanne Fleischman, 1–14. Amsterdam: John Benjamins. doi:10.1075/tsl.32
- Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World*. Chicago: University of Chicago Press.
- Cohen, David. 1984. *La phrase nominale et l'évolution du système verbal en sémitique. Études de syntaxe de syntaxe historique*. Paris: Peeters.
- De Haan, Ferdinand. 2006. "Typological Approaches to Modality." *The Expression of Modality*, ed. by William Frawley, 27–70. Berlin: Mouton de Gruyter.
- Eksell, Kerstin. 2006. "The Origin and Development of the Cursive *b*- Imperfect in Syrian Arabic". *Change in Verbal Systems: Issues on Explanation*, ed. by Kerstin Eksell & Thora Vinther, 73–98. Frankfurt am Main: Peter Lang.
- Frajzyngier, Zygmunt. 2002. "System Interaction in the Coding of Modality." *Modality*, ed. by Frits Beukema & Sjeff Barbiers, 165–184. Amsterdam: John Benjamins.
- Ingham, Bruce. 1994. "Modality in the Arabic Dialect of Najd". *Actes des Premières Journées Internationales de Dialectologie Arabe*, ed. by D. Caubet & M. Vanhove, 185–200. Paris: Publications Langues'O.
- Manfredi, Stefano. 2010. A grammatical description of Kordofanian Baggara Arabic. Napoli: Università degli Studi di Napoli 'L'Orientale' dissertation.
- Manfredi, Stefano. 2012. "Dialect Mixing and Dialect Levelling in Kordofanian Baggara Arabic (Western Sudan)". *Dynamiques langagières en Arabophonies: variations, contacts, migrations et créations artistique*, ed. by Alexandrine Barontini, Christophe Pereira, Angeles Vincente and Karima Ziamari. *Collección estudios de dialectología árabe*, 141–162. Zaragoza: Universidad de Zaragoza.
- Mitchell, Terence F. & Shahir al-Hassan. 1994. *Modality, Mood, and Aspect in Spoken Arabic with Special Reference to Egypt and the Levant*. London: Paul Kegan.
- Narrog, Heiko & Johan van der Auwera. 2011. "Grammaticalization and Semantic Maps". *The Oxford Handbook of Grammaticalization*, ed. by Heiko Narrog and Bernd Heine, 318–327. Oxford: Oxford University Press.
- Nuyts, Jan. 2005. "The Modal Confusion: On Terminology and the Concepts behind It". *Modality: Studies in Form and Functions*, ed. by Alex Klinge & Henrik H. Müller, 5–38. Oakville: Equinox Publishing.
- Nuyts, Jan. 2006. "Modality: Overview and Linguistic Issues." *The Expression of Modality*, ed. by William Frawley, 1–26. Berlin: Mouton de Gruyter.
- Owens, Jonathan. 1993a. *A Reference Grammar of Nigerian Arabic*. Wiesbaden: Harrasowitz.
- Owens, Jonathan. 1993b. "Nigerian Arabic in Comparative Perspective". *Sprache und Geschichte in Afrika* 14.85–175.
- Palmer, Frank R. 2001. *Mood and Modality*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139167178
- Palva, Heikki. 2008. "Sedentary and Bedouin Dialects in Contact: Remarks on Karaki and Salti (Jordan)". *Journal of Arabic and Islamic Studies* 8.53–70.

- Persson, Maria. 2008. "The Role of the *b*- Prefix in Gulf Arabic Dialects as Marker of Future, Intent and/or Irrealis". *Journal of Arabic and Islamic Studies* 8:4.26–52.
- Reichmuth, Stefan. 1983. *Der arabische Dialekt der Shukriyya im Ostsudan*, Zürich: Georg Olms.
- Rosenhouse, Judith. 2006. "Bedouin Arabic". *The Encyclopaedia of Arabic Language and Linguistics*, vol. 1, ed. by Kees Versteegh, Mushira Eid, Alaa Elgibali, Manfred Woidich & Andrzej Zaborski, 259–269. Leiden: Brill.
- Roth-Laly, Arlette. 1979. *Esquisse grammaticale du parler arabe d'Abbéché*. Paris: Geuthner.
- Simeone-Senelle, Marie-Claude & Martine Vanhove. 2003. "Le fonctionnement d'auxiliaires en afar". *Mélanges David Cohen. Etudes sur le Langage, les Langues, les Dialectes, les Littératures, Offerts par ses Élèves, ses Collègues, ses Amis. Présentés à l'Occasion de son Quatre-Vingtième Anniversaire*, ed. by Jérôme Lentin and Antoine Lonnet, 615–634. Paris: Maisonneuve et Larose.
- Tosco, Mauro. 1995. "A Pidgin Verbal System: The Case of Juba Arabic". *Anthropological Linguistics* 37:4.423–459.
- Van der Auwera, Johan & Vladimir A. Plungian. 1998. "Modality's Semantic Map". *Linguistic Typology* 2.79–124. doi:10.1515/lity.1998.2.1.79
- Vanhove, Martine, Miller, Catherine, & Dominique Caubet. 2009. "The Grammaticalization of Modal Auxiliaries in Maltese and Arabic Vernaculars of the Mediterranean Area". *The Grammaticalization of Modal Particles*, ed. by Johan Van der Auwera, 325–362. Berlin: Mouton de Gruyter.

**PAGE PROOFS**  
© JOHN BENJAMINS PUBLISHING COMPANY

**2nd proofs**

PART III

## Predication and beyond

**PAGE PROOFS**  
© JOHN BENJAMINS PUBLISHING COMPANY

**2nd proofs**

**PAGE PROOFS**  
© JOHN BENJAMINS PUBLISHING COMPANY

**2nd proofs**

# Insubordination in Modern South Arabian

## A common isogloss with Ethiosemitic?

Olga Kapeliuk

Hebrew University, Jerusalem

The term ‘insubordination’ is used to describe the case of a subordinate verb becoming the main verb of a sentence. In Modern South Arabian (MSA) a relative verb without an antecedent may act as the finite verb of an independent sentence. The MSA main relative verb is analyzed as the equivalent of an active participle together with copula zero, which leads to a comparison with Amharic and other modern Ethiosemitic (ES) languages in which a relative verb, followed by the copula – mandatory in these languages – is often used as the main predicate. After a review of the main points of convergence between modern ES and MSA in general, the function of the ES relative, plus copula, is analyzed as static predication as opposed to the dynamic predication of the ordinary verb. It appears in introductory passages of a literary text where the characteristics of the main character are presented.

**Keywords:** relative verb, predicalization of relative verbs, common features between MSA and MES

In his two papers *Le forme verbali pseudorelative: isoglossa strutturale del semitico sudoccidentale* (1993) and *L'impiego di frasi pseudorelative come verbi finiti* (2007), Pennacchietti deals with constructions in Modern South Arabian (henceforth MSA<sup>1</sup>) in which headless relative verbs in the perfective or imperfective verb forms, subordinated by the relative particle *d*, or what he calls ‘pseudorelative’, act as predicates of independent sentences. This is a classical case of ‘insubordination’ as defined by Evans in his paper “Insubordination and Its Uses”<sup>2</sup> where this term

---

1. List of abbreviations: Am – Amharic; ASA – Ancient South Arabian; CA – Classical Arabic; ES – Ethiosemitic; Gz – Gəʿəz; Gr – Gurage; Hr – Harari; KK – Kurmanji Kurdish; Me – Mehri; MES – Modern Ethiosemitic; MSA – Modern South Arabian; NA – Neo-Aramaic; Se – Semitic; SWS – South-West Semitic; Ta – Tigrinya; Te – Tigre.

2. I am indebted to Prof. Eran Cohen from the Department of Linguistics of the Hebrew University in Jerusalem for having drawn my attention to this important article.



is applied to “conventionalized main clause use of what, on *prima facie* grounds, appear to be formally subordinate clauses” (Evans 2007: 367). A good example of the insubordination process in Modern Ethiosemitic (MES) may be found in the actual use of the Tigrinya (Ta) gerund. Whereas in Gəʿəz (Gz) the gerund is a conjugated verb form which can only be used in a subordinate circumstantial clause and in Amharic (Am), the same gerund either acts as a subordinate verb form or demands a finite auxiliary in order to become the verb of a main sentence, in Ta the gerund alone, with no auxiliary, has become an independent verb, used freely in main sentences, especially in the spoken language, and indicating the recent past.

Cases of the relativized verb in insubordination in MSA had been analyzed exhaustively by Wagner 1953: in his *Syntax der Mehri Sprache*, based on texts collected by members of the scientific mission of the Imperial Austrian Academy of Sciences to South Arabia at the end of the 19th century. Pennacchietti tends to agree with Wagner who basically identifies such relative verbs of MSA as a replacement of the active participle, given that the original active participle was reduced in these languages to the status of a pure nominal (Wagner 1953: 120–121). On the other hand, Aaron Rubin, in his recent book *The Mehri Language of Oman* (2010: 143–149) based on texts collected by Johnstone and published by Stroemer (1999), tends to follow in the footsteps of Johnstone (1975: 1987: XIX), Simeon-Senelle (1997: 247–250) and Lonnet (2005) in identifying the construction as basically verbal, providing progressive or actual nuance to a past or present action.

Using argumentation involving Classical Arabic, Gz, and some Neo-Aramaic (NA) constructions with the relative particle *d-*,<sup>3</sup> Pennacchietti arrives at the important conclusion that the insubordinate use of the relative verb in MSA implies the presence of a zero copula (Pennacchietti 1993: 219, 2007: 141). Absence of an explicit copula is a common feature of MSA since these languages have no regular copula in the present tense and use freely nominal sentences. Consequently if a verb with the relative particle acts as the only verb in a sentence it can be analyzed as the main verb, composed of the relative verb corresponding to a participle and of a zero copula. The interesting point is that not far away to the West of the territory where MSA languages are spoken, just across Bab al-Mandeb in the Horn of Africa, we often encounter MES languages in which relativized verbs, followed by an explicit copula, form the predicate of independent sentences.<sup>4</sup> Considering that,

3. The NA construction led Pennacchietti to a typological comparison with relative clauses in insubordination in some dialects of Kurmanji Kurdish (KK). A particularly interesting use of the particle *d-* in insubordination has been revealed by Borghero (2015) in the NA dialect of ‘Ankawa. In this dialect a progressive present tense and a present perfect tense are created by prefixing the relative particle *d-* to the present and the past participle, apparently under the impact of KK.

4. A parallel between the frequent use of relative verbs in MSA and in such MES languages as Am or Harari (Hr) was mentioned for the first time by Wagner (2001: 341–342).

contrary to MSA, verbless nominal sentences don't exist in MES, we may surmise that a relative verb in an independent sentence, whether accompanied by a copula or without it, forms a common isogloss of 'semitico sudoccidentale'. In principle this shouldn't be surprising given that South Arabia is normally indicated as the place of origin of the Semitic migrants to East Africa but, considering that such constructions are not attested in Ancient South Arabian (ASA) and, anyway, that a direct relation between MSA and ASA is now strongly denied, or at least doubted by many experts in the field, the similarities on the synchronic level between MSA and MES raise certain serious questions about classification and chronology which still await their answers.<sup>5</sup>

Among additional common features between MSA and ES, ancient or modern, we may quote the glottalization of consonants pronounced emphatically in the other branches of Se; the similarity between the vocalic system of Gz and Mehri (Me) with five long phonemic vowels in the former and five or six in the latter (long *ē* being often only a variant of *ā*; see Johnstone 1987: XIII) and only two short phonemic vowels *a* and *ə* in both; the (unusual for MES) palatalization of the suffix of the 2nd person feminine singular *-ki* as *š* both in the southern branch of MES and in Me;<sup>6</sup> the transfer of the labialization into the root in the 3rd person plural in the perfective in Me, e.g., *'āmōr* "he said" but *'āmāwr* "they said" and in many similar cases in Gurage (Gr); the forms of the imperfective in the first stem, type A, with a vowel after the first radical, such as Gz *qatala/yəqattəl* against the vowelless jussive *yəqtəl* and Me *kātōb/yəkōtəb/yəktəb* (Voigt 1994;<sup>7</sup> Lonnet 2005: 187–188); the presence of the vowel *ī* before the pronouns suffixed to the plural of nouns in Gz and in the dual and the plural in Me; the creation of an independent possessive pronoun by preposing the relative particle to independent personal pronouns, e.g., Me *q-ho* "mine" (Rubin 2010: 33) and Am *yä-ne* (Leslau 1995: 56); the wealth of forms of the broken plural in addition to those attested in the ancient languages both in MSA and in Tigre (Te) and in Ta, and even such idiomatic expressions as the use of the interrogative pronoun "who" when asking about the name of a person, unknown elsewhere in Semitic (Kapeliuk 2017: 58), e.g., Gz *mannu səmə-ka* "What (lit. who) is your name?", Ta *mān šəm-ka?* "id.", Am *səm-əh man təbbalalläh?* "What name (lit. who name) are you called (M.SG)?", Gr (Čäxa) *šəm-x' m'an yəurk'* "What name (lit. who name) one calls you (F.SG.)?" and in Me *hāmm-ok mōn?* "What (lit.

5. For a comprehensive treatment of the origin of ES and its classification see Bulakh & Kogan (2013: 13–141).

6. A palatalized *k* in MSA results in *š* but it is rare in MES. In Am *k* is rarely palatalized at all and only dialectally and resulting in *č*, not in *š* (Cohen 1936: 35). Only the Am suffix of the 2nd p. f. sg. has the sound *š*.

7. This paper is a thorough comparative investigation into the morphology of the Me and Gz verb systems; see also Zaborski (2007: 198–199) and (Wagner 1993).

who) [is] your name?” (Nakano 1986: 52), or the use of “one” with the meaning of “the other” as in Examples (1a) and (1b):

- (1a) Me *’amma táyt, šarūt; w-’amma táyt wəkəbūt bərk amkōn* (Stroomer 1999: 2/3)<sup>8</sup>  
 “As for one, she stood outside; as for the other (lit. one), she went into the place.”
- (1b) Am *and-u s-iwädq and-u yənnäss-all* (Kane 1990: 1230)  
 “When one (lit. the one) falls the other (lit. the one) rises.”

As for the verbs subordinated by the relative particle *d*, or in short relative verbs, in MSA and by parallel particles in ES,<sup>9</sup> they have received in both sub-branches of South-West Semitic (SWS) various and much diversified functions and attained great frequency not attested in the other branches of Se.<sup>10</sup> As nominalized forms of the verb they may play the role of *nomina agentis*, of certain adjectives, of verbal abstract nouns in cleft sentences, and of *that* clauses. This phenomenon in MSA has been attributed to the restricted use of participles and their replacement by relativized verbs (Wagner 1953: 120–122). We encounter almost identical phenomena also in Gz where original morphological participles are mainly used as substantives, e.g., Gz *qatāli* “murderer” (Dillmann 1865: 441; Kapeliuk 2003: 177–179, 2009: 224–226). The great frequency and diverse functions of the relative constructions in Gz did not escape the attention of linguists already in the 19th century. Both Dillmann (1907: 527–542 = § 200–203) and Praetorius (1886: 29–33) pointed out the unusually common use of relative clauses in Gz as equivalents of Se adjectives and participles and the occasional use of *za-* as the conjunction *that*. Praetorius even imputed the scarcity of adjectives and participles in Gz to the fact that “der Gebrauch von Relativsätzen ist im Äthiopischen ausserordentlich beliebt” (1886: 31–32).<sup>11</sup> Thus we find in Gz certain relative verbs which are already lexicalized as adjectives and are presented as such in the dictionaries, e.g., *za-yənə’əs* “small, minor, lesser”, *za-yəmaššə’* “future”, *za-’i-yəməwwət* “immortal” etc. (Leslau 1987). This trend was strengthened with time, and in his *Amharische Grammatik* Josef Hartmann lists not less than 40 such “Verbaladjektivische Relativformen” (1980: 238–239; Edzard 2001).<sup>12</sup> This tendency is shown in the Am rendering of the Gz adjective in Example (2a):

8. The examples from Stroomer 1999 are quoted by text and sentence number.

9. For the relative particles in ES in general see Bausi (1990).

10. On relative verbs, besides their usual function as attributes of nominal headnouns, see Wagner (1953: 116–120) for Me, Garad & Wagner (1998: 221–270) for Hr, and Kapeliuk (1988: 70–100, 1994: 69–75) for Am.

11. “The use of relative clauses in Ethiopic is unusually liked” (my translation).

12. “Verbo-adjectival relative forms” (my translation).

- (2a) Gz wa-’am-za’ aḥaza ’Iyasus yaḥis-on la-’ahgur ’alla ba-wəstet-on gabra ḥayla  
*bəzuḥa* (Matthew 11/20)  
 “And then Jesus started to rebuke the cities, in which He made many wonders.”
- (2b) Am bā-zziyan gize yāmm-ibāzu (lit. which are many<sup>13</sup>) tā’ammərat yātādārrägä-  
*bb-aččāw-n kätāmawočč ... l-inäqf-aččāw ḡāmmärä* (Matthew 11/20)  
 “And then Jesus started to rebuke the cities, in which He made many wonders.”

But far more intriguing in MSA is the use of headless relative verbs in insubordination, as predicates of main sentences. As mentioned above, we learn from most descriptions of these languages, and of Me in particular, that in independent sentences verbs in the imperfective with the relative particle *d* indicate an actual or progressive present or a progressive past tense and that with the perfective they render some kind of present perfect, functions probably originating, according to Pennacchietti (1993: 213–219), in constructions of concomitance. Many examples in Harry Stroomer’s edition (1999) of Thomas Johnstone’s Me texts, which were recorded in the late 1960s and early 1970s, confirm the conclusions of the Austrian Expedition (Wagner 1953) and prove that this aspect of the language has changed little during more than half a century. A few remarks may, however, be added from the contextual and pragmatic point of view.

The Me relative verb may be the only verb of the sentence or one of two verbs or more. When accompanying, in narration, another independent verb, the relative verb often indicates a concomitant action, somewhat like the constructions with *waw al-ḥāl* of Arab grammarians, or with an asyndetic imperfective or with an accusative participle of *ḥāl* in CA (Wagner 1953: 120–121; Pennacchietti 2007: 143–145; Rubin 143–145), as shown in (3)–(4) with the imperfective and in (5) with the perfective. Strictly speaking what is concomitant is not always a real action but rather the state in which the subject of the ‘main’ verb finds himself during its action. The nuance of state is particularly prominent with relative verbs accompanying the verbs “to find that ...”, “to see that...”, the construction meaning literally “he found him who ...”, “he saw him who ...”, as in (6)–(7), but with other verbs, such as verbs of motion, it is also quite frequent, as in (3)–(5). With enunciative verbs, on the other hand, *d*- introduces a content clause<sup>14</sup> (Rubin 2010: 291–292) as in (8), even a verbless one as in (9):

- (3) Me w-aḡāyḡ rəd tawōli sekənəḥ *d-ikətōmāḥ* w-a-*d-ixtəyōb* (12/14)  
 “And the man went back to his encampment, despairing and disappointed.”
- (4) Me w-a-sōfər aḡəggən *d-a-yəḥōm* yəsīr h-arḥəbət *d-bīs tēt* (22/37)  
 “The boy went away willing to go off to the country in which the woman was.”

13. In the literal translation of the examples the original word order is not always respected.

14. Similar to Aramaic and post-biblical Hebrew.

- (5) Me *wəʃələm tīn káwb d-śálūl bəkəwbēt* (22/85)  
 “There has reached us a dog carrying a bitch.”
- (6) Me *tōli kūsəm təh əd-ğəyōb, ’əl bih ḥas lā* (17/8)  
 “Then they found him fainted. He was not conscious.”
- (7) Me *wə-kūsa ḥáybəh d-áywər* (24/50)  
 “And he found his father blind (having gone blind).”
- (8) Me *aškərək bə-ḥnōfi d-ho dāləmk* (19/24)  
 “I confess of myself that I have acted unjustly.”
- (9) Me *ğarəbk dā hāh ašədáyki məxləs šay* (18/17)  
 “I knew that he [was] my sincere friend.”

We find a somewhat similar case in ES, in Gz and in all the MES languages, with the verb *masala* “it seems” used impersonally, which is basically the only verb in ES to be consistently completed by a relative clause (Kapeliuk 1981, Kapeliuk 2017: 579), e.g., (10)–(13). The verb *mäsälä* in Am is sometimes used personally, and thus we may see the real value of the relative form (14). The same may be said about the verb *tayyā* “to be seen”; in Example (15) it is accompanied by *hono*, the gerund of the copula, to strengthen the predicative link. This example is significant because here, in MES, the zero copula discerned by Pennacchietti in the MSA relative constructions becomes explicit:

- (10) Gz *’asma masal-o za-’amsaṭu muquḥān* (Actes 16/27)  
 “For it seemed to him that the prisoners have escaped (lit. it seemed to him the prisoners [were] who escaped).”
- (11) Am *əsräññoč-u y-amälläṭu mäslo-t* (id.)
- (12) Ta *’atom ’asurat zə-mäläqu mäsila-wwo* (id.)
- (13) Gr *yä-näffägätt-u mäsälä-m* (Hetzron 1977: 99)  
 “It seemed that she was begrudging him.”
- (14) Am *ləğagäräd-wa-mm... and ṭəru nägär əndä-mmigätm-at täsfə y-alla-t təməsl-alläčč* (quoted in Kapeliuk 1981: 52, 2009: 320)  
 “The girl seems [as one] who hopes that a good thing will happen to her.”
- (15) Am *kä-ñña huneta yä-sswa yä-käffa hono yəttayy-all* (id. 56 [323])  
 “Her situation looks worse than ours (lit. from our situation hers is seen being which is bad).”

When we turn to Me sentences in which the relative verb is the only verb of an independent sentence, we should keep in mind, in our comparison with MES, the zero copula as reconstructed for MSA by Pennacchietti. As pointed out above, MES does not allow verbless sentences; consequently the very common Am construction

in which the relative verb followed by an explicit copula functions as the main and only verb of a sentence may be considered as an equivalent of the MSA construction under discussion, and the same may be said of Ta (Kapeliuk 1980a: 100–103, 2009: 313–316; Kapeliuk 1988: 146–155, 1980b: 19–20, 2009: 470–472) and probably of all the other MES languages as well. This construction may easily be compared on the formal level with the MSA constructions containing a relative verb with zero copula in a main sentence. To define the true nature of this construction in MES, let's look at a few sentences from Am and Ta texts in which the opposition between a regular verb and a verb decomposed into its relative form and the copula is exemplified, as in (16)–(17); even a change of subject is allowed in (18)–(20):

- (16) Am *səmməntānna kəfəl yä-čərräsäčč nat, əsswa gən yəmmə-ttawāra-w asra hullätānna čərrəšš-allāhu əyy-aläčč nəw* (quoted in Kapeliuk 1988: 151)  
 “She finished the 8th grade (lit. she is who finished) but it is ‘I have finished the 12th’ that she proclaims.”
- (17) Am *planetočč bə-šəhay zuriya yəmm-izoru māhon-aččāw kə-kopernik gəmməro yä-tawwāqə nəw ... planetočč bəšəhay zuriya sələ-mm-izoru šəhay yä-fəträt əmbərt nəčč* (id. 146)  
 “It is known since Copernicus that planets revolve around the sun (lit. their being who revolve is known) ... . Since the planets revolve around the sun, the sun is the center (lit. navel) of the universe.”
- (18) Am *mannəmm b-irəgətəh yəmm-a-yəqʷäräqqʷərə-h nəh. -ləməñ? -dəngay nəh. a-yəqʷäräqqʷərū-h-mm əñgi təqʷäräqqʷər-alläh* (id. 151)  
 “‘If anyone tramples on you, you are not hurt (lit. you are who one doesn’t hurt you)’. – ‘Why?’ – ‘You are a stone. One does not hurt you, rather you hurt.’”
- (19) Ta *səb bə-kəlattə ‘aynät ‘aššəwawta yəššawät... ‘əti nay təmhərti šəwāta gəna bə-bəlhat-n bə-‘aggəbab-ni z-iššawät ‘iyu* (quoted in Kapeliuk 1980b: 20)  
 “Men play in two kinds of games.... But educational game is played (lit. is which is played) with intelligence and a proper conduct.”
- (20) Ta *nab sə rəhəz-ä-yə... afəru səbat tə‘ə nna-om zə-təmalä’ä’yyu nə -malät yä-šəggəmə’yyu* (Yediot Negat no. 52, March 2009)  
 “It is hard to say that the health of people who don’t go to work is perfect (lit. is which was accomplished).”

These constructions seem to be formally similar in MES and in MSA, but on the syntactico-semantic level they differ at first sight. Whereas in MSA, according to most experts, they render an actual or progressive action, in MES they define the state or the properties of the subject. In MES the opposition between a regular verb and a verb decomposed into its relative form and the copula consists in the nature of the predication (Kapeliuk 1980a: 100–103, 1980b: 19–21, 1988: 146–155,



2009: 313–316, 472–474). The predication of the regular verb is usually conceived under a dynamic angle, whereas in the decomposed construction it is mostly static and corresponds to the definition of the subject as somebody or something characterized by the properties inherent in the meaning of the verb. Here the attributive nature of the relative verb as a possible replacement of the participle is still recognizable. When I described this construction for the first time at the 5th International Conference of Ethiopian Studies in Nice in 1977, I suggested that it may be considered as a living confirmation of the Aristotelian definition of the verb, namely that *'anthropos badízei* “a man walks” equals *'anthropos badízon 'esti* “a man is walking, un homme est marchant” (Sandmann 1954: 11–12; Aristoteles 1963: Chapter 12; Kapeliuk 1980a: 100–103, 2009: 313–316). The combination of the relative verb together with the copula is mostly found in MES in the written language, in literary texts and in the language of the press. In works of fiction it is usually encountered in initial descriptive passages which provide the framework within which the narration of the story will proceed. It is particularly common in sections in which a character is introduced for the first time or a new location or situation is depicted. In the language of the press it supplies the background data and the setting of an event. It corresponds to what is called in cinematic technique ‘an establishing shot’.

As for MSA, at least in its concomitant use beside another verb, the relative verb often also has the meaning of state, be it temporary, continuing as long as the action of the main verb lasts. In general we may say, based on examples collected from Stroomer’s book, that in Me narration the use of an independent relative verb not connected somehow to another verb is not so frequent, but when it does it may stand at the opening of a text as in (21)–(22) or introduce a change in the situation as in (23), thus reminding us of the Am use:

- (21) *Me w-ámma kəwɪt dā-bā nəwās, nəhōr ɬayt d -isyūr, wə-šəh dəgərīt ɬayt* (1/1)  
 “As for the tale of Abu Nuwās one day he was traveling and he had one bean.”
- (22) *Me xəɬərāt gayg d-yəgáwlək mən səkənəh* (32/1)  
 “Once a man was looking for his community.”
- (23) *Me ɖomeh aɡərōy dā-ħəynít wə-sənnáwrət thámsən. wə-ləkən sēh d-əktəwbōt tēt sáħrət* (6/9)  
 “This [was] the talk of the women while the cat was listening. But she turned into a sorceress.”

Most of the Me examples, however, are in direct speech and dialogue, and I have the feeling that many of them could be interpreted as cleft sentences. It should be stressed that there is no restriction on using this construction even in languages that do not use a copula (Polotsky 1944: 66) as is the case, for instance, in Gz in positive cleft sentences (Kapeliuk 1985: 201–204, 2009: 195–198) both in statements

(24)–(26) and in questions (27) and (28). In the MES languages such sentences carry a mandatory copula, as in (29)–(31):

- (24) Gz *ba'anta zəntu tasfā za-yəṯhassas-omu 'ayhud* (Acts 26:7)  
 “[It is] because of this hope that the Jews are seeking to accuse them.”
- (25) Gz *'əmənna tə'əyanta 'əsra' el 'ana za-dəḥənku* (II Samuel 1:3)  
 “[It is] I who escaped from the camp of Israel.”
- (26) Gz *wa-'aḥzāb-ni za-yəzabbəḥu la-'agānənt wa-'akko la-'əgzi' abḥer* (I Corinthians 6:5)  
 “And [its is] to demons that peoples sacrifice and not to God.”
- (27) Gz *mannu za-yāsta' arrəyā la-ṣədq məsla ḥāṭi' at* (II Cor. 6:14)  
 “Who [is it] that compares justice with sin?”
- (28) Gz *anta-nu za-yəmaṣṣə?* (Matthew 11:3; Lucas 7:20)  
 “[Is it] you who will come?”
- (29) Am *yāmm-imāṭa-w anta nāh-nə?*  
 “Are you the [one] who will come?”
- (30) Ta *'əti zə-māṣṣə-ss nəssəxa d-ixa?*  
 “Are you the [one] who will come?”
- (31) Gr *bā-ruq gānṇ-u yā-čānṇāhu* (Hetzron 1977: 126)  
 “[It] is from a distant country that I came.”

In Stroomer’s texts many cases of Me direct speech may be translated by English cleft sentences, as in (32)–(35). Some of the examples are even better rendered by French thetic sentences (Pennacchietti 1993: 217 note 7)<sup>15</sup> in which, from a pragmatic perspective, the rheme is constituted by the whole relativized clause. Anyway, it is a subject to be explored:

- (32) Me *'āmāwr: 'kēf ḥályək təh līn wə-ṣərōməḥ 'āmərək: 'sīnək təh lā!' hēt d-ḥrəkək abáyrən!* (23/10)  
 “They said: ‘How did you describe it [the camel] to us and now you say I have not seen it! [It is] you who stole our camel!’” (French cleft: “[C’est] toi qui a volé notre chameau!” French thetic: “[C’est] que tu as volé notre chameau!”)
- (33) Me *ḍoməḥ hē aḡīgēn d-žāṭ atetk, wəlākən ḥōm aṣédk: mōn d-xəyūn būk?* (22/89)  
 “This was (lit. he)<sup>16</sup> the son who took your woman, but I want to check who [was it] that betrayed you” (French: “Qui [est-ce] qui t’a trahi”)

15. But not in English.

16. Here the personal pronoun *hē* “he” already acts as a kind of copula.



- (34) Me 'āmōr: *"hibō aḡərōy dōmāh? d-əḥtəwēk aw hibō?"* – 'āmōr: *"lā', al hō d-əḥtəwēk lā, wəlākən ḥōkəm yəḥōm dráyyət"* (20/5–6)  
 "He said: 'What kind of talk [is] this. [Is it] you who have gone mad or what?'"  
 (French cleft: "Est-ce toi qui es devenu fou ou quoi?" or Frenchthetic: "[Est-ce ] que tu es devenu fou ou quoi?") – "He said: 'No, [it is] not I who am gone mad'" (French cleft: "C'est n'est pas moi qui suis devenu fou mais..." or Frenchthetic: "[Ce n'est] pas que je suis devenu fou mais...") but the ruler desires an offspring."
- (35) Me 'āmōr *hə-bā nəwās: "hēt 'ār tṣəḥōk līn!"* – 'āmōr: *"we-kō? hō el hō d-əḥgəlōl lā?"* (36/25–26)  
 "He said to Abu Nuwas: 'You are just making fun of us' – "He said: 'Why? I, [is]n't [it] I who have lit the fire?'" (French cleft: "Pourquoi? N'[est-ce] pas moi qui ai allumé le feu?" or Frenchthetic: "N'[est] pas que j'ai allumé le feu?")

The extreme frequency of cleft sentences in ES is a well-known and prominent feature of this group of languages (Kapeliuk 1980b: 17–19, 2009: 465–470, 1985, 1988: 101–146, 2009: 285–294;). Since Cerulli's presentation entitled "Le mode relatif en couchitique" in June 1936 in Paris before a meeting of the GLECS (Cerulli 1934–1937: 61–63; Cerulli 1938: 97, 136; Cerulli 1951: 134–136 for branches of Cushitic other than Agaw), it was accepted among the Ethiopianists that the frequency of cleft sentences and other special relative constructions in ES has to be attributed to Cushitic influence, or in Marcel Cohen's words during the ensuing discussion: "le développement des constructions relatives [en éthiopien] est peut-être un fait 'africain'" (id.). Also Tubiana's statement on this subject (1960: 122; see also Leslau 1945: 76–78) is worth quoting:

Le substrat agaw de l'amharique explique très largement, sinon totalement, les traits non sémitiques de la syntaxe amharique. En ce qui concerne en particulier la phrase relative, on constate que toutes les phrases relatives de l'amharique peuvent avoir leur correspondant exact en agaw.<sup>17</sup>

I myself have adopted this explanation not only for the cleft sentence (Kapeliuk 1985: 191–194, 2009: 185–188) but also for the constructions with the relative verb followed by the copula discussed here (Kapeliuk 1980b: 19–22, 2009: 470–474; Kapeliuk 2002: 44–53, 2009: 419–428). I found clear evidence of the use of relative verbs in independent sentences in descriptions of such Agaw languages as Kemant (Conti-Rossini 1912: 78) where, according to Appleyard (1975: 345), "77% occurrences of the relative in the material [checked] is in main verb position"; in

17. "The Agaw substrate of Amharic largely, if not fully, explains the non-Semitic features of the Amharic syntax. For what concerns the relative clause in particular, one notices that all the Amharic relative clauses may have their exact parallels in Agaw" (translation by Olga Kapeliuk).

Awgni for the definite non-past (Hetzron 1969: 12, 1976: 56); in Xamir (Reinisch: 1884: 89); in Quara (Reinisch 1885–1887: 1.67–80); and above all in Bilin, where even constructions of the relative verb followed by the copula are attested (Reinisch 1882: 649). Similar cases also seem to be found in other branches of Cushitic.

I confess that reading Pennacchiotti's paper was a revelation to me. I was not aware of the existence of relative verbs in insubordination in MSA or in any other branch of Semitic. I was convinced that the case of MES was an areal characteristic of the Horn of Africa. His reconstruction of a zero copula in MSA allowed me to discern the common features of the two constructions in MSA and MES and, consequently, to include them both in a common 'semitico sudoccidentale' isogloss.<sup>18</sup> Now we have to face some new and very serious questions as to what is the real source of the relative constructions with the copula in MES. Is it South Arabia, and from there it wandered to Ethiopia? Was it attested already in any of the ASA languages? If so, why there is no sign of it in Gz? And above all, how can we explain the unusually frequent use of the relative verb in Agaw and other branches of Cushitic? Is it due to an independent evolution? To contamination? To an ancient Afroasiatic undercurrent?

## References

- Appleyard, David L. 1975. "A Descriptive Outline of Kemant". *Bulletin of the School of Oriental and African Studies* 37.316–335. doi:10.1017/S0041977X0014248X
- Aristoteles. 1963. *Categories and De Interpretatione*, translated with notes by J. L. Ackrill. Oxford: Clarendon Press.
- Bausi, Alessandro. 1990. "La frase relativa nelle lingue semitiche d'Etiopia". *Quaderni di Dipartimento di Linguistica. Firenze* 1. 105–123.
- Borghero, Roberta. 2015. "The Present Continuous in the Neo-Aramaic Dialect of 'Ankawa and Its Areal and Typological Parallels". *Neo-Aramaic and Its Linguistic Context*, ed. by Geoffrey Khan & Lidia Napiorkowska, 187–206. Piscataway, Gorgias Press.
- Bulakh, Maria, Leonid Kogan & Olga Romanova (eds.). 2013. *Semitic Languages: Ethiosemitic Languages*. Moscow, Academia: 13–141 (in Russian).
- Cerulli, Enrico. 1934–1937. "Le mode relatif en couchitique". *Groupe Linguistique d'Études Chamito-Sémitiques* 2.61–63.

18. In using the term 'insubordination' I deviate somehow from the rules established by Evans (2007: 370) for the consecutive stages leading to insubordination. According to him, the first stage consists in the elision of the element which lent full predicativity to the expression. In the case of Me there was no elision because the copula is a virtual one and, besides, my comparison takes place between two languages within the same isogloss and not between two stages of the same language.

- Cerulli, Enrico. 1938. *Studi Etiopici II – La lingua e la storia dei Sidamo*. Roma: Istituto per l'Oriente.
- Cerulli, Enrico. 1951. *Studi Etiopici IV – La lingua caffina*. Roma: Istituto per l'Oriente.
- Cohen, Marcel. 1936. *Traité de langue amharique*. Paris: Institut d'Ethnologie.
- Conti-Rossini, Carlo. 1912. *La langue des Kemantes en Abyssinie*. Wien: Kaiserliche Akademie der Wissenschaften.
- Dillmann, August. 1865. *Lexicon Linguae Aethiopicæ*. Leipzig: Weigel (reprinted in 1955).
- Dillmann, August. 1907. *Ethiopic Grammar*, translated by J. A. Crichton. London: Williams & Norgate.
- Evans, Nicholas. 2007. "Insubordination and Its Uses". *Finiteness: Theoretical and Empirical Foundations*, ed. by Irina Nikolaeva, 366–431. Oxford: Oxford University Press.
- Edzard, Lutz. 2001. "Adjektive und nominalisierte Relativsätze im Semitischen: Versuch einer Typologie". *New Data and New Methods in Afroasiatic Linguistics* ed. by Andrzej Zaborski, 39–52. Wiesbaden: Harrassowitz.
- Garad, Abdurahman & Ewald Wagner. 1998. *Harari-Studien. Texte mit Übersetzung, grammatischen Skizzen und Glossar* (Semitica Viva 18). Wiesbaden: Harrassowitz.
- Hartmann, Joseph. 1980. *Amharische Grammatik* (Aethiopistische Forschungen 3). Wiesbaden: Steiner.
- Hetzron, Robert. 1969. *The Verbal System of Southern Agaw*. Berkeley: University of California Press.
- Hetzron, Robert. 1976. "The Agaw Languages". *Afroasiatic Linguistics* 3:3.31–75.
- Hetzron, Robert. 1977. *The Gunnän-Gurage Languages*. Napoli: Istituto Orientale.
- Johnstone, Thomas M. 1975. "The Modern South Arabian Languages". *Afroasiatic Linguistics* 1:5.93–121.
- Johnstone, Thomas M. 1987. *Mehri Lexicon with English-Mehri Word List*. London: School of Oriental and African Studies.
- Kane Thomas L. 1990. *Amharic-English Dictionary*. Wiesbaden: Harrassowitz.
- Kapeliuk, Olga. 1980a. "Évolution de la phrase amharique : la nominalisation du verbe". *Proceedings of the 5th International Conference of Ethiopian Studies, Session A – Nice 1977*, 97–106. Rotterdam: Balkema. [id 2009: 310–318].
- Kapeliuk, Olga. 1980b. "Sur le rôle de la forme relative en tigrigna" – Review of R.M. Voigt: *Das tigrinische Verbalssystem*. *Bibliotheca Orientalis* 37:1–2.16–23 [id 2009: 463–479].
- Kapeliuk, Olga. 1981. "Il semble que' ou 'Il semble qui' – un problème de syntaxe amharique". *Studies Presented to H.J. Polotsky*, ed. by D. H. Young, 51–67. Gloucester: Pirtle & Polson. [id 2009: 319–331].
- Kapeliuk, Olga. 1985. "La phrase coupée en guèze". *Mélanges linguistiques offerts à Maxime Rodinson*, ed. by Christian Robin, 191–204. Paris: Guethner. [id 2009: 185–198].
- Kapeliuk, Olga. 1988. *Nominalization in Amharic* (Aethiopistische Forschungen 23). Wiesbaden: Steiner.
- Kapeliuk, Olga. 1994. *Syntax of the Noun in Amharic* (Aethiopistische Forschungen 37). Wiesbaden: Harrassowitz.
- Kapeliuk, Olga. 2002. "The Relative Verb in Amharic in Areal Perspective". *Afrikanistische Arbeitspapiere* 71.33–54 [id 2009: 408–429].
- Kapeliuk, Olga. 2003. "*Za-motna nəḥna wa-za-nagasna nəḥna*: Notes on Relative and Correlative Constructions in Gə'əz". *Aethiopica* 6.177–191 [id. 2009: 224–238].
- Kapeliuk, Olga. 2009. *Selected Papers in Ethio-Semitic and Neo-Aramaic Linguistics*. Jerusalem: The Hebrew University Magnes Press.

- Kapeliuk, Olga. 2017. "Studies in Ethiopian Languages, Literature, and History – Festschrift for Getatchew Haile, ed. by A. C. McCollum, 575–585. Wiesbaden: Harrassowitz.
- Leslau, Wolf. 1945. "The Influence of Cushitic on the Semitic Languages of Ethiopia". *Word* 1.59–82. doi:10.1080/00437956.1945.11659246
- Leslau, Wolf. 1987. *Comparative Dictionary of Ge'ez*. Wiesbaden: Harrassowitz.
- Leslau, Wolf. 1995. *Reference Grammar of Amharic*. Wiesbaden: Harrassowitz.
- Lonnet, Antoine. 2005. "Quelques réflexions sur le verbe sud-arabique". *Studi Afroasiatici – XI Incontro Italiano di Linguistica Camitosemitica* ed. by Alessandro Mengozzi, 187–202. Milano: Angeli.
- Nakano, Akio. 1986. *Comparative vocabulary of Southern Arabic*. Tokyo: Institute for the Study of Languages of Asia and Africa.
- Pennacchietti, Fabrizio A. 1993. "Le forme verbali pseudorelative: isoglossa strutturale del semitico sud-occidentale". *Semitica – Serta Philologica Constantini Tsereteli Dicata*, ed. by Riccardo Contini, Fabrizio A. Pennacchietti & Mauro Tosco, 213–225. Torino: Zamorani.
- Pennacchietti, Fabrizio A. 2007. "L'impiego di frasi pseudorelative come verbi finiti". *Relative e Pseudorelative tra Grammatica e Testo*, ed. by Federica Venier, 133–148. Alessandria: dell'Orso.
- Polotsky, Hans J. 1944. *Études de syntaxe copte*. Cairo: Société d'Archéologie Copte.
- Praetorius, Franz. 1886. *Aethiopische Grammatik*. New York: Ungar (reprinted 1955).
- Reinisch, Leo. 1882. *Die Bilin-Sprache in Nordost-Afrika*. Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften: Phil.-Hist. Classe: 9.583–718.
- Reinisch, Leo. 1884. *Die Chamirsprache in Abessinien*. Wien: Kaiserliche Akademie der Wissenschaften.
- Reinisch, Leo. 1885–1887. *Die Quarasprache in Abessinien*. Wien: Kaiserliche Akademie der Wissenschaften.
- Rubin, Aaron D. 2010. *The Mehri Language of Oman* (Studies in Semitic Languages and Linguistics 58). Leiden: Brill.
- Sandmann, Manfred. 1954. *Subject and Predicate*. Edinburgh: Edinburgh University Press.
- Simeone-Senelle, Marie-Claude. 1997. "The Modern South Arabian Languages". *The Semitic Languages*, ed. by Robert Hetzron, 378–423. London: Routledge.
- Stroomer, Harry. 1999. *Mehri Texts from Oman* (Semitica Viva 22). Wiesbaden: Harrassowitz.
- Tubiana, Joseph. 1960. A Remark in a Discussion. *Proceedings of the 1st International Conference of Ethiopian Studies – Rome April 1959*. Rome: Accademia Nazionale dei Lincei.
- Voigt, Rainer. 1994. "Neusüdarabisch und Äthiopisch". *Arabia Felix: Festschrift Walter W. Müller zum 60. Geburtstag*, ed. by Rosemarie Richter, Ingo Kottsieper & Mohammed Maraqtan, 291–307. Wiesbaden: Harrassowitz.
- Wagner, Ewald. 1953. *Syntax der Mehri-Sprache*. Berlin: Akademie Verlag.
- Wagner, Ewald. 1993. "Gedanken zum Verb des Mehri aufgrund der neuen Materialien von Johnstone". *Zeitschrift für arabische Linguistik* 25.316–339.
- Wagner, Ewald. 2001. "Neues Material zum Studium des Neusüd-arabischen: Die Stroomer'sche Edition der Mehri-Texte von Johnstone". *Wiener Zeitschrift für die Kunde des Morgenlandes*. 91.331–346.
- Yediot Nogat. A bi-monthly magazine in Hebrew, Amharic and Tigrinya published in Jerusalem by the Israeli daily Yediot Aharonot.
- Zaborski, Andrzej. 2007. "Notes on the Reconstruction of the 'Prefix-Conjugation' in the Modern South Arabian Languages". *Lingua Posnaniensis* 49.196–206.

**PAGE PROOFS**  
© JOHN BENJAMINS PUBLISHING COMPANY

**2nd proofs**

# Possessive and genitive constructions in Dahālik (Ethiosemitic)

Marie-Claude Simeone-Senelle

LLACAN–INALCO, CNRS. France


There have been few comparative studies of possessive or genitive relationships within the Modern South Arabian (MSA) group, but Dahalik, an Ethiosemitic language, had never been the subject of any study on this topic. In fact the language itself was completely unknown until 1997 and could not be fully investigated because of the inaccessibility of the region since 2006. This article attempts to give details on these constructions in Dahalik to compare the different strategies for determining the noun in some Southern Semitic languages of Arabia and the Horn of Africa.

By providing such comparisons this chapter highlights the common features of these languages and the characteristics of Dahalik within the Southern Semitic group as a specific Afrosemitic language.

All the Dahalik data have been collected during my fieldworks in Eritrea, on the three inhabited islands of Dahlak Archipelago, on the continent in Massawa and suburbs. They are compared with the available data on MSA, Tigre and Tigrinya (see references below and Simeone-Senelle 2014: 686–687).

**Keywords:** Southern Semitic, analytic/synthetic structures, definiteness, comparison

## 1. Introduction

Dahalik [dahālik] is spoken exclusively in Eritrea, on three islands of the Dahlak archipelago: Dahlak Kebīr, Nōra, and Deḥīl (see map). It belongs to the Ethiosemitic group (also named Afrosemitic) together with two languages on the continent: Tigre, spoken on the coast, and Tigrinya, which is the language of the national majority in Eritrea. In spite of a close relationship with the former and many similarities with the latter, Dahalik can be considered as an original variety within the Northern Ethiosemitic group (Simeone-Senelle 2006, 2008a,b, 2010a). 

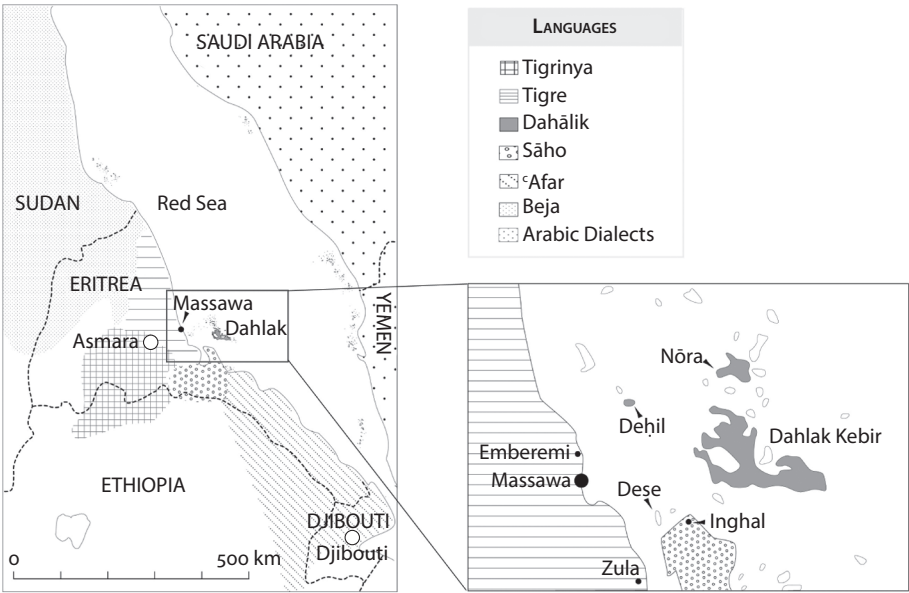


Figure 1. Languages in and around the Dahlak Archipelago

Before delving into details of the possessive and genitive constructions, it is worth giving a brief overview concerning the typology of the Dahalik language.

Both of the following word orders are attested in the sentence: vso and sov. Usually, the dependent clause is before the matrix.

The same speaker uses both orders in the sentence, although (1b) is less usual:

- (1a) o                      v  
    *ħidra                naʔadabbi*  
    holothurians 1PL.sell.IPFV  
    ‘We sell holothurians’
- (1b) s                      v                      o  
    *aġár=e                elħaqa                dibbaħar*  
    foot=SUF.PR.1SG touch.PFV.3M.SG in.sea  
    ‘My feet touch the bottom of the sea’

In the verb phrase, the auxiliary is before verb (AUX v) as in MSA or after verb (v AUX) as in many Ethiosemitic languages.

- (2a) v                      AUX  
*wadet*                  *kūnet*  
 do. PFV.3F.SG AUX.PFV.3F.SG  
 “She had done”

vs.

- (2b) AUX                      v  
*kūnet*                      *wadet*  
 AUX.PFV.3F.SG do.PFV.3F.SG  
 “She had done”

In the noun phrase (Simeone-Senelle 2005: 213; 2010b:110), the invariable definite article (*ya=* or *yā=*), specific to Dahalik, is proclitic to the determined noun:

- (3) *ya=bisa*  
 DEF=man  
 “The man”

The variable deictic determiner is most often post-clitic to the determined noun (4a), rarely proclitic (4b), and in some cases it is double and each part is a circum-fixed clitic (4c).

- (4a) *bisit=da* [bisidda]  
 woman=DEIC.F.SG  
 “This woman”  
 (4b) *di=fuyúl*  
 DEIC.M.SG=business  
 “This business”  
 (4c) *da=bisit=da* [dabisidda]  
 DEIC.F.SG =woman=DEIC.F.SG  
 “This woman”

The other determiners (adjective, relative, noun, pronoun) follow the determined noun or noun phrase. The attributive adjective follows the qualified noun.

- (5) *saheb=na*                      *meskin*  
 colleague=SUF.PR.1PL unfortunate  
 “Our unfortunate colleague”

In Dahalik, as in the other Semitic languages, there are two possibilities for marking possessive or genitive relation between two nouns or between a noun and a noun substitute: (1) a synthetic or direct construction with apposition of both



constituents (head-noun and its determiner/modifier) and (2) an analytic construction or indirect construction with a connective morpheme, a linker<sup>1</sup> which is the marker of the genitive relation and connects the two constituents of the phrase; it is proclitic to the determiner.

As in MSA languages, both constructions may coexist but with different semantic values. The synthetic construction is less common and constrained by the semantics of the determined noun and the type of relation linking the two components of the phrase.

## 2. Possessive and genitive constructions

The order is as follows:

determined (modified) N (+ linker=)determiner (modifier) N, and (modified) N=SUF.PERS.PR

### 2.1 Synthetic construction

The construction is called synthetic, direct, or construct state. The two constituents of the NP are juxtaposed, and the word order is crucial, since the function of each element is marked by its place in the phrase. In all MSA and in DK the order is always determined N + determiner N. The construct state is the less used of the two constructions, generally restricted to the semantic field of belonging (to a family, tribe, place, or ethnic group).

#### 2.1.1 N + N

In DK, as in MSA, the construct state is attested in the following situations.

##### a. With kinship terms

- (6) *aw walét abu haʔán*  
 father girl father boy  
 “Girl’s father, boy’s father”

- (7) *wáddi sultán*  
 child sultan  
 “The sultan’s son”

1. This genitive marker is glossed GEN (genitive) in Simeone-Senelle 2014.

- b. With the head-noun meaning “in charge of”, “responsible”: *baśal* as in MSA (Simeone-Senelle 2014: 669–670). Unlike MSA, this noun is used only in singular.

- (8) *baśal marša*  
in\_charge\_of wedding  
“Master of ceremonies at a wedding”

The construction is the same in Tigre: *bāśal saylāt* “Originary from ‘Aylat” (Leslau 1945: 172) to express the geographic origin.<sup>2</sup>

In the plural, the collective *sab* “people” is used. Followed by a determiner referring to a place or a community, the NP forms a relation noun expressing the origin, the belonging to a community.

- (9) *sab sad*  
people village/island  
“Villagers, islanders”

This construction is rare with other nouns in the Dahalik lexicon. Only three examples are attested in the corpus.

In (10) it is used in the title of a tale. The direct construction focuses on the exclusive relationship between the hero (Radod) and the reported events: it is his own story.

- (10) *muḥuddúma radōd*  
story/tale Radod (Proper name)  
“Radod’s tale”

Two others, (11) and (12), express the constituent material of an object to which the determined noun refers and a particular type of an object (here a ball).

- (11) *kussúset daháb*  
ball gold  
“Gold ball”  
(12) *kussúset ḥalaq*  
ball rags  
“Rag ball”

The only example of lexicalisation of a synthetic construction is attested in a toponym: both nouns form a simple stressed unit.

2. As in MSA, in Te, the word may also mean “native, originating from”.

- (13) *sīla ʿawāzim* [sīlaʿawāzim]  
 well ‘Awāzim  
 “Ila’awāzim” (proper noun; toponym)

(13) is ultimately used as a proper noun naming an actually uninhabited water-deprived place on the island of Deḥil. It tells us something about the topology of the island together with the history of its settlement: ‘Awazim are Arabic speakers originally from the Arabian Peninsula.

It should be noted that in some MSA languages the process of lexicalisation of the synthetic constructions can result in a fossilised NP. One of the most significant examples, including a complex synthetic construction, is in Mehri *barkəndufhikáyt*<sup>3</sup> “wading birds” such as heron, flamingo, curlew, or egret.

In Tigre the process is exemplified by more examples where the NP in a genitive relation forms a lexematic unity.

- (14) Te  
*məfgār                      ṣaḥay* (Elias 2005: 209)  
 coming\_out sun  
 “East”
- (15) Te  
*beddagge                      < bet dagge* (Palmer 1962: 5)  
 house.town  
 “Town house”

### 2.1.2 N=SUF.PR

The determiner is a dependent personal pronoun, clitic to the head-noun. The determined noun refers to something or someone belonging to or being part of the “personal sphere” of the possessor to whom the pronoun refers. This construction is also attested in all the MSA languages, although it is rare in Soqotri.

The determined noun can be any of the following.

#### a. A kinship term

- (16) *abuhát=na                      u=sáb=na* (our ancestors)  
 fathers=SUF.PR.1PL and=people=SUF.PR.1PL  
 “Our fathers and parents”

3. Cf. Simeone-Senelle 2014: 668. The gloss and the explanation of the lexeme should have been analyzed as: *bər=kəndūf=hikayt* <{having=beak}=shore>. *bər=kəndūf* is grammaticalised as relation noun: “one with beak, beaked one”. A (long) beak is one prominent characteristic of the wading birds, which are frequent on the shores.

- (17) *hawat=o*                      *hu=he*                      *u=bisit=o*  
 brothers=SUF.PR.3M.SG brother=SUF.PR.1SG and=woman=SUF.PR.3M.SG  
 “His brothers, my brother, and his wife”

- (18) *walat=ka* [walattika]  
 girl=SUF.PR.2M.SG  
 “Your daughter”

The consonant initial clitic induces gemination of the consonant final of the determined noun and epenthesis of [i].

b. A body part

- (19) *raʔs=e*                      *ʕayān=e*  
 head=SUF.PR.1SG eyes=SUF.PR.1SG  
 “My head, my eyes”

- (20) *qulb=un* [χulbun]  
 heart=SUF.PR.3M.PL  
 “Their heart”

c. A noun referring to someone’s or something’s identity

- (21) *súm=u*  
 name=SUF.PR.3M.SG  
 “His name”

- (22) *ded=dahlák=na* [deddahlákna]  
 in=Dahlak=SUF.PR.1PL  
 “On our Dahlak (island)” (the speaker and his family are native)

- (23) *ʕad=e* [ʕádde]  
 village=SUF.PR.1SG  
 “My homeland ” (village, island)

- (24) *fuyúl=na*  
 business=SUF.PR.1PL  
 “Our business” (referring to their specific, traditional business: fishing)

- (25) *luyet=na*  
 language=SUF.PR.1PL  
 “Our language”

d. A noun referring to something which is considered as the indisputable property of the possessor

- (26) *balámat=na*  
 sardines=SUF.PR.1PL  
 “Our sardines” (product of our fishing)

- (27) *may=na                      zifár=na*  
 water=SUF.PR.1PL shell\_opercula=SUF.PR.1PL  
 “Our water, our shell-opercula” (supply of water on our boat, shell-opercula  
 fished by us)

### 2.1.3 *Definiteness degree within the synthetic structure*

In this construction, one constituent or both constituents of the NP can be marked as definite: N1 by the definite article or a determinative deictic or suffix pronoun and N2 (the modifier) by a determinative suffix pronoun.

#### 2.1.3.1 *The definite article*

The article is the definite marker of the noun phrase whatever the determiner, a noun or a noun phrase (noun and suffix pronoun).

##### a. DEF=N1 + N2

The definite article is proclitic to the first constituent, it determines the whole noun phrase or the first constituent only.

- (28) *ya=aw [yâw] walet*  
 DEF=father girl  
 “The father of a girl, the girl’s father”
- (29) *ya=madrâset melil* (village on Dahlak island)  
 DEF=school Melil  
 “The school of Melil”

Dahalik differs from the MSA languages having an article. In these latter only the second constituent of the construction can be marked by the article (Simeone-Senelle 2014: 669).

##### b. DEF=N=SUF.PR

As in Tigre, the noun determined by a suffix personal pronoun with a possessive value may be defined by the article (Elias 2005: 92, 106; Raz 1983: 35; Simeone-Senelle 2005: 217). In the MSA languages having a definite article, the pronoun is always suffixed to a definite noun (Simeone-Senelle 2014: 672–673).

- (30) *yâ=salâdi=na*  
 DEF=money=SUF.PR.1PL  
 “Our money”
- (31) *yâ=walât=a                      yâ=ulâd=a*  
 DEF=girl=SUF.PR.3F.SG DEF=children=SUF.PR.3F.SG  
 “Her daughter, her children”

Compare with the following:

- (32) Te  
*la-kaləbka* (*la=kaləb=ka*) (Elias 2005: 92)  
 DEF-dog.SUF.PR.3M.SG  
 “Your dog”  
*la-kətbā* (*la=kətb=ā*) (Elias 2005: 106)  
 DEF-books.SUF.PR.3F.SG  
 “Her books”
- (33) Te  
*məsəl* (*məs=əl*) *wasāyfa* (*wasāyf=a*) (Raz 1983: 108)  
 with.DEF maid-servants.SUF.PR.3F.SG  
 “With her maid-servants”

### 2.1.3.2 *Deictic as a definite marker for N1: N1=DEIC + N2*

Only one example (in a tale) has been collected where the head-noun in a synthetic construction is specifically marked by the clitic determinative deictic. The structure connects an object and the material from which it is made.

- (34) *kussúset=da daháb* [*kussúsetta daháb*]  
 ball= DEIC.F.SG gold  
 “The/ this gold ball”

### 2.1.3.3 *The suffix pronoun as a determinative marker of N1 or N2*

- a.  $N1 = \text{SUF.PR} + N$

- (35) *ʔarf=u ensi*  
 smell=SUF.PR3M.SG human\_being  
 “His human scent, his scent of human being”

- b. The determiner noun is marked as definite by a suffix pronoun:

$$N1 + N2 = \text{SUF.PR}$$

This construction is attested when the determiner is a kinship term, as in (36).

- (36) *be(t) ʔat=u* [*beʔattu*]  
 house maternal\_aunt=SUF.PR3M.SG  
 “His aunt’s house”

### 2.1.3.4 *Concluding remarks*

Examples of complex synthetic construction where the determined is a NP {N1 + DET.N2} are not attested in my corpus, and a perusal of Elias (2005) and Raz (1983) corpora shows they are rare in Tigre.

- (37) Te  
       *ʔakara dāwal ǧəwār* (Elias 2005: 306)  
       farmers districts proximity  
       “The farmers of neighboring districts”

To conclude, the direct genitive construction is attested as being restricted to specific semantic fields as in the MSA languages, in Tigre, and in Tigrinya. It is more infrequent in Dahalik, and it is subject to virtually no process of lexicalisation. Whatever the determiner (noun or dependent pronoun), the word order is fixed: the determined element is followed by the determiner. The construction usually expresses an inalienable possession, an inherent property. Like in Tigre but unlike in MSA, the determiner in the construct-state NP is never marked by the definite article or deictic.

## 2.2 Analytic construction

The indirect construction occurs in Dahalik as in MSA, Tigre, and Tigrinya languages. Currently used when both nouns are not in a relationship of inherence, it can rarely have the same semantic value as the synthetic construction. Given that there is no lexical constraint and any noun may be involved in such construction, the analytic construction is more common than the synthetic one. When the determiner is a pronoun referring to the possessor, the indirect construction prevailing in Soqotri and virtually absent in the other MSA languages is common in Dahalik. In this respect Dahalik also differs from Tigre and Tigrinya where the determinative possessive is clitic to the determined noun (Leslau 1941: 51; Raz 1983: 37–38; Elias 2005: 90–91).

In such a construction, the relationship between both constituents is overtly marked by a morpheme, a linker, always proclitic to the determiner noun or pronoun.

In Dahalik the linker *na=* is invariable and exclusive to this function,<sup>4</sup> as *nay* “independent genitive particle” (Elias 2005: 209), “mostly used as preposition of appurtenance” in Te (Leslau 1945: 189) or as marker of membership status in Ta (Leslau 1941: 42).

4. As *nay* in Tigre (Elias 2005: 209) and Tigrinya (Leslau 1941: 41), only when two nouns are involved in the construction. In the MSA languages the linker is variable in number (Simeone-Senelle 2014: 675).

### 2.2.1 Indirect construction: $N + \text{LINK} = N/\text{SUF.PR}$

The order is the same as in the synthetic construction whether the determiner is a noun or a suffix pronoun. The pronoun has the same dependent form as in the direct construction.

#### 2.2.1.1 Semantic values

Usually the analytic construction is used to express a non-permanent possession, an alienable, temporary property:

- (38) *dabbābat na=maḥmūd*  
 moped LINK= Mahmud  
 “Mahmud’s moped”
- (39) *makkína na=he*  
 car LINK=SUF.PR.1SG  
 “My car”
- (40) *nōra sukkān na=ha*  
 Nora inhabitants LINK=SUF.PR.3F.SG  
 “Nora (island), its inhabitants”
- (41) *ḡilud na=ha*  
 skin LINK=SUF.PR.3F.SG  
 “Her skin”

The suffix pronoun refers to the girl born wrapped in a she-camel skin (cf. 43); the construction implies that this skin is not her own.

The indirect genitive constructions involve many other semantic relationships such as identification, origin, function, measure, and partition.

The determiner noun (second constituent) gives information about the origin, as in (42)–(44) and (48), or the the content, as in (45)–(47).

- (42) *ḥaka na=dahlāk, na=dahālīk*  
 speech LINK=Dahlak, LINK=dahalik  
 “The language of the Dahlak islands, of the Dahalik people”
- (43) *ḡilūd na=insatet*  
 skin LINK=she\_camel  
 “The skin of a she-camel”
- (44) *aqbūr na=fūrs*  
 tombs LINK=Persians  
 “Persians tombs”<sup>5</sup>

5. These tombs are reported to have been built by Persians.



- (45) *alābis na=dafēs*  
clothes LINK=wedding\_present  
“Clothes of the wedding present”
- (46) *śādat na=mōt*  
rituals LINK=death  
“Funerary rituals”
- (47) *sawāri na=may*  
cisterns LINK=water  
“Cisterns of water”
- (48) *kissar na=kendur*  
bread LINK=oven  
“Bread of the oven”

*kissar* names bread cooked in a traditional oven, by contrast with *inġera muḡla* “bread cooked on plate”.

### 2.2.1.2 Analytic construction vs. synthetic construction

When direct and indirect constructions are possible, they have different semantic values. However, their nuances are now and then difficult to disentangle.

In DK, when the second constituent is a suffix pronoun, both constructions may occur but with different semantic values. There is no notion of possession in the indirect one, and the determiner expresses origin (see also (42)–(44) and (48)):

- (49) *śāsa na=ha* vs. *śāsa=ha*  
fish LINK=SUF.PR.3F.SG. fish=SUF.PR.3F.SG  
“Her fish” (given to her) vs. “Her fish” (which she fished herself)
- (50) *sab na=śad* vs. *sab śad*  
people LINK=village/island people village/island  
“People of the village, of the island” vs. “Villagers, islanders”

*sab na-śad* refers to people living temporary in the village and from different origins (village, country, other region ...), *sab śad* (9) refers to the native and permanent inhabitants of the village.

- (51) *ḥaka na=dahālik* vs. *ḥaka dahālik*  
language LINK=dahalik language Dahalik  
“(The) language of (the) Dahalik people” vs. “(The) Dahalik language”

In *ḥaka dahālik* the determiner functions as attributive adjective.

- (52a) *kussušet na=dahab* vs. *kussušet dahab*  
ball LINK=gold ball gold  
“Ball made of gold” vs. “Gold ball”

- (52b) *kussuṣet na=ḥalaq* vs. *kussuṣet ḥalaq*  
 ball LINK=rags ball rags  
 “Ball made of rags” vs. “Rag ball”

The nuance is tenuous. It seems that the synthetic construction should name a type of ball, while the analytic construction focuses on the material the ball is made of. In the same text, the speaker gives equivalents as an explanation for *kussuṣet na-dahab* and *kussuṣet na-ḥalaq*, with the preposition *min*, marker of origin: *kussuṣet min-dahab*, *kussuṣet min-ḥalaq*. The synthetic construction occurs at the end of the listing, as the more appropriate word.

### 2.2.2 Definiteness degree within the analytic structure

The suffix pronoun can determine only the second constituent:

- (53) *raʔis na=ʕad=na* [naʕaddína]  
 chief LINK= village=SUF.PR.1PL  
 “The chief of our village”

As in the synthetic construction, only the first constituent can be marked as definite by the article or a deictic (proximal one). *na=SUF.PR* has a value of possessive pronoun:

- (54) *noway=di na=he tu*  
 sheep=DEIC.M.SG LINK=SUF.PR.1SG COP.M.SG  
 “This sheep is mine”

The second constituent (modifier) can be determined by a numeral:

- (55) *ḥikayat=da* [ḥikayadda] *na=hente besit*  
 story=DEIC.F.SG LINK=one woman  
 “This story (is that) of a women (who ...)”

When the head-noun is marked as definite by the article, the suffix pronoun referring to the possessor is always in analytic construction.

- (56) *ya=kissar na=hum* (their (daily) bread)  
 DEF=bread LINK=PR.3M.PL.  
 “Their livelihood”
- (57) *ya=singʕar na=na*  
 DEF=escort LINK=PR.1PL  
 “Our escort ship”

In Dahalik, the NP in the indirect construction never forms a semantic and syntactic unit. Actually, the attributive adjective (58) or the verb (59) can be inserted between the head-noun and the determiner.

- (58) *sawāri buduḥ na=may*  
 cistern many LINK=water  
 “Many cisterns of /for water”
- (59) *sādat ellilun na=marša*  
 rituals have.PFV.3M.PL LINK=marriage  
 “They have marriage rituals”

### 2.2.3 Remarks on word order in Dahalik, MSA, Tigre, and Tigrinya

In Dahalik the word order is fixed whatever the type of construction or the category of the modifier (noun or pronoun). The synthetic construction is the same in all the languages. However the word order differs in Tigre and Tigrinya when the construction is analytic.

In Tigre there are more examples of the reverse order in the Mensa variety (Raz 1983), than in Tigre of Habab where “the modifying noun usually follows the modified noun” (Elias 2005: 209).

- (60) Te  
*nāy bun ṛabqālāt* (Raz 1983: 125)  
 of coffee plants  
 “Coffee plants”
- (61) Te  
*nāy faf tyārotāt* (Raz 1983: 131)  
 of fight planes  
 “Fighter planes”
- (62) Te  
*sadāt nay təgra* (Elias 2005: 209)  
 culture of Tigre  
 “The Tigre culture”

The same applies in Tigrinya, where both orders occur:<sup>6</sup>

- (63) Ta  
*nay ṛabbay ṛəngera* vs. *ṛəngera nay ṛabbay* (Leslau 1941: 41)  
 of father.SUF.PR.1SG bread bread of father. SUF.PR.1SG  
 “My father’s bread” (“Le pain de mon père”)

The order LINK=Determiner + Determined is more frequent in Leslau’s examples (1941: 40, 42).

6. “Si on emploie la particule *nay*, le groupe *nay* et complément peut précéder ou suivre le complété” (When the particle *nay* is used, the group including *nay* and complement may precede or follow the completed) (Leslau 1941: 52).

Moreover in Tigrinya there is a construction similar to Soqotri where the suffix pronoun referring to the possessor is clitic to the linker *nat=*. As in Soqotri, this linker varies with the number of the determined noun – (N.SG) *nat=*SUF.PR; (N.PL) *natat=*SUF.PR (Leslau 1941: 52) – and the pronoun phrase usually stands before the modified noun:

- (64) Ta  
*nat=ki qwālsa* (Leslau 1941: 52)  
 LINK=2F.SG son  
 “Your son” (“Ton fils”)
- (65) Ta  
*nat=om bəṣray* (Leslau 1941: 52)  
 LINK=3PL bull  
 “Their bull” (“Leur boeuf”)

In Soqotri the pronoun has an independent form:

- (66) *di=hoh káser*  
 LINK=PR.1SG house  
 “My house”

DK is closer to the usual Te word order (according to Elias 2005: 209) than to that of Ta where *nay=N* occurs indifferently before or after the head-noun (Leslau 1941: 41).

The following comparative tables summarise the typology of the genitive and possessive constructions<sup>7</sup> attested in some languages belonging to the same Southern Semitic group. They highlight the common features and the divergences between the Ethiosemitic sub-group and the Modern South Arabian sub-group. Furthermore they provide evidence of the originality of Soqotri within MSA as concerns the prevalence of indirect construction with a pronoun and its word-order in such a construction.

**Table 1.** Possessive and genitive constructions in some Southern Semitic languages

	N + N	N=SUF.PR	N + {LINK=N}	N + {LINK=SUF.PR}
DK	+	+	N + <i>na=</i>	N + <i>na=</i>
Te	+	+	N + <i>nay=</i>	Ø
Ta	+	+	N + <i>nay=</i>	N + <i>nay/nat=</i>
MSA	+	+	N + <i>d/l=</i>	Ø
Soq	(+)	(+)	N + <i>d/l=</i>	<i>d/l/mən=</i> PR + N (66)

7. The rare occurrences are put in brackets.

Table 2. Word-order in possessive and genitive constructions

	N + N	N=SUF PR	N + {LINK=N}	N + {LINK=SUF.PR}
DK	+	+	+	+
Te	+	+	+	Ø
			& ({LINK=N} + N)	
Ta	+	+	+	N + {LINK=SUF.PR}
			& {LINK=N} + N	
MSA	+	+	+	Ø
Soq	(+)	(+)	+	{LINK=PERS.PR} + N

3. Conclusion

The structure of the possessive and genitive constructions in the MSA languages and in DK depends on the meaning of the nouns that constitute it and on the grammatical category of the modifier/determiner. Both constructions, direct and indirect, are attested (see Table 1). However, generally speaking, in all the languages, it is the indirect construction that is usually used, the direct construction being limited to a small number of lexemes. The indirect construction can be used with any personal pronouns and any nouns. It can have some semantic values similar to the direct construction. When, for the same constituents and in the same language, the two possibilities coexist, each one expresses a different type of relationship between the two elements of the phrase. The synthetic construction marks a definite and specific relationship of possession, the determiner referring to something or someone considered as belonging to the personal sphere of the determined noun. The analytic construction, on the other hand, marks an alienable relationship.

In the indirect construction, Dahalik, like Tigre and Tigrinya, has a linker exclusive to this construction and rather similar in form (*na* vs. *nay*, *nat*). *na* and *nay* are invariable while *nat* in Tigrinya, dedicated to the indirect construction with suffix pronoun, varies with the number of the determined noun. By contrast, in all the MSA languages the linker is polyfunctional and always variable (Simeone-Senelle 2014: 675 et sq.).

On the syntactic level (Table 2), in the direct construction and in the indirect construction with a noun, the word order is similar in all the languages. This order is not fixed in Tigre and Tigrinya where the determiner NP may precede or follow the determined noun. The indirect construction with personal pronoun is attested in Dahalik, Tigrinya, and Soqotri but not in Tigre and in the other MSA languages. Dahalik and Tigrinya have the same word order in this construction. However, and only with the linker *nat*=, Tigrinya has the same word order as Soqotri. The Dahalik word order, fixed and similar to that of MSA (except Soqotri concerning

the construction with a pronoun determiner) is a discriminative point within the Northern Ethiosemitic group.

The survey of examples and the comparison of the analytic construction highlight the relationship of Dahalik with the Northern Ethiosemitic group on the morphological level but with the MSA group on the syntactic level.

Furthermore, concerning the lexicalisation of the genitive constructions, Dahalik stands out in the group of Southern Semitic languages. No example in the corpus attests lexicalisation in Dahalik, while in the other languages the genitive constructions, both direct and indirect, can be fixed in fossilised constructions, integrated in the lexicon and sometimes in the morphology of the language.

## Abbreviations

### Languages:

DK	Dahalik/dahālik
MSA	Modern South Arabian
Soq	Soqotri
Ta	Tigrinya
Te	Tigre

### Gloss:

AUX	auxiliary	NP	noun phrase
DEF	definite article	Ø	no occurrence
DET	Determiner	PERS	personal
F	feminine	PFV	perfective
GEN	genitive	PL	plural
IPFV	imperfective	PR	pronoun
LINK	linker	SUF	suffix
M	masculine	V	verb
N	noun		

With the exception of Elias, previous authors did not use a glossing system or systematically segment their examples. In these cases I have completed the segmentation and added my own glosses.

## Acknowledgements

This paper is the second part of my study of possessive and genitive constructions in some Southern Semitic languages, dedicated specifically to the Dahalik language. For comparison with the Modern South Arabian languages the reader should consult the first part in (Simeone-Senelle 2014).

This research was supported by CNRS, the Alliance française in Asmara, and the Ministry for Education and Research in Asmara. Their support is deeply appreciated. I am also grateful to Eritrean authorities and to the native speakers in Massawa and on the islands for their invaluable cooperation and their active part in this research.

## References

- Elias, David Lyndon. 2005. *Tigre of Habab: Short Grammar and Texts from the Rigbat People*. Cambridge, MA: Departement of Near Eastern Languages and Civilizations, Harvard University.
- Leslau, Wolf. 1941. *Documents tigrigna (Ethiopien septentrional)*. *Grammaire et textes*. Paris: Klincksieck.
- Leslau, Wolf. 1945. *Short Grammar of Tigré: The verb in Tigré (North-Ethiopic), Dialect of Mensa*. Pub. of the American Oriental Society. Offprint 18. New Haven, CN: American Oriental Studies.
- Palmer, Frank Robert. 1962. *The Morphology of the Tigre Noun*. London: Oxford University Press.
- Raz, Shlomo. 1983. *Tigre Grammar and Texts*. Malibu: Undena Publications.
- Simeone-Senelle, Marie-Claude. 2005. "La détermination du nom en dahaalik". *XI Incontro Italiano di Linguistica Camito-Semitica. 5-7 giugno 2003, University of Bergamo*, ed. by Alessandro Mengozzi, 03-217. Milano: Franco Angeli.
- Simeone-Senelle, Marie-Claude. 2006. "Some Characteristics of Dahalik, a Newly Discovered Afro-Semitic Language Spoken in Eritrea". *Proceedings of the 15th International Conference of Ethiopian Studies, Hamburg, July 2003*, ed. by Siegbert Uhlig, Maria Bulakh, Denis Nosnitsin & Thomas Rave, 861-869. Wiesbaden: Harrassowitz.
- Simeone-Senelle, Marie-Claude. 2008a. "A Survey of the Dahalik Language". <http://halshs.archives-ouvertes.fr/halshs-00319729/fr/>.
- Simeone-Senelle, Marie-Claude. 2008b. "Up to Date Assessment of the Results of the Research on the Dahalik Language". (December 1996-December 2005) <http://halshs.archives-ouvertes.fr/halshs-00320383/fr/>.
- Simeone-Senelle, Marie-Claude. 2010a. "The Specificity of the Dahalik Language within the Afro-Semitic Languages". *History and Languages of the Tigre Speaking Peoples*, ed. by Gianfrancesco Lusini, 127-145. Napoli: Università di Napoli L'Orientale.
- Simeone-Senelle, Marie-Claude. 2010b. "The Relative Clauses in Dahalik". *Current Trends in Eritrean Studies*, ed. by Gianfrancesco Lusini, 109-123. Napoli: Università di Napoli L'Orientale.
- Simeone-Senelle, Marie-Claude. 2014. "Expression de l'appartenance et de la possession dans le syntagme nominal en sudarabique moderne". *Linguistic, Oriental and Ethiopian Studies in Memory of Paolo Marrassini*, ed. by Alessandro Bausi, Alessandro Gori & Gianfrancesco Lusini, 661-687. Wiesbaden: Harrassowitz.

# The characterization of conditional patterns in Old Babylonian Akkadian

Eran Cohen

The Hebrew University of Jerusalem

In this paper, a common paratactic conditional construction, found in the letter corpus of Old Babylonian Akkadian (~18th century BCE), is given a syntactic characterization so as to differentiate it from other potential sequences. Several distinctive syntactic and semantic features unique to the construction are identified and discussed: polar lexical resumption between the protasis and its preceding co-text; negative polarity items in the protasis; special semantics of verbal forms; and divergence from the common modal-congruence. In addition, the structural variables are formulated, and eventually the construction itself is compared with another construction, the circumstantial construction.

The importance is twofold: first, to exemplify a relatively simple characterization of a construction, which is usable for the identification of the construction in question; second, to add a description which pertains to the non-uniquely-marked conditional constructions in the languages of the world, which are often ignored or underdescribed.

**Keywords:** conditional constructions, paratactic conditionals, construction characterization, Old Babylonian Akkadian syntax, conditional vs. circumstantial constructions

## 1. Preliminaria

### 1.1 General background

Old Babylonian is the ‘classical’ phase of Akkadian, and the oldest copiously attested Semitic language, written mostly between the 18th and 17th centuries BCE. This paper describes a conditional pattern, characterizing it in a way that distinguishes it from other sequences.

The language at the center of this inquiry is the epistolary language, found in thousands of letters and hence well known. It features verb-final order, where all



verbal arguments (except suffixes and clitics) occur PRECEDING the verbal forms, whereas all other modifiers (e.g., those occurring with nouns) FOLLOW.

The clauses are either verbal or non-verbal. Subordination always means that the clause functions as an attribute to some nucleus, be it substantival, pronominal or adverbial (see Table 1).

Table 1. Attribution in Old Babylonian

nucleus type	nucleus	attribute	gloss
substantival	<i>bīt</i> house.NUC	<i>ab-ī-šu</i> father-GEN-GEN.3MS	“house (of) his father”
		<i>ī-puš-u</i> 3CS-make.PST-SBJV	“house (which) he made”
pronominal	<i>ša</i> PRON.NUC	<i>ab-ī-šu</i> father-GEN-GEN.3MS	“that (of) his father”
		<i>ī-puš-u</i> 3CS-make.PST- SBJV	“that (which) he made”
adverbial	<i>kīma</i> as.NUC	<i>ab-ī-šu</i> father-GEN-GEN.3MS	“as his father”
		<i>ī-puš-u</i> 3CS-make.PST- SBJV	“as he made”

The subordinate clause is generally marked by a special verbal form, called subjunctive. Such clauses always function as PART of the main clause. Note that, in this capacity, they are interchangeable with an entity in genitive status.

1.2 The domains

Besides this exceptionally clear and consistent type of subordination, Akkadian syntax has a special means to interconnect clauses, the connective particle *-ma*. It creates an ASYMMETRICAL INTERCONNECTION between the clauses, thereby marking a logical or temporal sequence between the clauses. The order it creates is linguistically pertinent, since the sequence is not reversible.

The clausal chains interconnected by the particle *-ma* are characterized by a formal, morphological unity common to all the forms in the chain: they all belong to one group, e.g., they are all subjunctive forms or all directive (that is, the forms which denote a command). This feature is termed “modal congruence” (Cohen 2005: 123–137). It serves, *inter alia*, as the cohesive factor among the clauses, which together constitute the syntactic unit above the clause level which may be called ‘domain’ (Cohen 2014).

The directive domain is exemplified in (1):<sup>1</sup>

- (1) *ina šāb PN<sub>1</sub> u PN<sub>2</sub> 100 šāb-um ittī-šu*  
 from army.NUC and troupe-NOM with-GEN.3MS  
*l-i-llik=ma* 5 *ūm-ī* {*adi* PN<sub>1</sub>u PN<sub>2</sub> *ištu* GN  
 JUSS-3CS-go=CONN 5 day-OBL.PL until and from  
*i-llakūnim#*} *ina āl-ānī* *l-i-ptarrik-ū=ma*  
 3-come.NPST.MP in cities.OBL.PL JUSS-3-ITER.trouble-MP=CONN  
*ḥarrān-ātim* {*ša* *ī-ten-errubānim#*} [*i*]*šteat ū šitta*  
 caravan-OBL.PL PRON.NUC<sup>2</sup> 3-ITER.NPST-come\_in.FP one or two  
*l-i-dūk-ū=ma* *l-ī-dur-ā #*  
 JUSS-3-strike-MP=CONN JUSS-3-fear-FP

“Let one hundred troops from the troops of PN<sub>1</sub> and PN<sub>2</sub> go with him, and let them cause continuous difficulties in the cities for five days {until PN<sub>1</sub> and PN<sub>2</sub> come from GN}, and let them strike at one or two caravans {that come in regularly} so that they be afraid” (11, 193: 13–23<sup>3</sup>)


Inside two individual directive clauses, one finds two separate occurrences of the attributive domain (in curly brackets), both marked (1) by a nucleus (*adi* and *ša*) and (2) by forms which are clearly not directives (they do not begin with *l-* as do 3rd person directive forms). Note that the subjunctive forms of the verb are not interconnected with the directives but are rather literally embedded WITHIN the directive clause. Note, in addition, that there is a strong notion of finality (e.g., between the last two verbal forms). This is one of the semantic notions common with chained directives.

The domains are conceived of as the next definable syntactic entity after the clause. Their existence may be justified from yet another angle: they each seem to have different complement syntax: the directive domain is markedly different in this respect from the syntax of the indicative domain:

1. PN=proper name; GN=geographical name; CN=canal name; [...] encloses a broken part; an underlined vowel signals functional lengthening. The glossing follows the Leipzig rules, with the following differences: DIR.SP= direct speech converter; pronouns appended to nouns and prepositions alike are glossed GEN (followed then by person, gender, number) rather than POSS. NVC is an abbreviation for non-verbal clause.

2. See Table 1 above.

3. The source for the examples is the *Altbabylonische Briefe* (see references). I note merely the numbers of the volume, letter and lines.

- (2)  *qi[b]i-ma ma[mman l]ā udabbab-šu* (=negative directive)  
 order.IMP.2MS=CONN anyone NEG (3)CS.harass.NPST-3MS  
 “O[r]der that (lit. and) n[o one] should harass him” (12, 13: 17–18)
- (3) *mamman lā dubbub-šu [i]-qtabī-šinūšim* (ibid. 12–13)  
 anyone NEG harass.INF-GEN.3MS [3CS]-say.PF-DAT.3MP  
 “[He] ordered them that no one should (lit. anyone not to) harass him”

The same events are represented in two different domains: (2) is the directive domain and (3) is the indicative domain. Whereas in (3) the indicative verbal form (“he ordered them”) is formally complemented by an INFINITIVE in accusative status (“not to harass”), acting as a formal object, the directive domain works differently: here, both finite complement clauses and infinitive complements are truly rare. Instead, chains are normally used, where the content of the order is conveyed via a chained (rather than subordinate) clause, using an entirely different strategy ((2) lit. “order and let no one harass him”).

Note that this particle *-ma* does not, in itself, represent or impose any particular meaning. The semantic notion among the chained clauses could be neutral, final, indirect volitive (“je veux qu’il fasse ...”), conditional, or even circumstantial. For this reason, the EXISTENCE of a consistent CONDITIONAL PATTERN is not self-evident: the examples may be explained away as sequences which merely incidentally denote condition. I go one step further and show that a carefully defined pattern consistently signals conditional function. This paper describes several parameters which seem to be pattern-specific and can hence serve as distinctive features.

### 1.3 Literature review

This paper characterizes a conditional pattern which does not have an overt, clear conditional marking. Of the instances where a conditional structure is not marked by an exclusive, unequivocal conditional exponent, most cases still have forms which have a clear-enough, relatively specialized exponent of the protasis (and in some cases, the apodosis). For instance, the CONDITIONAL (or PSEUDO-) IMPERATIVE (Fortuin & Boogaart 2009), the CONDITIONNEL verb form in French (Borillo 2009: 115, 2010: 5–8) or verb-first order (Hilpert 2010; see also Reis & Wöllstein 2010, who describe this construction within the system of other constructions). Such conditionals are somewhat easier to identify and describe. Less overtly marked forms, and hence more challenging ones, are discussed by Thumm 2000 (referring to many types of paratactic conditionals in English) and Borillo (2008, 2009, 2010), who covers the entire array of paratactic patterns in French, involving several

combinations of different verbal forms as well as several types of interconnection (on the term ‘paratactic’, see below).

Old Babylonian is an ancient language, a written medium, and the structures in question mostly feature the form denoting non-past, which may occur anywhere and is hence not very distinctive. However, the language has an array of conditional patterns (described as a group in Cohen 2012: 29–120 (Chapter 2)); the cases discussed below are relatively frequent in the text and are consistently consensual from a semantic point of view, interpreted by the philological experts as denoting conditionality.

#### 1.4 Terminology

The term to describe these conditional sequences is ‘parataxis’. This is the usual term in Akkadian to refer to the interconnection of clauses via the particle *-ma*, but it calls for an explanation: In French linguistic terminology, *parataxe* generally means the state of affairs in [P, Q] rather than [P (and/so/then) Q]. However, in the English-written treatment of conditionals, the term ‘paratactic conditionals’ has quite a consensus. The first use of this term is found in Haiman 1983 and refers to the structure  $S_1$  (*and*)  $S_2$  which may be interpreted as *If*  $S_1$   $S_2$ . Thumm 2000: 3 describes this conditional type within the framework of contextualization, triggering conditional interpretation. The term ‘paratactic’ is used for the comprising clause combinations that contain coordinating conjunctions such as ‘and’, ‘or’, or ‘but’, as well as in referring to asyndetic juxtaposition, that is, constructions without any explicit exponent between the clauses (ibid., n. 2). Declerck & Reed (2001: 401) define the term as follows:

PARATACTIC CONDITIONALS are constructions that are interpreted as conditionals but in which the P-clause is coordinated with the Q-clause (rather than being syntactically subordinated to it).

Last but not least, Fortuin 2011, describing similar patterns in Russian, uses the term ‘paratactic conditional’ for conjunctionless constructions as well as for instances where the apodosis is introduced by a resumptive form (*как* “so, then”, *тогда* “then”) or a coordinative conjunction (*и* “and”). The term “parataxis” is used because the protasis and apodosis are juxtaposed without an explicit exponent conjunction that denotes condition in the protasis (ibid., 90 and n. 2).

To avoid this difficulty, these patterns in French are termed “hypothétiques non-marquées” (Corminboeuf 2008: 12 n. 19). The point is that, when the overt

connective does not contribute any specific meaning, it is not considerably different than finding no overt exponent between the clauses.

These ‘paratactic conditionals’ are here defined as structures SEEMINGLY devoid of external characterization, or which appear to be minimally characterized. Hence they need some kind of characterization to set them apart, if one wants to regard them as a category in their own right. Such characterization should be as concise as possible and should focus mainly on the formal side, but not exclusively.

As one last remark, the term ‘pattern’ is used here as a cover term for a non-compositional complex and hence an irreducible exponent of some linguistic function. It is not essentially different than the term ‘construction’ as used within construction grammar, except that its characterization is simpler, as it is meant to be more practical, or ready-to-use.

The context is the entire array of complex conditional forms in Old Babylonian. The pattern discussed below is merely one out of several, marked in different ways, and not always by the incontestable conditional particle *šumma* “if”.

To facilitate following the examples, I adduce a short reader of forms in Table 2.

Table 2. Form reader

form	name	gloss	group name
<i>iprus</i>	“preterite”	PST	indicative
<i>iptaras</i>	“perfect”	PF	
<i>iparras</i>	“present-future”	NPST	
<i>paris</i>	“stative”	STV	
<i>purus</i>	“imperative”	IMP	directive
<i>liprus</i>	“precative”	JUSS	

(note that neg. particle *la* precedes the present-fut (NPST) function as a prohibitive

2. Parameters

The main part of this paper describes distinctive and other features which characterize the paratactic conditional pattern.

2.1 Preceding polar directive: polar lexical resumption

The most instructive feature of the current paratactic conditional pattern is POLAR LEXICAL RESUMPTION. This is a phenomenon where in the immediate co-text preceding the protasis there is a directive (e.g., an imperative or a precative form, used to issue a command) which has the same (or a related) verbal lexeme as the one in the protasis. Occasionally, the arguments and other complements are resumed as well. This form and the protasis following it exhibit opposite polarity:

- (4) *ina* 4 *ūm-ī* *dalt-ī* *lū* *kams-at* ... DIRECTIVE  
 in day-OBL.PL door-GEN.1CS MOD.PTCL complete.STV-3FS  
*ina* 4 *ūm-ī* *dalt-ī* *ul* *kams-a[t]=ma* NEG. RESUMPTION  
 in day-OBL.PL door-GEN.1CS NEG complete.STV-3FS=CONN  
*lū* [t]-īde-ā  
 MOD.PTCL 2-know.STV-PL  
 “Let my door be complete within four days ... (If) my door is not complete within four days, be warned” (3, 34: 19, 37–38)
- (5) *rēq-ūs-su* *lā* *i-llakam* NEG. DIRECTIVE  
 empti-ness-GEN.3MS NEG 3CS-come.NPST  
 [r]ēq-ūs-su *i-llakam=ma* RESUMPTION  
 empti-ness-GEN.3MS 3CS-come.NPST=CONN  
 [b]īt-ī *i-šabbat=ma* *ana bāb-im*  
 house-GEN.1CS 3CS-seize.NPST=CONN to gate-GEN  
*ušešše-anni=ma*  
 (3)CS-take\_out.NPST-1CS=CONN  
 “Let him not come empty-handed; should he come [em]pty-handed, he would seize my [hou]se and put me out to the gate ...” (6, 140: 20–25)

The preceding directives generally assume the execution of the commanded action. There is hence a plain logical incompatibility between the two forms in the same utterance (for instance: *do not go* vs. *you go* ... namely, the command and its obvious opposite), which is resolvable only by interpreting the structure as an unequivocal conditional.

The parts containing the polarity opposite to the directive (“it is not ready” in (4) and “he comes” in (5)), open up a channel of likelihood that the command may not be carried out as originally expected. Such likelihood is at the basis of any epistemic, non-factual expression. This part is the protasis of the construction.

## 2.2 Negative polarity items: (otherwise) negative expressions and *arhiš ul*

‘Scale reversing contexts’ are negative or negative-implying environments – such as a question or a conditional structure – which host special expressions, termed ‘negative polarity items’ (or rather ‘opposite polarity items’). An example for this phenomenon in English would be the behavior of the indefinite pronoun *anyone*: One could compare the occurrence of the phrase #you see anyone# alone (which is very irregular) with its normal occurrence in negation, question or condition:

---

negation: #You do not see anyone#

question: #Do you see anyone?#

condition: #If you see anyone...#

---

Curiously, expressions which generally occur negated in the Old Babylonian epistolary corpus are chiefly “be negligent” expressions: *aḥam nadûm* “be negligent”, *nîdi aḥim rašûm* “be negligent”, and *idam šuršûm* “be careless, raise objections”. In the pattern in question, these expressions conspicuously occur in the AFFIRMATIVE.

### 2.2.1 The expressions “be negligent”

#### 2.2.1.1 *aḥam nadûm*

The following (6) is a transition between the former parameter, polar lexical re-sumption, and an opposite polarity item.

- (6) *ana ša a-špur-akkum aḥ-ka lā*  
 to PRON.NUC 1CS-write.PST-DAT.2MS hand-GEN.2MS NEG  
*t-anaddi ⇒ aḥ-ka t-anaddi=ma*  
 2MS-throw.NPST hand-GEN.2MS 2MS-throw.NPST=CONN  
*di<h>ti-ka ašâ[l]*  
 information-GEN.2MS 1CS-ask.NPST  
 “Do not be negligent regarding what I wrote you;  
 should you be negligent, I will inquir[re] after you” (12, 64: 31–33)

The prohibitive form of this expression *aḥam nadûm* (around 80 tokens) is the common form of this lexeme, and its affirmative use is typical either in questions or in this very pattern. In (6), the expression occurs twice, once as a preceding negative directive and immediately thereafter in the affirmative, which indicates a scale reversing context, and, not being a question, we are left with the conclusion that this indeed is an indication for a conditional protasis.

#### 2.2.1.2 *nîdi aḥim rašûm*

- (7) *nîdi aḥ-im t-arašši-ā-šim=ma šumma*  
 throw.NUC arm-GEN 2-have.NPST-CP-DAT.3FS=CONN IF  
*elepp-um šī i-mtūt ḥamuttam=ma ša kīma*  
 boat-NOM PRON.NOM 3CS-sink.PF quickly=FOC PRON.NUC like  
*šâti n-irtêb* (3, 35: 26–28)  
 PRON.OBL 1CP-replace.NPST  
 “Should you be negligent towards it, if this boat were to sink, would we be able to replace it for ourselves soon with one just like it?”

When occurring without the preceding directive, as in (7), it can be interpreted as implicitly responding to the commonly negative expression, which is missing (since its occurrence as a protasis is enough). It turns out to be a special mark for a conditional environment.

### 2.2.2 *arḥiṣ ul*

Another case consists of the adverb *arḥiṣ* “quickly”, which occurs around 200 times in this letter corpus. This adverb generally occurs with prospective verbal expressions (non-past, directive, and infinitive). It is rarely compatible with negation. The syntagm [*arḥiṣ ... ul t-aparras*], that is, its compatibility with the negated 2MS non-past, is generally found **ONLY** in the protasis of this paratactic pattern. It is therefore considered as yet another distinctive feature:

- (8) *kanikk-am arḥiṣ ul t-ušezziḫ-šu=ma*  
 document-ACC quickly NEG 2MS-CAUS.leave.NPST-3MS=CONN  
*lū t-īde* (12, 25: 19–20)  
 MOD.PTCL 2MS-know.STV  
 “Should you **not** have him draw up a document **quickly**, be warned”
- (9) *rakb-ī arḥiṣ ul t-appal-ā=ma ...*  
 rider-OBL.PL quickly NEG 2-satisfy.NPST-CP=CONN  
*pān-ū-kunu ul i-bbabbal-ū* (4, 11: 29–33)  
 face-NOM.PL-GEN.2MP NEG 3-PASS.carry.NPST-MP  
 “Should you **not** satisfy the(se) riders **quickly**, you will not be forgiven ...”

### 2.3 Special semantics: the temporal frame of *ul iprus*

One form which occurs in the protasis of this pattern is IPRUS, what is traditionally termed ‘preterite’. It is attested in the negative only (UL IPRUS):

- (10) *ištu inanna ana 5 ūm-ī ina mahri-ka wašbā-ku adi*  
 from now to 5 day-OBL.PL in front-GEN.2MS dwell.STV-1CS till  
*a-llakam ul t-ukillaš-šu=ma* PN<sub>1</sub> u PN<sub>2</sub>  
 1CS-come.NPST NEG 2MS-prepare.PST-3MS=CONN PN<sub>1</sub> and PN<sub>2</sub>  
*t-appal* (6, 73: 10–17)  
 2MS-answer.NPST  
 “Within 5 days from now I will be with you. If you **have not prepared** him by the time I come, you will (have to) answer to PN<sub>1</sub> and PN<sub>2</sub>.”

The example makes it clear that the verbal form UL IPRUS does not have past or even present perfect value, as it normally does in declarative main clauses:

- (11) *[k]as[p]-am ana qāti-šu uttēr 1 uttēt*  
 silver-ACC to hand-GEN.3MS (1)CS.return.PF 1 grain.NUC  
*[k]as[p]-am ul ukī[l]* (7, 145: 8)  
 silver-ACC NEG (1)CS.keep.PST  
 “I have returned to him the silver, I **did not** (even) **keep** (or: **have not even kept**) one grain (of) silver”



The explicit temporal frame in (10) (“five days from now”), within which preparation is to take place, is located in the realm of the FUTURE. The temporal reference of the form *ul* IPRUS is similar to that of another form, *lā*, inside the more prevalent *šumma* conditional patterns, which often denotes the FUTURE PERFECT. This point is very important: the value of *lā* (which is otherwise regarded as the negated preterite or perfect) in the conditional pattern in question is CONSIDERABLY DIFFERENT THAN THE FUNCTION IT HAS ANYWHERE ELSE. Another example is (12):

- (12) ... *alkam=ma* PN *apulli* (sic) ⇒ *ul t-allikam=ma*  
 come.IMP.2MS=CONN PN pay.IMP.2MS NEG 2MS-come.PST=CONN  
*nāq mē ina bīti-ka ul i-zzibūni*  
 pourer.NUC water.OBL.PL in house-GEN.2MS NEG 3-leave.NPST.MP  
 “... Come and pay PN; if you have not come they will not leave a water pourer  
 in your house” (7, 67: 13–18)

The form *ul tallikam=ma* in Example (12) has, again, a future-perfect value, as it is expected to have occurred only AFTER the imperative “pay”.

## 2.4 Diverging from modal congruence

As noted above, one of the prominent syntactic phenomena of Old Babylonian is MODAL CONGRUENCE (Cohen 2005: 123–137). This means that forms which are chained, or interconnected via the connective *-ma*, form a DOMAIN consisting of one group of forms – directive forms, indicative forms, subjunctive forms, etc. These forms, in principle, DO NOT INTERMIX within the confines of the *-ma* domain. The pattern in question, however, allows such odd interconnections between different groups of forms, specifically the interconnection between an indicative and directives (e.g., imperative, jussive or prohibitive).

In the following pair of examples such peculiar sequences are found, which clearly contradict this modal congruence:

- (13) *ina nār* CN *nārt-a muḥur=ma eql-am*  
 from canal-NUC CN dike-ACC accept.IMP.2MS=CONN field-ACC  
*ša PN mē mullī=ma ana errēš-im*  
 PRON.NUC PN water.OBL.PL fill.IMP.2MS=CONN to cultivator-GEN  
*idīn=ma nār-am šāti {ul t-amḥur=ma}*  
 give.IMP.2MS=CONN canal-ACC DEM.OBL NEG 2MS-accept.PST=CONN  
*ul awāt-ī* (11, 175: 5–10) [IMPERATIVE-*ma* PRETERITE]  
 NEG affair-GEN.1CS

“Draw off a canal from the CN canal and fill PN’s field with water and give (it) to the cultivator; {if you have not drawn off the canal}, (it will) not (be) my affair”

In (13), the preceding directive is interconnected with the protasis form (in curly brackets). In (14), the protasis, as usual, is interconnected with the directive apodosis.

- (14) *pīhāt-ī*                      *apāl-am*                      *u[l] e-[l]e"i=ma*  
 obligation-PL-GEN.1CS    answer.INF-ACC    NEG    1CS-be\_able.NPST=CONN  
*ana mi[mm]û-ya* (sic)      *kīma i-baššû*                      *lā*  
 to    property-GEN.1CS    as    3CS.exist.NPST.SBJV    NEG  
*t-eggī*                                      (6, 148: 27–29) [PRESENT-*ma* IMPERATIVE]  
 2MS-be\_negligent.NPST  
 “Should I no[t] be [a]ble to meet my obligations, do not be negligent towards  
 my pro[per]ty, as much as there is”

This ‘digression’ is found in either direction from the protasis, that is, the protasis, consisting of forms which are elsewhere considered indicative, is nevertheless willing to interconnect with the directive group of forms. It may be taken as formal corroboration to the special modal, or more specifically, conditional function of these so-called indicative forms.

## 2.5 The pattern: forms and structure

The most common form in both protasis and apodosis is the form IPARRAS, generally denoting the non-past. However, in dialogue environments such as the letter corpus, these forms are very common, and as such they can hardly constitute a distinctive feature for the conditional sequence. The other verbal forms occurring in this conditional pattern, although less frequent, are better for characterization.

### 2.5.1 Forms and compatibility

In the protasis one finds, less frequently, UL IPRUS forms (discussed in § 2.3), and in one case a PARIS form. The apodosis mostly has IPARRAS (non-past) but also LIPRUS (directive), PARIS (stative), and non-verbal clauses:

Table 3. Forms in the protasis and in the apodosis of the pattern

protasis	connective	apodosis	(the negative particle is added here to the form when the form is attested only in the negative)
<i>iparras</i>	<i>-ma</i>	<i>iparras</i>	
<i>ul iprus</i>		<i>liprus</i>	
( <i>ul paris</i> )		non-verbal clause	
		<i>paris</i>	

A point worth mentioning *qua* forms is that, curiously, all the cases of negated 2nd person IPRUS or IPARRAS (that is, the forms *ul t-aprus-ma* and *ul t-aparras-ma*), which are connecting forward to a legitimate apodotic form (to wit, neither IPRUS nor IPTARAS forms; see below), are *always* a part of this pattern.

The forms in question are an important part of the characterization of the pattern, and eventually the entire group serves as a distinctive feature vis-à-vis other patterns. Table 4 shows the forms and their compatibility across the particle *-ma*; protasis forms are compatible only with apodosis groups with which they share a side in the table (e.g., IPARRAS is compatible with all forms):

Table 4. Compatibility of protasis forms with the apodosis forms

protasis	apodosis
( <i>ul paris-ma</i> )	<i>liprus</i>
<i>iparras-ma</i>	<i>paris</i>
<i>ul iprus-ma</i>	<i>iparras</i> non-verbal clause

### 2.5.2 The connective particle *-ma*

The protasis and apodosis are mostly interconnected via the connective particle *-ma*. The particle is occasionally absent (and the gap is indicated by the arrow):

- (15) *maškan-a ša 5 mana id-î-šim=ma kil-î-ši*  
 fetter-ACC PRON.NUC 5 mina put.IMP-2FS-DAT.3FS=CONN hold.IMP-2FS-3FS  
*amt-um uзн-ā-ša anni[š]=ma [NVC-ma]*  
 maid.NOM ear-DU-GEN.3FS there=CONN  
*i-halliq-ki ⇒ maškan-u*  
 3CS-get\_lost.NPST-DAT.2FS fetter.NOM  
*l-i-battiq-ši* (1, 27: 24–28)  
 JUSS-3CS-hinder-3FS  
 “Put fetter(s) of 5 mina on her and detain her; the maid is intelligent; should she (try to) escape from you, let the fetter(s) hinder her”

The absence of the particle *-ma* between the protasis and the apodosis is not an obstacle, as is clear from (15), and it is occasionally attested in other chains as well. Note, however, that the common Semitic connective *u* does not occur under any circumstances here between the protasis and the apodosis:

- (16) *umma anāku=ma kīma erēb-i-ka kasp-am qāti*  
 DIR.SP 1CS.NOM as enter.INF-GEN-GEN.2MS silver-ACC hand.NUC  
*awīl-im usuḥ # kasp-am qāti awīl-im ul*  
 man-GEN remove.IMP.2MS silver-ACC hand.NUC man-GEN NEG  
*t-assuḥ u kasp-um šibt-am*  
 2MS-remove.PST CONN silver-NOM interest-ACC  
*uṣṣab* (12, 53: 5–8)  
 (3)CS.add.NPST  
 “I (said) as follows: ‘As soon as you come in, take the silver from the man’. You **did not take** (or: **have not taken**) the silver from the man and the silver (now) bears interest.”

(16) is not a conditional (note that *usuḥ* “remove” and *ul t-assuḥ* “you did not remove” do not originate in the same utterance), and it illustrates very well the different nature of the connection by *u*, whose mere occurrence between the clauses is enough to rule out the conditional pattern.

## 2.6 The respective function of the forms inside the pattern

Table 5 summarizes the possible forms and lists their respective functions in both parts of the paratactic conditional pattern. An important fact which should be emphasized is that, since the forms in the apodosis occur in the apodosis of ORDINARY conditions, everything is *a priori* **CONDITIONED**. That is, the basic function of the apodosis – the common denominator of all the forms occurring in it – is **CONDITIONALITY**. For instance, the forms in the apodosis, whatever they may be, are subject to whether the clauses in the protasis take place or not. For this reason, the forms in **BOTH** protasis and apodosis lie somewhere on the epistemic scale (Akatsuka 1985), analogous to modal particles such as ‘perhaps’. The other functions are secondary and depend on the pattern.

Table 5. The functional values of the forms in the pattern

protasis			apodosis	
forms	values	connective	forms	values
<i>ul iprus</i> NEG 3CS.PST	(non-past) perfect	<i>-ma</i> (or $\Rightarrow$ )	NVC	indicative
<i>±iparras</i> 3CS.NPST	(dynamic) non-perfect		<i>paris</i> (STV.3MS)	
<i>(ul paris</i> NEG STV-3MS	resultative)		<i>iparras</i> (3CS.NPST)	directive
			directive (JUSS)	
			question	
			conditional	condition

Note that in the protasis, only negative forms can denote the future perfect, so the opposition with this perfect function is to be found only in the negative. Consequently, the affirmative non-past (IPARRAS) is not opposed to any perfect. It should be mentioned that the stative is marginally attested and is hence not very significant.

Considering the forms in the apodosis, two additional possibilities are found, hitherto not discussed, being less germane to the characterization of the pattern:

1. A question: when the entire conditional pattern is produced as question (e.g., (21)), such a question is occasionally marked, but only on the apodosis;
2. An entire additional conditional pattern could stand for the apodosis (see (7), where, in fact, question and condition co-occur).

## 2.7 Summary

To conclude the characterization, the following example illustrates some of the peculiarities of the paratactic conditional pattern:

- (17) *šāb* *našpak-ātim* *apul=ma*  
 army.NUC cargo\_boat-PL.OBL supply.IMP.2MS=CONN  
*našpakāt-im* *l-i-puš* *arhiš šāb*  
 cargo\_boat-OBL.PL JUSS-3CS-make quickly army.NUC  
*našpak-ātim* PN *ul t-appal-šu=ma*  
 cargo\_boat-OBL.PL PN NEG 2MS-supply.NPST-3MS=CONN  
*pīhat-um* *ši* *ana muḥḥi-ka*  
 responsibility-NOM DEM.NOM.FS to top-GEN.2MS  
*i-ššakkan* (2, 59: 14–21)  
 3CS-PASS.put.NPST  
 “Supply PN with cargo boat workers so that he can build the cargo boats ...;  
 should you not supply him quickly with cargo-boat workers, the responsibility  
 will be laid on you”

The biclausal sequence *ul tappalšu=ma ... iššakkan* (IPARRAS–*ma* IPARRAS) is the core of the pattern, constituting the protasis and the apodosis respectively:

1. The directive *apul* “supply” is the co-textual preceding directive, which precedes the protasis, and has a tight connection with the form in the protasis, where it is in fact resumed: it has the same lexeme (*apālum*, “pay, satisfy”) but opposite polarity, viz., the directive is affirmative whereas the protasis is negative (§ 2.1).
2. A second feature of the pattern, evident in (13), is that this DIRECTIVE is occasionally interconnected via the particle *-ma* to the IPARRAS form in the protasis, in a way which contrasts the otherwise obligatory modal congruence (§ 2.4 (13) and § 2.5.2).

3. A third peculiarity of the pattern found in (17) is the compatibility of the adverb *arhiš* “quickly” with the negative particle *ul* and the form IPARRAS. Again, this occurs only in the protasis of the paratactic pattern in question (§ 2.2.2).
4. A fourth feature, manifest in (10) and (12), is the special temporal value of UL IPRUS (§ 2.3). The sequence of forms in (17) answers fully to the most common combination, but the other features give it away. Note, however, that no example actually exhibits the entire list of characteristics discussed and that the identification is carried out based on different combinations of these features.

### 3. Distinction from other analogous patterns

The conditional structure has been fully characterized as a pattern in Old Babylonian enumerating several unique parameters, only some of which are external. However, other patterns considered, this characterization does not cover every conceivable difficulty in the identification of this pattern, especially when it comes down to superficial features, that is, mere form. In this respect, some confusion may ensue in attempting to distinguish the paratactic conditional pattern from ostensibly similar patterns. A case in point is the interconnected CIRCUMSTANTIAL CLAUSE (see Cohen 2015).

The circumstantial clause often consists of the stative form, PARIS, forward connected by the particle *-ma*, as is illustrated in the following example:

- (18) *maruṣ=ma*                      *ul*    *i-llikam*                                      PARIS-*ma*  
 sick.STV. 3MS=CONN NEG 3CS-come.PST  
 “Being sick (lit. he is sick and) he has not come”                                      (2, 212: 9–10)

However, it is of a wider nature, as is clear in (19):

- (19) *ana* GN *gerr-um*                      *ul*    *i-mqut=ma*  
 to GN caravan-NOM NEG 3CS-fall.PST=CONN  
*ul*    *a-llik*    UL IPRUS-*ma*  
 NEG 1CS-go.PST  
 “A caravan **did not arrive** in GN so I did not go”                                      (2, 77: 4–6)

Note that a circumstantial clause (the group of forms is presented in Table 6, where the functions are on the left hand side) may consist, in addition to the common stative form, of NEGATED PRETERITES and NON-VERBAL CLAUSES.<sup>4</sup> Ordinarily, the precise shade of meaning in the circumstantial clause is often causal. The circumstantial

4. A conditional non-verbal circumstantial clause is underlined in (15) above: *amtum uznāša anni[š]=ma* lit. “the maid, her ears are there”.

expression is COMPATIBLE with our conditional pattern, which apriori means that they constitute DIFFERENT categories. In (20) and (21) the circumstantial clause immediately precedes the conditional pattern:

- (20) *u kallat* PN *napi-at=ma* (circumstantial) *PARIS-ma*  
 CONN daughter\_in\_law-NUC PN be\_taken\_as\_pledge.STV-3FS=CONN  
*ina nakkamt-im t-u<še>ššî-ši=ma* (conditional) *IPARRAS-ma*  
 from storehouse-GEN 2MS-CAUS.take\_out.NPST-3FS=CONN  
*maḥrî-ki l-i-šib*  
 before-GEN.2FS JUSS-3CS-dwell  
 “Furthermore, the daughter-in-law of PN is taken as pledge, so should you release her from the storehouse, let her stay with you” (9, 270: 10–15)
- (21) *kâtî išû-ka=ma* *šani-am*  
 you.OBL (1)CS.have.STV-2MS=CONN another-ACC  
*e-še’î=ma tâb-kum* (9, 226: 4–6)  
 1CS-look\_for.NPST=CONN be\_good.STV.3MS-DAT.2MS  
 “Me having you, should I look for someone else, would it please you?”

The circumstantial expression is less easily characterizable than the paratactic conditional pattern (see Cohen 2015). Its most important distinctive features, namely, the forms at play and their respective functions, are compared with the right hand column of Table 6, which contains the forms of the protasis clause in the paratactic conditional pattern:

- The negated preterite (UL IPRUS) is the only form which is attested in both circumstantial expression and paratactic conditional; note, however, the marked difference in temporal values.
- The most common form in the circumstantial expression is the stative *PARIS* (attested only once, in the negative, in the conditional pattern).
- The form *IPARRAS*, on the other hand, is attested only as conditional and is the most common form with this function.
- The non-verbal clause is native as circumstantial but never attested as a protasis in the paratactic pattern.
- The last resort in differentiating between the constructions is the second clause in each construction (i.e., the ‘main clause’ of the circumstantial and the conditional apodosis). This clause is in principle unrestricted in the case of the circumstantial expression but it often consists of the preterite *IPRUS* (as in (18) and (19)), whereas the conditional apodosis never has this form.

It is important to note how the different forms at play characterize each of the patterns differently by presenting a distinct distribution for these forms in each function.

Table 6. Circumstantial vs. conditional

circumstantial clause	forms	- <i>ma</i> conditional protasis
functions as a <b>negated preterite</b>	<i>ul iprus-ma</i>	functions as a <b>future perfect</b> ; (2nd person is always conditional)
<b>the most common form</b>	<i>paris-ma</i>	(occurs only once, in the negative)
(marginal) <sup>5</sup>	<i>iparras-ma</i>	<b>the most common form</b>
recurring	(S)-P- <i>ma</i> (Non verbal Cls)	non-occurring

#### 4. Conclusions

The paratactic conditional pattern is described based upon several syntactic and semantic criteria:

1. A directive carrying the same semantics as the protasis but showing the opposite polarity often precedes the protasis;
2. Negative polarity items occur only in the protasis (e.g., specific verbal expressions);
3. The form *UL IPRUS* has the functional value of future-perfect;
4. Combinations which conflict with the otherwise strict modal congruence allow for special combinations;
5. A specific set of forms makes up the protasis.

These descriptive data suffice, it seems, to change the status of these superficially ‘notional’ conditional chains to a carefully defined and easily identifiable conditional pattern.

Similar information with regard to paratactic conditionals in any language, wherever they may occur, is crucial for a cross-linguistic understanding of this conditional type.

5. The form *IPARRAS* occurs in circumstantial function only with *adīni ul* meaning “until now”, but in combination with *IPARRAS* it functions as a **negated present perfect**.



## References



- Akatsuka, Noriko. 1985. "Conditionals and the Epistemic Scale". *Language* 61:3.625–639. doi:10.2307/414388
- Borillo, Andrée. 2008. "Structures de corrélation hypothétique et situations de dialogue". *Recherches ACLIF: Actes du Séminaire de Didactique Universitaire* 5.9–25.
- Borillo, Andrée. 2009. "Quelques formes de corrélation hypothétique caractéristiques des situations de dialogue". *Langages* 174.113–128. doi:10.3917/lang.174.0113
- Borillo, Andrée. 2010. "La corrélation hypothétique et la construction parataxique." *La parataxe: Structures, marquages et exploitations discursives*, vol. 2 ed. by Marie-José Béguelin, Mathieu Avanzi & Gilles Corminboeuf, 3–29. Bern: Peter Lang.
- Cohen, Eran. 2005. *The Modal System of Old Babylonian* (Harvard Semitic Studies 56). Winona Lake: Eisenbrauns.
- Cohen, Eran. 2012. *Conditional Structures in Mesopotamian Old Babylonian*. Winona Lake: Eisenbrauns.
- Cohen, Eran. 2014. "The Domain: A Formal Syntactic Unit above Sentence Level." *Strategies of Clause Linking in Semitic Languages*, ed. by Bo Isaksson and Maria Persson, 233–252. Wiesbaden: Harrassowitz.
- Cohen, Eran. 2015. "Circumstantial Clause Combining in Old Babylonian Akkadian." *Clause Combining in Semitic: The Circumstantial Clause and Beyond*, ed. by Bo Isaksson & Maria Persson, 365–406. Wiesbaden: Harrassowitz.
- Corminboeuf, Gilles. 2008. "'Tu m'embrasses encore, et c'est mon pied dans les pompons!' Comment construit-on le sens?" *Discours* 3. doi:10.4000/discours.4173
- Declerck, Renaat & Susan Reed. 2001. *Conditionals: A Comprehensive Empirical Analysis* (Topics in English Linguistics 37). Berlin: Walter de Gruyter. doi:10.1515/9783110851748
- Fortuin, Egbert. 2011. "Iconicity, Economy, and Frequency: The Paratactic Conditional Perfective Present Construction in Russian." *Russian Linguistics* 35.89–123. doi:10.1007/s11185-010-9071-6
- Fortuin, Egbert & Ronny Boogaart. 2009. "Imperative as Conditional: From Constructional to Compositional Semantics." *Cognitive Linguistics* 20:4.641–673. doi:10.1515/COGL.2009.028
- Haiman, John. 1983. "Paratactic If-Clauses." *Journal of Pragmatics* 7.263–281. doi:10.1016/0378-2166(83)90014-0
- Hilpert, Martin. 2010. "What Can Synchronic Gradience Tell Us about Reanalysis? Verb-First Conditionals in Written German and Swedish." *Gradience, Gradualness and Grammaticalization*, ed. by Elizabeth Closs Traugott & G. Trousdale, 181–201. Amsterdam: John Benjamins. doi:10.1075/tsl.90.10hil
- Reis, Marga & Angelica Wöllstein. 2010. "Zur Grammatik (vor allem) konditionaler V1-Gefüge im Deutschen." *Zeitschrift für Sprachwissenschaft* 29.111–179. doi:10.1515/zfs.2010.004
- Thumm, Markus. 2000. "The Contextualization of Paratactic Conditional." *Interaction and Linguistic Structures* 20.1–38. <http://www.inlist.uni-bayreuth.de/issues/20/index.htm>

# Locative predication in Chadic

## Implications for linguistic theory

Zygmunt Frajzyngier

University of Colorado at Boulder

The aim of this study is to demonstrate that Proto-Chadic had a category ‘locative predication’ that was formally and semantically distinct from all other predication in the language. The proposed hypothesis, combined with the principle of functional transparency (Frajzyngier & Shay 2003), allows us to answer the following theoretical questions: (1) why lexical items with the same reference have different properties across languages; (2) why a given form in the same language is sometimes used and sometimes not used in the coding of what appears to be the same situation; (3) why some languages have only one locative preposition and other languages have many; (4) why some languages deploy serial verb constructions and others do not; and (5) why some languages deploy verbal extensions for some functions and others do not. These questions are examined using data from locative expressions in Chadic languages.

The existence of a grammaticalized predication may imply that some lexical items are compatible with semantic features of the predication and others are not. The existence of locative predication, as proposed for Proto-Chadic, means that, if the semantic feature ‘locative’ is present in the complement and/or in the predicate, no other formal means are required to mark either component for locative predication. If the feature ‘locative’ is absent in the predicate, languages with locative predication have either lexicalized the category ‘locative predicator’ or have grammaticalized other means, such as serial verb constructions and locative extensions, to mark the predicate as locative. If the complement is not inherently locative, languages have lexicalized the category ‘locative preposition’ to code the locative complement. In some Chadic languages, subsequent changes have resulted in the replacement of locative predication by narrower semantic categories, such as movement toward a goal, movement from the source, or presence in a place.<sup>1</sup>

**Keywords:** Chadic, locative predication, types of prepositions, verbal extensions, semantic compatibility

---

1. The work on this study was conducted while I held the Chaire Régionale de Chercheur Étranger de la Région des Pays de la Loire. The work on Wandala was supported by NSF Grant

## 1. Introduction

### 1.1 The aim and scope of the study

The main aim of this study is to demonstrate that Proto-Chadic had a category ‘locative predication’, semantically and formally distinct from all other predications in the language. A ‘predication’ is a grammaticalized meaning. A locative predication has a general locative meaning that may subsume much narrower characteristics such as presence at a place, movement toward a place, or movement from a place. Significant traces of locative predication still exist in a number of contemporary Chadic languages. Once the existence of locative predication in Proto-Chadic is demonstrated, I discuss changes that locative predication has undergone, including the loss of locative predication as a category and the emergence of semantically narrower locative expressions.

The proposed hypothesis, combined with the principle of functional transparency (Frajzyngier & Shay 2003; Frajzyngier 2004), allows one to explain the following facts, which were previously unexplained or which have not been viewed as needing explanation: (1) why some locative expressions have prepositions and others do not; (2) why some languages have only one locative preposition; (3) why some languages must have a locative preposition even though they have locative predication; (4) why some languages have serial verb constructions coding locative relations, and others do not; and (5) why some languages have verbal extensions coding locative relations and others do not. The study thus explains why locative expressions have different forms across related languages.

As an illustration, consider coding movement toward a place in Hausa and Mupun (both West Chadic). In Hausa, the expression may consist of the subject, the verb, and a locative complement without a preposition (1). In Mupun the expression consists of the subject, the serial verb construction, and a preposition preceding the locative complement as in (2).

- (1) *yaa tàfi Kano* (Hausa)  
 3M:PRF go Kano  
 “He went to Kano.”

---

Grant Nr. 0439940. The work on Mupun, Lele, and Hdi was supported by grants from NSF and NEH. The work on Pero was supported by the University of Colorado. Without the help of native speakers of these languages, this work would not have been possible. I am grateful to the Humboldt Foundation for the award that allowed me to collaborate at the University of Bayreuth with Mohammed Munkaila on the locative predication in Hausa. Some results from this work are used in the section on Hausa in the present study. I am most grateful to Erin Shay for substantial and editorial comments on this study and to Joe Salmons for valuable editorial comments.

- (2) *wu se siam n-panksin* (Mupun)  
 3M depart descend PREP-Pankshin  
 “He went to Pankshin.”

The questions posed by the two examples are: (a) Why does Mupun (like English) require a preposition, in this case *n*, before locative complement, while Hausa does not? (b) Why does Mupun have the serial verb construction (*se siam* “depart descend”) and Hausa (or English) does not?

Another illustration raises the questions of why the same preposition in the same language is used in clauses having opposite meanings with respect to directionality and why there is a locative postposition in one case and no locative postposition in another:

- (3) *η kil-iy dà bónú-m ni*  
 1SG buy-3M PREP brother-2M LOC  
 “I bought it from your brother.”
- (4) *se è gé dà túgú pói kúsíge-η kè-y Ø*  
 INCEPT go 3PL PREP home Poi Kusige-DEF GEN-3M  
 “They went to the house of Poi Kusige.” (Lele, Frajzyngier 2001)

The coding of locative expressions in Chadic turns out to be a testing ground for the semantic structure of the language and for the forms coding the semantic structure, i.e., for the syntax of the language. The study also contributes to the explanation of why lexical items having the same reference across languages appear to have distinct syntactic properties. It is thus also a contribution to lexical semantics.

A byproduct of the present study is that it provides an explanation for the existence of serial verb constructions with respect to locative expressions. As an illustration, consider the fact that in some languages one can say:

- (5a) “Who was the man that *ran from Marathon to Athens?*”

In other languages, such as Mupun, one might say something like:

- (5b) “Who was the man that left Marathon ran went reached Athens.”  
 (Lexical glosses are at best a gross approximation of the actual meanings of verbs deployed.)

Most studies take the presence of the serial verb constructions as given, with no attempt to explain their existence. Others attribute the presence of serial verb constructions to different meanings of verbs and to the absence of prepositions in the language, and still others to the different perceptions of reality by the speakers.

The rest of this study is organized as follows. A brief review of the state of the art with respect to locative expressions in Chadic is followed by a statement of the

methodology used to determine whether Proto-Chadic had a locative predication. This is followed by a demonstration that in all three branches of Chadic there exist traces of locative predication. The study then demonstrates that in some languages the locative predication was replaced by narrower semantic categories. The study concludes with the discussion of implications of the existence of locative predication for linguistic theory.

## 2. State of the art with respect to locatives in Chadic

Comparative studies of locative predication in Chadic are limited to Frajzyngier (1987) and Pawlak (2003). Frajzyngier 1987 is concerned with the reconstruction of locative prepositions in Chadic. In that study it is postulated that Proto-Chadic had only one locative preposition, *a*, that may have had different tones. The study also demonstrated that the use of the locative preposition correlated with the feature [+locative] in the complement, in that if the noun in the locative complement was inherently locative the preposition was not used. The preposition was used only if the noun was not inherently locative. That study did not take into consideration the properties of the predicate, a gap that is filled in the present study. Moreover, Frajzyngier 1987 did not take into account the presence of the preposition as motivated by the principle of functional transparency, another omission corrected in the present study. Pawlak 2003 is concerned with the grammaticalization of prepositions. Every grammar of a Chadic language contains some discussion of locative expressions, most often couched in terms of better-known categories of IE languages, such as ‘prepositional phrase’; frequent semantic categories in Chadic such as ‘ventive’ and ‘allative’; and formal categories such as ‘locative extensions’ and ‘particles’. With very few exceptions (Frajzyngier et al. 2005) no interaction among these categories is mentioned.

One of the questions for comparative study is why the locative expressions differ across related languages spoken in the relative geographical proximity. The default expectation here would be to have similar structures for the same semantic function. The specific questions for Chadic languages are as follows:

- Do contemporary Chadic languages have locative predication?
- How does locative predication differ from other types of predication in the language?
- What are the coding means for the locative predication in the Chadic family?
- Why do some structures and languages deploy adpositions while others do not?
- Why do some structures and languages deploy more than one general locative adposition?
- How can one explain the rich variation in locative expressions in a group of related languages?

### 3. The terms

A **LOCATIVE PREDICATION** is a grammaticalized coding of the existence of an element or event at some location or the movement of an element to or from some location. A language has a locative predication if the predication differs in at least one formal characteristic from all other types of predications that the language might have.

A **LOCATIVE EXPRESSION** is any expression, not necessarily grammaticalized, describing movement toward a location, movement from a location, presence in a location, etc. The mere fact that every language can have such expressions does not imply that these functions are coded differently from other predications in a given language. A locative expression is thus different from the locative predication.

A **LOCATIVE COMPLEMENT** is a nominal complement of the locative expression.

A **LOCATIVE PREDICATE** is a verbal or non-verbal predicate of the locative expression.

A **LOCATIVE PREDICATOR** is a predicate whose sole function is to serve as a locative predicate when the predicate of the clause is not inherently locative and the clause aims to convey the locative predication. This function does not imply coding of specific directionality, manner, or any other attributes of locative expression. It turns out that some Chadic languages have one or two locative predictors.

A **SERIAL VERB CONSTRUCTION** is a deployment of two or more verbs to mark a grammatical function. These verbs share the same set of arguments and have the same modality, polarity, tense, and aspect.

**VERBAL EXTENSIONS** are affixes to the verb coding categories other than arguments and tense, aspect, and modality. Some verbal extensions participate in the coding of locative expressions.

A **PREPOSITION** is a marker of the grammatical or semantic relation of a noun that precedes the noun.

A **POSTPOSITION** is a marker of grammatical or semantic relation of noun that follows the noun.

**PARTICLE** is a term devoid of an association with any specific function; it refers to independent grammatical morphemes whose functions are yet to be discovered.

The **PRINCIPLE OF FUNCTIONAL TRANSPARENCY** states that the role of every element in an utterance must be transparent with respect to the functions grammaticalized in the given language. The principle of functional transparency does allow ambiguity.

#### 4. The hypotheses

The following hypotheses are explored:

1. Some languages have grammaticalized locative predication.  
The evidence for the grammaticalization of this type of predication consists of the following. In a language that has grammaticalized locative predication, there exist:
  - a. A class of predicates that, without any other markers, code locative predication. Such a language has grammaticalized formal means to code the locative function of inherently non-locative verbs. These include locative predicators, serial verb constructions, and verbal extensions.
  - b. A class of nouns that, without any other markers, are interpreted as locative complements. Such a language has grammaticalized formal means, e.g., locative prepositions, to code the locative complement function for use with inherently non-locative nouns.
  - c. Even if a language has grammaticalized the locative predication, and even if the language has a class of inherently locative nouns, the locative complement may still be marked by a preposition as required by the principle of functional transparency, as explained below.
2. Some languages do not have a locative predication distinct from other predications and consequently do not have inherently locative predicates and inherently locative nouns. In such languages, there may exist markers of narrower semantic categories, such as movement toward a goal or away from the source, presence at given place, etc.
3. Proto-Chadic had the category locative predication.
4. The category locative predication was replaced in some languages by narrower semantic functions.

#### 5. Synchronic and diachronic methodology required for locative predication

The evidence for the existence of locative predication consists of demonstrating that there is a class of inherently locative complements which appear in locative predication without any additional markers, such as prepositions or postpositions. The non-inherently locative complements must have some marker. There exists a class of inherently locative predicates that without any additional markers code locative predication. Any other predicate must have additional markers of locative predication. The study demonstrates that there exists a cross-linguistic complementarity of forms used to code locative predication with inherently non-locative predicates.



The evidence for the hypothesis that Proto-Chadic had the category locative predication consists of demonstrating that:

1. In some languages from all three branches of Chadic, there exists locative predication, as described above.
2. Given the rarity of locative predication in unrelated languages, the existence of a similar category with specific formal characteristics in three branches of the family makes it unlikely an independent innovation.
3. Coding of the same function through different formal means argues against the function being borrowed. If it had been borrowed it would have been borrowed with some coding means. Hence, if conditions (2) and (3) are met, that argues for the predication being a retention from Proto-Chadic.

It is possible to reconstruct the overt markers involved in locative predication, such as the predicator or the preposition *a*, the prepositions *n* or *t*, and various types of locative extensions on the verb. A reconstruction of the phonological forms alone does not provide, however, any information about their function(s) or even of their categoriality. Many grammars of Chadic languages note the presence of the form *a* in locative expressions. With the exception of the recent works by this writer they analyze it as a preposition, although in a number of cases, e.g., in Mina and Wandala (Frajzyngier et al. 2005, Frajzyngier 2012), they are not prepositions but rather locative predicators. Given that the functions and syntactic coding means are at issue, rather than the phonological means, one has to design a methodology that would enable the reconstruction of the functional domains coded in Chadic.

The methodology for reconstruction of functions coded in the proto-language, in this case the function of locative predication, must involve an investigation of (1) whether the descendant language has distinct ways of coding locative expressions, which may involve prepositions, adpositions, verbal extensions, and serial verb constructions and (2) whether the deployment of locative markers correlates positively or negatively with the inherent locative features of nouns and verbs.

Pursuing this methodology, I first provide the evidence that in all three branches of Chadic there are languages with distinct ways of coding locative predication and that these ways correlate with the inherent locative features of nouns and predicates. I start with Mina, a Central Chadic language, which fully displays the presence of locative predication and where some nouns and some predicates are inherently locative. I then demonstrate the existence of the locative predication in another Central Chadic language, Hdi, where different phonological forms of a preposition provide evidence for the existence of inherently locative nouns. The West Chadic data are represented by Pero, which displays serial verb constructions, prepositions, and postpositions, each of which is a formal means coding a different facet of the



locative predication. East Chadic is represented by Lele, which has serial verb constructions and postpositions, both in complementary distribution with the inherent locative features of predicates and nouns. The presence in each branch of at least one language with morphological or syntactic evidence of the locative predication is powerful argument that such a predication was coded in Proto-Chadic.

6. Complementarity of lexical and grammatical means in locative predication in Mina

6.1 The system

The system of coding locative expressions in Mina (Central Chadic) includes inherently locative predicates, inherently locative complements, linear order, the predicator *á*, and the prepositions *n* and *ká*. Evidence for the existence of locative predication is provided by two forms: (1) the predicator *á*, which codes locative predication when the predicate is not inherently locative, and (2) the locative prepositions *n* and *ká*, used when the complement is not inherently locative. The following functional distribution obtains:

	Pred	Coding	Compl	Coding
Locative	+	Ø	+	Ø
Locative	–	á	–	n

The term ‘coding’ in the table refers to the coding by overt grammatical means. The symbol ‘Ø’ refers to the absence of overt coding. ‘Locative’ refers to the inherent semantic locative feature of the lexical item, ‘Pred’ is the predicator, and ‘Compl’ is the complement.

Inherently locative verbs include directional verbs of movement such as “go to”, “come from”, and a few others. Inherently locative nouns include place names; terms for “compound”, “house”, and “room”; and locative adverbs for “here” and “there”. Animate nouns, including human nouns, cannot be inherently locative. The system of locative predication in Mina operates as follows:

1. If the complement is inherently locative, no marker of locative complement are needed.
2. If the complement is not inherently locative, it is coded as such by the preposition *n*. The structure of the locative predication is Predicate *n* Noun.
3. If the predicate is inherently locative it is not coded by any other means.
4. If the predicate is not inherently locative, the locative predication has the form Predicate *á* Complement or Predicate *ká* Complement.

## 6.2 Inherently locative predicate and inherently locative complement: coding through juxtaposition

Inherently locative verbs in Mina include: *ndà* “go”, *tsú* “went”, the borrowed verb *nástà* (Fula) “enter”, and a few others. Toponyms are inherently locative, as are the words *dámù* “uncultivated area, bush” and *wùtá* “village, compound”:

- (6) *íbà ndà tètè wùtá*  
ASSC:PL go 3PL.POSS village  
“They went home.”
- (7) *séy mà ngùl ngùl tiy á tiy-ù wàl tsú zà dāmù*  
so REL husband see 3SG see-3SG wife went EE bush  
“So the husband saw that the wife went to the bush.”

Direct objects can also follow the verb without any additional marking. The evidence that the noun following the verb belongs to the locative predication rather than to what I shall provisionally call transitive predication is provided by the behavior of the end-of-event marker *za* (phrase-internal forms: *z*, *zà*). This marker occurs BEFORE a locative complement, as in (8), but AFTER a direct object, as in (9):

- (8) *tséy hidi wàcín tán z wùtá à bá á n médíg*  
so man DEM return EE house 3SG say PRED PREP neighbor  
*ngàñ wàcín ngámbug há kà déb-é-ñ dāl nà hídà wà*  
3SG DEM friend 2SG INF bring-GO-3SG money PREP man DEM  
*dāl vānū*  
money how much  
“When the man came back to the house, he said to his neighbor, ‘Friend, you brought money to this man. How much money?’”

The end-of-event marker occurs after a direct object:

- (9) *áá wàl nà ká dzán-á skàn pàr zà bàdáp*  
ah wife 1SG INF find-GO thing another EE again  
“Ah, my wife found another thing again.”

## 6.3 Locative predicate and non-locative complement: Predicate *n* Noun

An inherently non-locative noun phrase is marked for its locative role by the preposition *n*. Inherently non-locative nouns include [+human] nouns and pronouns:

- (10) *mìnjée mbà mà mármàr ká nàz-á kw-yii zà nà láy*  
now boy REL pasture INF abandon-GO goat-PL EE PREP field  
“Now the shepherd left the goats in the field.”

The evidence that *láy* is inherently locative is that it can be used as a direct object of the clause:

- (11) *guzak nan kə vl-a-k lay za*  
 uncle 1SG INF give-GO-1SG field EE  
 “My uncle gave me a field.” (written sources, hence no tonal notation)<sup>2</sup>

The preposition *n* does not code spatial specification, such as “at”, “in”, etc., nor does it code a distinction between the directional and stative meaning, as illustrated in the examples above and below.

#### 6.4 Non-locative predicate and locative complement: Predicate *á* Noun

A locative predication whose predicate is inherently non-locative must be marked by the particle *á* (‘predicator’), which follows the direct object, if present:

- (12) *nd-á yà ngùl ngàn á bìn*  
 go-GO call husband 3SG PRED room  
 “And [she] called her husband into the room.”
- (13) *ñkwà tá làvéñ hì ká skàm-á zà hì fàt kà á káyàk*  
 goat GEN black 2PL INF buy-GO EE 2PL skin POS PRED ground  
 “A black goat, when you have bought it, you skin it on the ground.”

Non-directional verbs of motion such as *tíl* “leave, move”, *yàn* “move house”, and *déb* “carry” are inherently non-locative and require the predicator *a* in locative predication:

- (14a) *tíl ngən a wta* (written sources)  
 leave 3SG PRED house  
 “He returned home.”
- (14b) *bày ñ kádām ngən bāt déb á déb ká á idá*  
 chief PREP calabash 3SG take carry 3SG carry POS PRED home  
 “The chief took his calabash and carried it home.”
- (14c) *èe hìd-yî wà í-bà yàn tètà á màcín*  
 eh, man-PL DEM PL-ASSC move 3PL PRED there  
 “Those people moved over there.”

2. Written examples come from transcriptions by Adrian Edwards, which do not indicate tone.

The predicator *á* codes locative predication when the main verb is *dáhá* “exist”:

- (14d) *háḷàm dáhá á biŋ ngèn*  
 daughter exist PRED house 3SG  
 “There is a girl at her house.”

If the clause has no predicate, the locative predication is coded by the predicator *á* and a locative complement. This is evidence that the particle *á* alone functions as a locative predicate:

- (15a) *kwáyàŋ zá ḷi mà m̀ats-yí bàytán á dàmù*  
 squirrel COMP meat REL die-STAT large PRED bush  
 “The squirrel said, “There are a lot of dead animals in the bush.”

As demonstrated in Frajzyngier et al. 2005, the preposition *ká* is also a locative predicator with the meaning approximating “be behind”, used when the predicate is not inherently locative:

- (15b) *à zá wàcín nék skù náz ká dùwáŋ dà*  
 3SG COMP DEM good NEG throw PRED behind house  
 “He said this isn’t good. He threw it behind the house.”  
*gimíḷíá káts wál ngəŋ-yí ábà wəz-yí tán nd-á ciké*  
 monkey gather wife 3SG-PL ASSC children-PL DED go-GO all  
*ká dùwáŋ dà kwáyàŋ*  
 PRED behind compound squirrel  
 “The monkey gathered his wives and children. They all stayed behind the squirrel’s house.”

## 6.5 Non-locative predicate and non-locative complement: Predicate *á* *n* Noun

If neither the predicate nor the complement is inherently locative, that results in the predicator *á* and the preposition *n* occurring in the same clause:

- (16) *há tán tówər á nə fálà tətàn*  
 2SG DED suffer PRED PREP among 3PL  
 “You suffer [a lot] among them.”

## 7. Locative predication in Hausa

The data on Hausa (West Chadic) are discussed here as part of a demonstration that in every branch of Chadic there exists at least one language that has the category locative predication. The system is as follows. The predicator *a* ('locative preposition' in the Hausa grammatical tradition) is deployed when either the predicate or the complement is not inherently locative. If both elements are inherently locative no other means to code locative predication have to be deployed. The inherently locative verbs include the allative and ventive forms of *je* "go" and *zoo* "come" (probably the same verb), *shiga* "enter", *tàfi* "go", and a number of other verbs. Toponyms are inherently locative nouns, as are nouns signifying "house", "room", and "compound". Example (1) above illustrates the fact that, when the verb is inherently locative and the complement is inherently locative, no other means to code locative predication are required. Compare also the following:

- (17) *sun sa mota gareji*  
 3PL put car garage  
 "They put a car in the garage."

The imperfective/continuous aspect, coded by *suna* and its so-called 'relative' counterpart *suke*, is inherently locative, corresponding to the meaning "to be at", as will presently be demonstrated. If the complement is inherently locative, no other means have to be used to code locative predication in the imperfective:

- (18) *su waa sukee daki*  
 3PL who 3PL:be room  
 "Who are the people inside the room?"

The particle *a*, can, however, be used:

- (19a) *su wa sukee a daki*  
 3PL who 3pl:BE PRED room  
 (19b) *su wa suke (a) cikin daki*  
 3PL who 3PL:be PRED inside room  
 "Who are those people in the room?"

The distinction between the inherently locative predicate and the non-locative predicate is that the latter requires the use of the particle *a* in a locative predication, regardless of the presence or absence of the feature [locative] on the complement. The verb *akwai* "exist" is inherently non-locative, as evidenced by the fact that the predicator *a* is required in locative predication, even with inherently locative complements (recall from (1) that inherently locative complements are not marked if the predicate is inherently locative):

- (20a) *akwai mutane dà yawa a kano*  
 exist people ASSC many PRED Kano  
 “There are a lot of people in Kano.”
- (20b) *akwai makaranta a nan garin*  
 exist school PRED DEM town  
 “There is a school in this town.”

We can therefore postulate that, contrary to the widespread analysis of *a* as a preposition in Hausa, it is a predicator, similar to the one described for Mina. Unlike in Mina, the predicator in Hausa is used to mark a locative predication when either the predicate or the complement is not inherently locative. Thus, using a non-locative noun, such as *akwati* ‘box’, as a locative complement of the verb *sa* or *ajiye* ‘put’ requires the use of the predicator *a*. The absence of the predicator makes the utterance ungrammatical. A locative complement must be marked as such by a preposition when it follows the object as in (21a):

- (21a) *sun sa/ajiye kaya a akwati*  
 3PL put thing PREP box  
 “They put the goods into the box.”

Compare the ungrammaticality of the clause without the preposition preceding the noun *akwati* ‘box’:

- (21b) \**sun sa/ajiye kaya akwati*  
 3 PL put thing box  
 “They put the goods into the box.”

With an inherently locative complement the verbs *sa/ajiye* may be used without the predicator *a*:

- (22a) *sun ajiye mota a gareji*  
 3PL put car PRED garage  
 “They put the car in the garage.”
- (22b) *sun ajiye mota gareji*  
 3PL put car garage  
 “They put the car in the garage.”

I conclude that Hausa has locative predication. If either the complement or the predicate is not inherently locative, the locative predicator *a* is used. If both the predicate and the complement are inherently locative, there are no other markers of locative predication.

## 8. Locative predication in Pero<sup>3</sup>

Pero (West Chadic) also has the category of locative predication. The system of coding locative predication in Pero is as follows. If both the predicate and the complement are inherently locative, no other means of coding locative predication are required. The evidence for the existence of the locative predication is provided by the fact that inherently locative nouns are not marked by a locative preposition in locative expressions. Among these nouns are *birá* “outside”, *bìcìrù* “there”, *pídì* “place”; all toponyms, e.g., Kano, Filiya, etc.; and a few ordinary nouns, such as *miná* “house”, *makaranta* “school” (Hausa), and *lāādì* “church”. Pero also has a class of inherently locative verbs, which include directional verbs of movement. When an inherently locative verb has as its complement an inherently locative noun, no other markers of locative predication are used:

- (23a) *n-kájà mándì pídì*  
 CONJ-move another place  
 “And she moved to another place.”
- (23b) *n-wáatò-n miná án-kúndùl ká yìppá*  
 CONJ-go-ANAPH house owner-kundul PREP night  
 “... and take it to the house of the owner of kundul at night.”

An inherently locative noun may be preceded by prepositions, but the function of prepositions is to code spatial orientation, such as “inside”, “on”, “under”, “behind”, rather than locative function:

- (24) *rú-kò yù-miná*  
 enter-PRF inside-house  
 “He entered the house.”

If the noun is not inherently locative there are two ways in which it can be made a part of the locative phrase. One is through the addition of the noun *pók* “mouth”:

- (25) *cìrèp dǒè tá-wáatò pók túccò ...*  
 women all FUT-come place thrashing  
 “All the women will come to the place of thrashing.”

The locative function of an animate noun may be coded through a serial verb construction or through the addition of the morpheme *cíg* “body”:

3. Although the data are based on Frajzyngier 1989, the analyses presented herein are substantially different and supersede those.

- (26) *wáat-ná cíg tójè*  
 come-PRF body horse  
 “He came near the horse.”

If the verb is not inherently locative, the locative predication is marked by the preposition *tí*. The final vowel of the preposition may change under the influence of the first segment of the following noun:

- (27) *kpàttín mì-tók-kò cínà ló-i lók-kò ló tí*  
 men REL-kill-PRF food animal-DEF hung-PRF animal PRED  
*pórò dàmbàn-i*  
 tree damban-DEF  
 “Men who kill animals hang them on the tree of damban.”
- (28) *má gbúr-kò kán mór n-cákù-n tí cíg kúndùl*  
 TEMP mix-PRF CON oil CONJ-rub-ANAPH PRED body kundul  
 “When they mixed it with oil they will rub into the body of kundul.”

## 9. Mupun (West Chadic)

Mupun has a variant of locative predication, although the coding means and their properties are significantly different from those in other languages with locative predication.

The locative complement must always be preceded by a particle *n* or by a particle *a*, both analyzed as prepositions in Frajzyngier 1993.<sup>4</sup> The particle *a* occurs only in stative locative predications, and it must therefore be reanalyzed as a stative locative predicator. The particle *n* is somewhat more difficult to make generalizations about, as it occurs with motion verbs and as a marker of relationship between noun phrases. I propose here that the particle *n* is a locative directional predicator when occurring in locative predication and a locative/dative preposition when occurring in a non-locative predication. Here is the evidence.

### 9.1 Predicator *a* in Mupun

The form *a* in Mupun is locative stative predicator. The presence of *a* does not depend on the function and the properties of the nominal complement but rather on the non-directional properties of the predicate. Consequently, in the examples below, *a* is glossed as PRED rather than as PREP as in Frajzyngier 1993:

4. The analyses in Frajzyngier 1993 are superseded by the present analyses.



- (29) *to, be mo maŋ wur a tok yol a pe kwat*  
 well, CONS 3PL take 3M PRED shoulder get up PRED place hunt  
*nə har ká mo dəm ká wur n-tulu*  
 DEF till (H.) CONJ 3PL go CONJ 3M PREP-home  
 “He will be taken on their shoulders from the place of the hunt back to the village.”
- (30) *nen Mupun mo kad mo yo kwat lusim ba me mat*  
 people Mupun PL when 3PL go hunt leopard NEG QUANT woman  
*taa tok sik ká mo a ar kas*  
 fall greet CONJ 3PL PRED road NEG  
 “People of Mupun, when they go to hunt for a leopard no woman will greet them on the way.”

## 9.2 The directional predicator *n*

The marker *n* precedes the complement whenever the locative predication involves movement rather than state:

- (31) *wur taa n-yil*  
 3M fall PREP-ground  
 “He fell down.”

The main question with respect to the particle *n* is why it occurs in all predications involving verbs of movement and locative complements. The explanation of this fact rests on the nature of verbs of motion in Mupun. Each verb represents lexicalization of only one parameter of motion. Some verbs code only the direction with respect to the speaker; others code spatial orientation of movement with respect to the source, such as on a plane, descending, or ascending; others code initiation or conclusion of movement; and still others code spatial orientation with respect to the goal (Frajzyngier 1993: 236). This pattern of lexicalization correlates with locative expressions being realized by a rich system of serial verb constructions, but without the indication of direction with respect to some overtly stated locative center, whether the source or the goal. The particle *n* marks the direction with respect to the locative center.

In the following example the particle *n* precedes a toponym. In languages that have grammaticalized locative predication, toponyms are inherently [+locative], hence the use of the particle *n* does not depend on the properties of the complement:

- (32a) *yak-sə mu dəm dī n-Germany nā*  
 time-DEM 1PL depart there:ANAPH PREP-Germany DEF  
 “Then we went to Germany.”

- (32b) *be mo n-dəm ká wur n-tulu miskoom*  
 CONS 3PL FUT-go CONJ 3M PREP-home chief  
 “He will be taken to the home of the chief.”

The preposition *n* can be used as a marker of relationship within the noun phrase, corresponding to ‘NP at NP’:

- (33) *ama makaranta dā n-nsəhən kamlu mo mo ret*  
 but school REL PREP-S. building 3PL 3PL good  
 “But the school in S. for sure, their buildings are nice.”

This preposition came to mark the general benefactive and purpose function:

- (34) *n-pus can nə be kən fua mo mbə siwa mwes dā*  
 PREP-day circumcision DEF CONS kin 2M PL FUT drink wine REL  
*puun fua cet n-mo*  
 father 2M cook PREP-3PL  
 “On the day of circumcision, your kin will drink the wine that your father cooked for them.”

## 10. Lele (East Chadic): coding locative predication by serial verb constructions

### 10.1 The interest of the situation in Lele

Lele (East Chadic) has the category locative predication, as evidenced by the fact that there are different coding means deployed in locative expressions depending on the feature [locative] in verbs and nouns. If the complement is not inherently locative, Lele uses a postposition to code locative function. The stative locative predicate is marked only by the locative postposition. If the predicate is not inherently locative, the directional predication is marked by serial verb constructions.

### 10.2 Inherently locative predicates and inherently locative complements

Each verb of movement in Lele has different properties with respect to the parameters it encodes. One verb has lexicalized direction away from a deictic center; another verb codes movement toward a deictic center; and a third verb codes movement from inside a deictic center. Other verbs code just the manner of movement, without the direction:

- (35) *è “go” jè “come”*  
*se “get up”*  
*án “leave”*

The evidence of the inherent locative meaning of these verbs is provided by clauses where one of the verbs is the only predicate. If the complement is inherently locative it is not marked by adpositions:

- (36) *wilèn dí è dí Bongor*  
 lack 3M go 3M Bongor  
 “He is not here, he left for Bongor.”

As in other Chadic languages that have locative predication, the noun *túgú* “home, compound, village” is inherently locative in locative predication:

- (37) *è-gé túgú sì-gé*  
 go-3PL village drink-3PL  
 (*túgú* is one’s home, so any movement home means “return”)  
 “They return home, they drink ...”

The importance of Lele is that the locative predication hypothesis allows one to explain the existence of the serial verb construction. If the predicate is not inherently locative, a serial verb construction is used to code locative predication. The serial verb construction has the main verb followed by the verb *è* “go” for direction toward the goal:

- (38) *àlá gír è yàá kolo-η bé kùrmbàlo*  
 but run go tell word-DEF DAT chief  
 “But instead she ran and informed the chief.”

Without a verb coding directionality, the verb *gír* cannot be used in locative predication:

- (39) *\*àlá gír yàá kolo-η bé kùrmbàlo*  
 but run tell word-DEF DEF chief  
 “But instead she ran and informed the chief.”

The combination of movement, direction, and manner involves the use of several verbs, in the order: Manner Movement Direction:

- (40) *gír è jè hàn*  
 run go come here  
 “He ran in here.”

Movement away from a source is coded by the verb *se* “get up, raise”. Arrival at a place is coded by the verb *jè* “come” albeit in most cases following the verb *è* “go to”.

### 10.3 Coding the locative complement through postposition

If the noun is not inherently locative, the locative complement must be marked by the postposition *ni* (*ní* in Weibegue & Palayer 1982), which follows the complement:

- (41) *na dú è sógú ni*  
 HYP 3F go toilet LOC  
 “She pretended she was going to the toilet.”

The marker *ni* cannot be added if the locative phrase contains a toponym, which indicates that the marker *ni* codes the preceding noun phrase as a locative complement:

- (42) \**wilèn dí è dí Bongor ni*  
 lack 3M go 3M Bongor LOC  
 “He is not here, he left for Bongor.”

An inherently locative noun followed by a possessive pronoun behaves like an inherently non-locative noun and must be marked by the postposition *ni*. This is the case with the noun *túgú* “home”:

- (43) *kà[w] ngà túg kò-m ni gólè kùní kùlè-ndì*  
 go:IMP 1DU.INCL home GEN-2SG LOC see:FUT home interior-3M  
 “Let’s go to your home [and] we will see the interior of the house.”

The locative postposition can be the only marker of locative predication in a stative locative:

- (44a) *kòbró dùgì ín-dí-gè kama ni*  
 pirogue sink ASSC-GEN:PL-3PL water LOC  
 “The pirogue sank with them in the water.”
- (44b) *ày bùgú kàb-iy ni an ná kumno kam-dì*  
 ake bag hand-3M LOC leave ASSC sky/God water-3M  
 “He took a bag into his hands [and] he left in the rain.”

### 10.4 Animate locatives

A [+animate] noun must be marked for the locative function by the preposition *dà*, by the postposition *ni*, and by another noun that does contain some locative characteristics. Such nouns serve as modifiers of the [+animate] noun or, alternatively, the animate noun serves as a modifier of the locative complement. Two nouns are deployed as locative coding means: *já* “side” in inalienable possessive constructions and *túgú* “village, home” in alienable possessive constructions:

- (45a) *ɲ se dà bónú-m já-y ni*  
 1SG leave PREP brother-2M side-3M LOC  
 “I come from your brother.”
- (45b) *ɲ se dà tǔgú bónú-m kè-y ni*  
 1SG leave PREP village brother-2M GEN-3M LOC  
 “I come from your brother.”

An animate noun may be coded just like other non-locative nouns when it is only in the domain *de dicto*, i.e., when no specific place is involved:

- (46) *gír go làmndá ba gùmnò ni* (a proverb)  
 escape REF elephant fall buffalo LOC  
 “Escape from an elephant, fall upon a buffalo.”

## 10.5 Summary of the coding of locative predication in Lele

Lele has the category of locative predication. If the predicate is not inherently locative, the locative predicate is coded by a serial verb construction using the verbs *è* “go” or *je* “come to”. If the complement is not inherently locative, it is marked by the postposition *ni*. If the predicate is inherently locative and the complement is also inherently locative, no other means of coding locative predication are used.

## 11. Hdi: Locative predication through locative prepositions

Hdi (Central Chadic) is verb-initial and uses the following means to code locative expressions: prepositions, verbal extensions, and tonal changes. Hdi distinguishes between the directional and stative locative predication through the use of different prepositions.

Extensions on the verb code a number of semantic categories, including direction toward a goal and movement away from the source (*ú*).

The evidence that Hdi has locative predication is provided by the fact that even though all locative complements are preceded by one of the two locative prepositions, the form of the directional preposition indicates whether the complement is inherently locative or not. The motivation for the deployment of the locative prepositions in Hdi lies not in the existence of locative predication but rather in the need to distinguish locative complements from other complements, motivated by the principle of functional transparency.

### 11.1 Prepositions *dá* and *dà*

Prepositions in Hdi are obligatory in locative predications, but the motivation for their existence comes from the default verb-initial linear order. The position after the verb is restricted to subjects or objects only. Any other role of the noun phrase must be marked by a preposition. The low tone on the directional preposition *da* indicates that the ensuing noun is not inherently locative. The high tone on the preposition *dá* indicates that the ensuing noun is inherently locative. This coding is an important piece of evidence that Hdi distinguishes between inherently locative and non-locative nouns and hence has a locative predication:

- (47) *lá-xà-dá dá xdí ...*  
 go-DOWN-1SG PREP Hdi  
 “Having gone to Hdi I ...”

The directional preposition *da* has no value with respect to the goal or source orientation, as evidenced by the fact that it can occur with both directions:

- (48) *yàgh-ká dá n`gh-ú dà sígà*  
 should not-2SG PURP look-SO PREP pot  
 “Do not look inside the pot.”

In Hdi, like in many Chadic languages, human and animate nouns are not locatives, but unlike in other languages they are not formally distinguished from other non-locative complements:

- (49) *kà lá-ghá-tsí dà ùvá*  
 SEQ go-D:PVG-3SG PREP cat  
 “And he went to Cat.”

If the verb is inherently non-directional there are two means of coding directionality. One is through a serial verb construction using a directional verb, e.g., *la* “go”, and the other is through verbal extensions.<sup>5</sup> What determines when one means is used rather than the other remains to be discovered.

Use of the serial verb construction with the verb *xwáyá* “run, escape” which is not directional. Directionality away from the source is coded by the directional verb *la* “go to” followed by the source-oriented extension:

- (50) *mbàd ká-’á kà xwáyá-úgh-tà lá-ghú dà zwán-ì*  
 then COMP-3SG SEQ run-SO-REF go-D:SO PREP child-PL  
 “Then he fled and he went to his children.”

5. Hdi has four series of verbal extensions, two of which have locative functions.

The directionality toward a goal is coded by the verbal extension *dá*, segmentally identical with the directional preposition *da*:

- (51) *dàgà rverè, dàgà gwi'yan kà ddà-dá-tá-xán dà*  
 CONJ (Hau) lion CONJ (Hau) elephant SEQ fall-ALL-REF-3PL PREP  
*vú mà xàdík*  
 fire PREP ground  
 “Lion and Elephant fell into the fire in the ground.”

Locative verbal extensions code spatial orientation with respect to the goal. In the following example the extension *m* codes movement toward the inner part of the goal:

- (52) *lá-mà pákává ghúvì dá xàdà mà tughwázàk kà*  
 go-IN hyena PREP place PREP hibiscus SEQ  
*hlà-ná-ghá-tá-tsí tá kři*  
 find-DEM-D:PVG-REF-3SG OBJ dog  
 “When Hyena entered the hibiscus, he found Dog.”

## 11.2 Stative locative predication in Hdi

The stative locative complement is coded by the preposition *tà* or by the segmentally identical suffix to the verb, analyzed in Frajzyngier & Shay 2002 as a referential marker and glossed REF. The preposition *tà* is just the locative marker, without any specification of spatial relationships. The preposition *tà* cannot be replaced by *da* in stative locative predications.

If the clause does not have a verb, the stative locative predication is marked only by the preposition *tà*:

- (53) *kí vli tà bèrék*  
 how place PREP Berek  
 (Berek ← barrack [Eng.] administrative quarter of Turu town)  
 “How are things at Berek?”

In the following example there are two locative complements: the noun *xàdì* “place” in the first line following the verb and *tughwázàk* “hibiscus” in the second, which is added as an afterthought. The first noun is not preceded by a preposition, but the verb ends in the marker *tà*. The second is preceded by the preposition *mà*, which codes the spatial specification “in”. The extension *tà* may well be the result of fusion of the preposition *tà* to the verb:

- (54) *lá-m-là dífà-úgh-tà xàdi yá, mà tùghwázàk xàd yá bá*  
 go-IN-go hide-SO-REF place DEM PREP hibiscus place DEM RESP  
 “Go hide yourself here, in the hibiscus, in this very place.”

### 11.3 Summary of the locative coding in Hdi

Hdi has a locative predication, as evidenced by the fact that it distinguishes between inherently locative complements and non-locative complements by the tone on the directional preposition. Hdi distinguishes between directional and stative locative predications. The non-directional predicates have a locative extension or serial verb construction to code locative predication.

## 12. Locative predication in East Dangla

All information on East Dangla comes from Shay 1999, although some interpretations in the present study may differ from those in that work. East Dangla appears to have the category of inherently locative predication, as evidenced by the fact that with inherently locative predicates the inherently locative complements, such as toponyms or the noun *gèr* “home”, there are no other markers of either stative or directional locative predication.

Directional predication:

- (55) *ɲaa-k élél-lúu káté Dyàmméen.*  
 COP-3M want.PRES-NEG go.VN N'Djaména  
 “He just wants to go to N'Djaména.”

Stative predication:

- (56) *lisín-àk tik-ga Dàñil-ik gòy gèr-tyò!*  
 Dadyo.PL-DEM let.PAST-3M.O Danglaman-DEM be.PAST home-3PL.POSS  
 “The Dadyo people let the Dangla man stay in their home!”

With inherently non-locative predicates, a number of prepositions (not locative predicators) are used to code locative predication. The preposition *mín* “from”, a borrowing from Arabic, indicates movement from a source. The preposition *ak* is a spatial specifier, indicating the space within the locative complement. Shay 1999 states that, when the argument marked by *mín* is not inherently locative, an additional coding device, either another preposition or the locative suffix *írá* must be used:



- (57) *yáa gùrì suù! miny-iy-tè mín ak kàa-y.*  
 even maggots INTERJ fall.PAST-3PL-REF PREP PREP head-3M.POSS  
 “There were even maggots falling out of his head.”

The preposition *ak* may have acquired the function of the locative predicator, as it can be followed by a noun “stomach”, whose function is to mark the inner space. It is used when the predicate is not inherently locative:

- (58) *kar nàa giy waar kó ak ad-èy*  
 SEQ 3M TOP dance.PRES already PREP stomach-3M.POSS  
*ka ger-tyò.*  
 DEM house-3PL.POSS  
 “Then he started dancing inside the house.”

Here is another example of the preposition *ak* with inherently non-locative predicates:

- (59) *à tyòóp-ga ak dyiimer.*  
 3M soak.PAST-3M.O PREP honey.OBL  
 “He soaked it in honey.”

The function of the suffix *írá* is not completely clear, as it may or may not occur with the same complement and the same predicate:

- (60a) *á-ye kát-íny-dyì sùgín-írá*  
 FUT-INCL go.VN-INF-3M.POSS market-LOC  
 “We will go to the market.”  
 (60b) *á-no káté sugine*  
 FUT-1SG go.VN market  
 “I’m going to go to the market.”

If the complement is animate it must be preceded by a marker making it locative, such as the noun *wer* “place”:

- (61) *diír no kát wer ka gàrpinàr*  
 yesterday 1SG go.PAST place DEM smiths.OBL  
 “Yesterday, I went to the place of the blacksmiths.”

The preposition *ku* appears to be used in directional predication when the complement is not inherently locative:

- (62) *mín-ik dòn-íy-tè ku and-ityò*  
 PREP-DEM return.PAST-3PL-REF PREP mush-3PL.POSS  
*kE, tèe-y-gà.*  
 DEM eat.PAST-3PL-3M.O  
 “They went back to their mush and ate it.”

### 12.1 Summary of the locative predication in East Dangla

If the predicate and the complement are inherently locative, the locative predication is coded by juxtaposition alone. If the predicate is not inherently locative, the preposition *ak* serves as locative predicator. There are, however, other prepositions and postpositions coding various types of locative expressions.

## 13. Summary of the evidence for the locative predication

The existence of locative predications in the languages discussed above is supported by the following facts. Some languages have inherently locative predicates and inherently locative complements. If a clause does not have an inherently locative predicate or inherently locative complement, other, compensatory means are used to code the locative feature. For the predicate, some languages have the predicator *a*, other languages use serial verb constructions, and still others have verbal extensions. For the complement, languages have grammaticalized a locative preposition that does not code any spatial relations. Given that, in each branch of Chadic there are languages that have the locative predication, we can postulate that a locative predication was a characteristic of Proto-Chadic. This conclusion is based on the reasoning that it is less likely that languages from three branches have independently grammaticalized locative predication, which is otherwise typologically rare, than the possibility that some languages from each branch have retained a function from the Proto-Chadic.

Presence of the locative predicator *a* in two branches of Chadic indicates that it may well be a retention from Proto-Chadic.

## 14. Further evolution of locative predication

### 14.1 The nature of the changes

In a number of languages, the locative predication ceased to exist as a separate predication and was replaced by a number of narrower locative expressions. The motivations for this change might have been language-internal and/or language-external, affected by language contact. Language-internal changes might have come about through the reanalysis of locative predicators as locative prepositions and thereby the loss of the notion of an inherently locative predicate.

Loss of locative predication might have also occurred as a result of language contact, whereby a Chadic language would borrow formal means and functions involving narrow locative expressions from neighboring languages. It is more than

likely that both factors might have been at work in the same language. In each branch of Chadic there are languages in which this process has taken place.

Although in each branch of Chadic there are languages without locative predication, this does not contradict the hypothesis that Proto-Chadic had a locative predication. In addition to the typological argument listed in the conclusion of the previous section it must be noted that even within languages that have lost the locative predication there are some features associated with locative predication that indicate that a locative predication used to be a part of grammatical system. One such language, Gidar (Central Chadic), is discussed below.

## 14.2 Gidar (Central Chadic)

The evidence that Gidar lost locative predication is provided by the fact that locative expressions are marked by prepositions regardless of the type of the predicate and regardless of the type of the complement in locative expressions.

The prepositions coding various locative expressions include *á* 'direction toward the goal and the presence at a place', *klà* 'through', *s* 'from the source', and the locative marker for animate nouns *za*.

Here is the evidence for the existence of locative expressions but the absence of the single domain of locative predication. Locative expressions with verbs that in other languages are inherently locative and nouns that in other languages are inherently locative nevertheless have a preposition:

- (63) *à-l-k á zà mǎlyà à-ná sǎ-nǎ-k ná dǎ-m*  
 3M-go-PRF PREP side chief 3M-say DAT-3M-PRF COMP go-1PL  
*á wrà nà-k ká-nà*  
 PREP bush DEM-DEM DEM-DEM

'He went to the chief, and he said to him, 'Let us go to that bush.'

- (64) *và-skà ná dǎf tà-y á wálànlà*  
 DEF-DEM COMP man PROG-3M PREP village  
 'Once there lived a man in a village ...'

Predicates that are not inherently locative also have a preposition preceding the non-locative complement:

- (65) *à-gǎmá-k gòrdú à-kpà-ǎ-k á màkrá-nì*  
 3M-take-PRF knife 3M-plunge-3M-PRF PREP heart-3M  
*mày à-mtǎ-kà*  
 ASSC.PL 3M-die-PRF

'He took a knife, plunged it into his heart, and died.'

- (65b) *ká-dé gón árahú-w á ddáf inkílè n-k-í*  
 2-go:VENT SUBJ find-1SG PREP in water DEM-DEM-PL  
 “You will come and will find me in this water here.”

Toponyms, which in other languages are inherently locative, are preceded by a preposition:

- (66) *mà-m tá-mbát-ák á gágám ákày sá-m ázámà*  
 mother-1PL 3F-go-PRF PREP Gagam search DAT-1PL to eat  
 “Our mother went to Gagam to look for something to eat.”

The locative source is coded by the preposition *sá* “from”. The fact that the preposition *sá* cannot co-occur with the preposition *á* provides evidence for its being in the same functional domain as *á*.

- (67) *à-mbát-ák sá jáabè*  
 3M-go-PRF PREP Jabe  
 “He went from Djabe.”

If the source is a human noun, the preposition is followed by *và* “hand”:

- (68) *sá zà à-gàpá-η á wrá dák àkàn dè hílíw*  
 from side 3M -reach-PL PREP field woman REL.F ASSC calabash  
*tà-ngálà-k hímbe á vá-t mbàdà-tá yáη tà-bón*  
 3F-ask-PRF ax PREP hand-3F in-law-3F CONJ 3F -refuse  
 “When they arrived at the field, the woman with the calabash asked her co-wife for the ax, but she refused.”

The preposition *sá* “from” can be used without any implication of movement of arguments; in other words, it can also have a stative meaning:

- (69a) *fàrádà sá zà wàlánglà*  
 far PREP side village  
 “Far from the village.”  
 (69b) *wín sá-t dèvè-t má-n ...*  
 boy PREP-3F inside-3F mother-3M  
 “A boy from inside his mother ...”

To express the locative predication with a noun that is not inherently locative, the noun must be preceded by a locative marker. One such marker is *zà* “side”:

- (70) *wín à-lá-k á zá fè-ní sòkò bbà sòkò bbà*  
 child 3M -go-PRF PREP side father-3M thanks dad thanks dad  
 “The child went and approached his father, ‘Thank you daddy, thank you daddy.’”

If the source or goal is a [+human] noun or pronoun and the goal is the permanent place of the noun, the locative complement is marked by the genitive preposition *ná* and the postposition *dà*:

- (71)

*à-gàp-án-k*

3M

-arrive-PL-PRF

*á*

PREP

*ná*

GEN

*áfə-t*

father-3F

*wán-k*

girl- F

*dà*

LOC

"They came to the father of the girl."

*à-dé-k*

3M-go:VENT-PRF

*sá*

from

*ná-w*

GEN-1SG

*dà*

LOC

"He came from me."
- 14.3 A summary of the locative predication in Gidar
- Gidar does not have a locative predication. Instead it has a variety of narrower locative expressions marked by prepositions, such as *à* marking direction “to” and stative “at”, *klà* marking movement through a space, and *sá* marking movement from a source.
15. Conclusions and implications
- Table 1 represents how the locative expressions have been grammaticalized in selected Chadic languages.
- Table 1. Grammaticalization of locative expressions in Chadic
- | Language    | Locative predication | Predicators     | Locative adposition (s) | Spat+LOC PREP     |
|-------------|----------------------|-----------------|-------------------------|-------------------|
| Hausa       | yes                  | <i>a</i>        |                         |                   |
| Pero        | yes                  | <i>ti</i>       |                         |                   |
| Mupun       | yes                  | <i>a</i>        | <i>n</i>                |                   |
| Mina        | yes                  | <i>a</i>        |                         |                   |
| Hdi         | yes                  | SVC, extensions | <i>dà, dá, tà</i>       |                   |
| Wandala     | yes                  | <i>a</i>        |                         |                   |
| Gidar       | no                   | <i>no</i>       |                         | <i>à, klà, sá</i> |
| Lele        | yes                  | SVC             | <i>ni</i>               |                   |
| East Dangla | yes                  | <i>ak</i>       |                         |                   |
- 2nd proofs

In languages from all three branches of Chadic there are traces of the locative predication. In particular, in all three branches there are languages in which, if the locative preposition is not used, the noun may still be interpreted as a locative complement. This interpretation obtains if the noun is inherently locative, such as toponyms and names for “house”, “compound”, and “village”.

In all three branches there are languages that have developed strategies to mark as locative a predicate that is not inherently locative. In some languages verbal extensions coding locative predication perform this function. In other languages locative predication is coded by serial verb constructions. The coding of locative predication is actually one of the motivations for the existence of serial verb construction, the existence of which has so far not received a satisfactory explanation. Finally, some languages have lexicalized the category locative predicator, whose only function is to mark predication as locative. Frajzyngier’s 1987 reconstruction of *a* as a locative preposition in Proto-Chadic should now be recast as locative predicator.

We can now explain why in some languages there is an obligatory preposition in stative locatives, different from prepositions used in directional locatives. The stative preposition is used regardless of whether the complement noun is inherently locative or not. In those languages, there are means to indicate the directional locative predication through serial verb constructions or through verbal extensions (Mupun and Hdi respectively), but these means do not code stative locative predication. The addition of a preposition is the only means to mark the locative predication.

This use of a preposition might have been one of the factors that led to the transformation from a language with locative predication to a language with multiple locative expression but no single domain of locative predication.

This analysis has further implications: In order to explain the forms of the utterances in a language it is important to discover what semantic functions have been grammaticalized. With respect to locative predications, discovering semantic functions that have been grammaticalized allows us to explain:

1. Why some languages have lexical categories that others do not have (case in point: the locative predicator of Mina, Hausa, and Pero);
2. Why lexical items having the same reference across languages have different properties;
3. Why in some languages there are only a few adpositions and in other languages there are many; and
4. Why an adposition is sometimes used and sometimes not used for the same meaning within the same language.

## Abbreviations

1	First person	HUM	Unspecified human subject
2	Second person	HYP	Hypothetical
3	Third person	IMPER	Imperative
ADDR	Addressee	IMPF	Imperfective
ADJ	Adjective	INCEPT	Inceptive
ALL	Allative	INCL	Inclusive
ANAPH	Anaphor	INF	Infinitive
AR.	Arabic	INTERJ	Interjection
ASSC	Associative	LOC	Locative
ATT	Attributive	M	Masculine
CAUS	Causative	N	Noun
COLL	Collective	NEG	Negative marker
COM	Comment marker	NG.	Ngambay
COM	Comment clause marker	NOM	Nominalizer
COMP	Complementizer	NP	Noun Phrase
COND	Conditional	OBJ	Object
CONJ	Conjunction	OBL	Oblique
COP	Copula	P	Proximate
D	Dependent	PART	Partitive
DAT	Dative	PAST	Past
DAT.OR	Dative orientation	PL	Plural
DEB	Debitive	POL	Marker of polite request
DED	Deduced reference	POS	Point-of-view of subject
DEF	Definite	POSS	Possessive
DEM	Demonstrative	PRED	Predicator
DIMIN	Diminutive	PREP	Preposition
DU	Dual	PRF	Perfective
DUB	Dubitative	PRO	Pronoun
EE	End of event marker	PROG	Progressive
EP	Epenthetic	PURP	Purpose
EXCL	Exclusive	Q	Interrogative marker
EXT	Extension	QUANT	Quantifier
F	Feminine	R	Remote
F.	Fula (Fulfulde)	REDUP	Reduplication
FOR	Preposition “for”	REF	Referential marker
FR.	French	REL	Relative marker
FREQ	Frequentative	REM	Remote previous mention
FUT	Future	SG	Singular
GEN	Genitive marker	SPEC	Spatial specifier
GO	Goal orientation	SQ	Specific question marker
H.	Hausa	STAT	Stative marker
HAB	Habitual	SUBJ	Subjunctive

T	Tone	UNSP	Unspecified
TOP	Topic marker	V	Verb
TOT	Totality	VENT	Ventive
TQ	Question about the truth	VN	Verbal noun

## References

- Frajzyngier, Zygmunt. 1987. "Encoding Locative in Chadic". *Journal of West African Languages* 17:1.81–97.
- Frajzyngier, Zygmunt. 1989. *A Grammar of Pero*. Berlin: Reimer.
- Frajzyngier, Zygmunt. 1993. *A Grammar of Mupun*. Berlin: Reimer.
- Frajzyngier, Zygmunt. 2001. *A Grammar of Lele*. Stanford Monographs in African Linguistics. Stanford: CSLI.
- Frajzyngier, Zygmunt. 2004. "Principle of Functional Transparency in Language Structure and Language Change". *Linguistic Diversity and Language Theories*, ed. by Zygmunt Frajzyngier, David Rood & Adam Hodges, 259–283. Amsterdam: John Benjamins.
- Frajzyngier, Zygmunt. 2012. *A Grammar of Wandala*. Berlin & New York: Mouton de Gruyter. doi:10.1515/9783110218411
- Frajzyngier, Zygmunt & Eric Johnston with Adrian Edwards. 2005. *A Grammar of Mina*. Berlin: Mouton de Gruyter. doi:10.1515/9783110893908
- Frajzyngier, Zygmunt & Erin Shay. 2002. *A Grammar of Hdi*. Berlin: Mouton de Gruyter. doi:10.1515/9783110885798
- Frajzyngier, Zygmunt & Erin Shay. 2003. *Explaining Language Structure through Systems Interaction*. Amsterdam: John Benjamins. doi:10.1075/tsl.55
- Pawlak, Nina. 2003. "Locative Prepositions in Chadic: Lexical or Grammatical Morphemes". *Motion, Direction and Location in Languages: In Honour of Zygmunt Frajzyngier*, ed. by Erin Shay & Uwe Seibert, 235–254. Amsterdam: John Benjamins. doi:10.1075/tsl.56.17paw
- Shay, Erin. 1999. *A Grammar of East Dangla*. Boulder, CO: University of Colorado dissertation.
- Weibegué, Christophe & Pierre Palayer. 1982. *Lexique lele-français*. Sarh: Centre d'études linguistiques.



**PAGE PROOFS**

© JOHN BENJAMINS PUBLISHING COMPANY

**2nd proofs**

# Unipartite clauses

## A view from spoken Israeli Hebrew

Shlomo Izre'el

Tel Aviv University

Within an integrative approach to the structure of spoken language, taking into account prosody, information structure, and syntax, a new model of **CLAUSE** is suggested, viz., a **UNIPARTITE CLAUSE**, where the only necessary and sufficient component is the **PREDICATE**, i.e., with no subject component required. By default, the predicate is viewed as the element carrying the informational load of the clause, the 'new' element in the discourse, and the focused component of the clause. The predicate carries the clause modality, where 'modality' is viewed in a broad perspective. A preliminary classification of unipartite clauses in Hebrew is also offered.

**Keywords:** syntax, clause structure, unipartite clause, predicate, context, spoken language, Israeli Hebrew

### 1. Introduction

In "Basic Sentence Structure: A View from Spoken Israeli Hebrew" (Izre'el 2012), I laid some foundations for a unified theory of clause structure in spoken Israeli Hebrew. A clause is defined as a unit consisting minimally of a predicate. Two main classes of clauses have been identified: (1) unipartite, consisting of a predicate only; (2) bipartite, where a clause consists – in its minimal manifestation – of a predicate and a subject. A clause includes only one predicate. Bipartite clauses show hierarchical structure, with three levels of predication, where the second and third levels include predicates which are in themselves complete clauses.

In this paper, I elaborate on the minimal disposition of a clause, viz., one that consists of only a predicate domain, where a subject does not form part of the clause. This type of clause will accordingly be labeled **UNIPARTITE**.

But before presenting my analysis of unipartite clauses in spoken Hebrew, I state again the premises that serve as a guide for my work on spoken language:

- Language is a tool of expression and communication. Its most frequent manifestation is human communication.
- Spoken linguistic varieties, notably the language of everyday conversation, are the most frequent of all linguistic systems. It is this capacity of spoken language that lends it the power to have its impact on all other linguistic systems and their development.
- Therefore, proper linguistic attention must be drawn to spoken language.
- Spoken language must be analyzed according to its own properties. We must detach ourselves from any preconceptions about the structure of language based on its written forms.
- Corpus data reflect the perceived language rather than the produced one. Therefore, linguistic description and analysis based on corpus data can lean solely on data as heard rather than as generated by the speaker.
- Linguistic analysis must regard language as a system on its own, notwithstanding its mutual-relationship with elements that are either external to the linguistic system or external to the immediate discourse.
- Prosody is a formal feature of spoken language no less than segmental features.
- Prosody is the main tool we use for spoken language segmentation.
- Syntax, information structure, and prosody integrate in spoken language structure, forming a coherent unity (cf. Izre'el forthcoming).

Using these premises as guidelines, I pursue the notion of UNIPARTITE CLAUSE and then sketch a preliminary classification to further establish the criteria for determining unipartite clauses. The data for this research is drawn from spontaneous speech recordings collected for The Corpus of Spoken Israeli Hebrew (CoSIH).

Let us begin with one example:

- (1) [1] sp2:  
       *kama            anafim   kiblu   elef        fekel   bevat\_axat ||*  
       how\_many people got.PL thousand shekel at\_once  
       “How many people got a bonus of thousand shekels?”
- [2] sp1:  
       *harbe ||*  
       many  
       “A lot.”
- [3] *harbe ||*  
       many  
       “A lot.”

- [4] *kəl ha=χameʃet alafim* |  
all the=five thousands  
“All five thousand;”
- [5] *kəl ha=mifmaʕot* ||  
all the=shifts  
“all shifts.”
- [6] sp3:  
*ələf ʃekəl kəl ɛχad* ||  
thousand shekel every one  
“A thousand shekels each.”
- [7] sp2:  
*ələf ʃekəl kuʕliti of lajf* /  
thousand shekel quality of life  
“A thousand shekels quality-of-life worth?”
- [8] *ɔ ələf ʃekəl meʕuman* ||  
or thousand shekel cash  
“Or a thousand shekels in cash?”
- [9] sp1:  
*ələf ʃekəl* |  
thousand shekel  
“A thousand shekels;”
- [10] *ələf ʃekəl le=malʕn* ||  
thousand shekel for=hotel  
“a thousand shekels for a hotel.”
- [11] sp3:  
*ələf ʃekəl meʕuman* ||  
thousand shekel cash  
“a thousand shekels in cash.”
- [12] sp1:  
*keʃ* ||  
cash  
“In cash.”
- [13] sp3:  
*le=malʕn* ||  
for=hotel  
“For a hotel.”
- [14] *le=bet\_malʕn* ||  
for=hotel  
“For a hotel.”

- [15] sp2:  
*waj* ||  
 wow  
 “Wow!”
- [16] *eze kef la=hem* ||  
 what fun to=them  
 “How wonderful for them.”  
 (OCD\_2\_sp2\_036–040; sp1\_017–023; sp3\_040–043)

In this exchange, quite typical of Hebrew casual speech, none of the units – except the first – conforms to common definitions of *CLAUSE* as a unit consisting of both subject and predicate. Utterances like the ones comprising this passage are amply attested in spoken languages (e.g., Biber et al. 1999: §§ 14.3.3–4; Cresti 2005; cf. Izre'el 2005: 4–5), but they are usually regarded as elliptical, reduced, or concise syntactic structures (for Hebrew, see Borochofsky Bar-Aba 2010: Chapter 2, § 5), if not excluded from the syntactic analysis altogether. Indeed, Carter & McCarthy (2006: 490) explicitly claim that “[t]he sentence is a unit of grammar, and must be grammatically complete (i.e., it must have at least one main clause). The utterance is a unit of communication. It ... does not need to be grammatically complete”. Biber et al. (1999: Chapter 14) use the term ‘non-clausal’ for units that do not conform to the traditional definition of a clause, yet they nevertheless feel the need to coin an “umbrella term ‘C-units’ for both clausal and non-clausal units; i.e., for syntactically independent pieces of speech” (Biber et al. 1999: 1070). This was done precisely because many of the units used in everyday speech do not fit in the “received receptacles”, to use Sinclair’s metaphor in his review of this magnum opus (2002: 357).

As mentioned, utterances that do not conform to the canonical definition of *CLAUSE* are usually regarded in the linguistic literature as if a virtual component is represented in the clause as a zero component or as if an allegedly missing component has gone through a process of ellipsis (e.g., Benayoun 2003; Foley 2006; Winkler 2006). This type of structure is so frequent among world languages (Givón 1983) that one wonders whether the phenomenon is indeed to be viewed as ellipsis. For Kibrik (2011: 44), “zeroes are not a theoretical construct but rather a convention of representation”. Nariyama (2007: 100) suggests an opposite way to look at ‘ellipsis’: “It is not that sentences are produced with ellipsis, but rather those words/information that are not retrievable from contexts are being verbalized”. Different paths of analysis have been tried. Vacillating between syntax, semantics, and pragmatics, debate over the analysis of so-called subsentences or fragments, elliptical structures, and the like has been going on since the outburst of generative grammar, putting aside what may be regarded as pre-structuralist statements over the nature of this type of units as forms of sentences (cf., inter alia, the discussions by Segel 2008: §§ 1–3; Hall 2009; Harnish 2009; with references to previous works).

Ironically, already Jespersen (1924: 306) notes that “[a]n old-fashioned grammarian will feel a certain repugnance to this theory of one-member sentences”. Although classifying such types of sentences as “inarticulate” or “semi-articulate” (op. cit., p. 308), he still notes (p. 306) that “a one-word sentence is at once a word and a sentence, just as a one-room house is from one point of view a room and from another a house, but not something between the two”.

A notable exception to recent traditions struggling with analyses of ‘fragments’ comes from French scholarship, where utterances that do not fit the concept of predication between two components can still be regarded as sentences (e.g., Tesnière 1966: Chapter 45; Le Goffic 1993: § 351; Lefeuve 1999: *Troisième partie*; Blanche-Benveniste 2006: § 3; cf. § 3 below).

Taking the point of view of the recipient, I do not refer to nonexistent elements as if elided or missing. I will try to find a path through which we can reach a unified theory that will encompass all the evidence provided by spontaneous speech data as regards units that do not include predication and therefore are usually not regarded as (complete) clauses (or sentences). In other words, I will incorporate unipartite clauses into a theory of clause structure (Izre’el 2012).

Before going into detail, a word on prosody and its interface with discourse structure and syntax is needed.

## 2. Prosody, discourse and syntax

As mentioned, I build on three premises as regards prosody:

1. Prosody is a formal feature of spoken language no less than segmental features.
2. Prosody is the main tool we use for spoken language segmentation.
3. Syntax, information structure, and prosody integrate in spoken language structure, forming coherent sequential units.

From the recipient’s perspective, prosody is a *sine qua non* when trying to delimit units of spoken language (Mettouchi et al. 2007; Izre’el & Silber-Varod 2009). For our needs, it will suffice to define two units in the prosodic hierarchy: PROSODIC MODULE and PROSODIC SET.

The PROSODIC MODULE (henceforth: PM; aka ‘intonation unit’, ‘tone group’, ‘prosodic group’, or the like), which has been determined as having a coherent intonation contour (Chafe 1994: 57–60), encapsulates a segmental unit of language to be termed SEGMENTAL MODULE, forming together an INFORMATION MODULE (IM) (cf. Tao 1996: §§ 9.1–2 for what he terms SPEECH UNITS). The boundaries of IMs are therefore defined by prosody. There are two main classes of boundaries: major (which indicates terminality) or minor (which indicates continuity). Both

are indicated by their respective boundary tones. A major boundary is also the boundary of a PROSODIC SET.

PROSODIC SET is defined as a stretch of speech ending – as its default manifestation – in a major boundary. A PROSODIC SET can consist of one or more PMs of which the last ends in a major boundary, whereas any (optional) previous PM ends in a minor boundary.

Whereas a PM encapsulates a SEGMENTAL UNIT and forming together an INFORMATION UNIT (IM), a PROSODIC SET encapsulates an UTTERANCE (cf. Cresti & Moneglia 2005: § 1.2). As regards syntax, it is suggested that the utterance is the default domain of the clause (or sentence), whether it consists of a single IM or more. An utterance can consist of more than a single clause. An IM can consist of either a phrase, being a component of a clause, or of a complete clause. The interface between prosodic and segmental units is outlined in Table 1.<sup>1</sup>

Table 1. The interface between prosodic, discourse, and syntactic units

Prosodic units	Discourse units	Syntactic units
Prosodic Module (PM) (one of two or more in a Prosodic Set)	Information Module (IM) (one of two or more in an utterance)	Phrase / Clause (/ Clause cluster)
Prosodic Set	Utterance	Clause / Clause cluster

3. What is a unipartite clause?

By default, the predicate is viewed as the element carrying the informational load of the clause, the “new” element in the discourse (cf. Chafe 1994: 108), and the focused component of the clause. The predicate can thus be identified with the rheme. Essentially, the predicate carries the modality of the clause. The view of modality as an inherent, indispensable characteristic of the clause follows the path of French linguistic schools (Bally 1965: 36; Le Goffic 1993: Chapter 4; Lefevre 1999: Chapter 1; Martin 2009: ch. II/1; convenient surveys can be found at Vion 2001; Johansson & Suomela-Sahni 2011). Modality thus has a much broader scope than it is usually conceived by other schools, notably Anglo-Saxon linguistic schools (e.g., Palmer 2001; Butler 2003: Chapter 9), and includes not only the commonly known, consensual types of epistemic and deontic modality but also assertion, polarity (cf. Halliday 2004: 147; Butler, loc. cit.), and beyond.

1. For further details see Izre'el forthcoming; cf. Izre'el & Mettouchi 2015 (where different terminology has been used).

By default, the predicate carries with it assertive modality. The traditional notion of assertion has always been central to the definition of predication (Goldenberg 1998: 156–157). The thesis advanced here is that a unipartite clause does not have to lean on an implicit subject (pace Lefeuve 1999: Chapter 5). Therefore, when a subject does not form part of the clause, the load of assertion is carried exclusively by the predicate. The same can be said of other types of modality as it is conceived here and indeed of modality in its entire gamut.

To the range of modality types such as assertive, affirmative, negative, interrogative, evidential, injunctive, and so on, the framework offered here requires that more types of modality be introduced. One example of such extra types is VOCATIVE, usually regarded as a case or as a form of expression excluded from syntactic analysis ('extragrammatical', as labeled by Daniel & Spencer 2009; for English vocatives see Biber et al. 1999: § 14.4.1; Halliday 2004: § 4.3.4, who describes vocatives as outside the scope of the mood system; Carter & McCarthy 2006: §§ 116–118). That an address or calling attention like "Jack!" or "Sir!" should be regarded as modal will be understood if we realize that it is in fact a REQUEST to pay attention. If an address like these forms an entire utterance or comprises in itself an IM, it would carry its own independent intonation contour, forming an independent PM. In such cases, the intonation contour will be observed as indicating the modality of the IM. Of course, such an IM carries informational load with it, if it forms a separate PM it will usually be focused, and in some cases it will manifest 'newness' of the address form in terms of the discourse flow (cf. Chafe 1994: Chapter 9). Therefore, vocatives such as these will be regarded as unipartite clauses.

(2) exhibits some typical unipartite clauses:

(2) [1] sp2:

*mɔʁuʃ* ||  
Morush  
"Morush,"

[2] sp1:

*ma mɔʁɛk* ||  
what sweetie  
"What, sweetie?"

[3] sp2:

*aʁbaa jamim* |  
four days  
"(For) four days –"

[4] *fva mɛʁt fɛkɛl lɛ=zug* ||  
seven hundreds shekel to=couple  
"(the cost is) seven hundred shekels for two."



- [5] sp1:  
*bli kesef* ||  
 without money  
 “(This is) very cheap.”
- [6] sp2:  
*naḥan* /  
 right  
 “Isn’t that so?”
- [7] sp1:  
*εjf* /  
 where  
 “Where?”
- [8] sp2:  
*bε=hɔlidej\_in ha=χadaf* ||  
 in=Holiday\_Inn the=new  
 “At the new Holiday Inn.”
- [9] sp1:  
*daj* ||  
 enough  
 “Wow!”  
 (OCD\_2\_sp2\_057–061; sp1\_029–030)

Each of the utterances in lines [1], [2], [5], [6], [7], [8], [9] (which in this case each consists of a single IM) meets the requirements of the definition of a predicate and thus constitutes a (unipartite) clause: each conveys new information and each carries modality: vocative (IM [1]), interrogative (IMs [2], [6], [7]), assertion (IMs [5], [8]), or exclamation (IM [9]). Also, each of the predicates is focused. IMs [3]–[4] make an interesting case. IM [3] recalls a short exchange regarding a weekend at a hotel which took place almost two minutes before returning to this issue here. At this point in the conversation it is invoked not by repeating the exact words used before (“weekend”) but by indicating the time span of the hotel stay, viz., “four days”. Therefore, this IM seems to introduce a new piece of information into the discourse, thus constituting a unipartite clause. The modality carried by this predicate is somewhat obscured by the minor boundary tone. Had it been a major boundary tone, there would be no doubt about the assertion expressed by this IM, making it into a clear declarative clause, meaning something like “(It is) four days” or “(We have) four days (at the hotel)”. The minor boundary tone, which indicates continuity, is needed for showing the link between this IM and the following one (IM [4]), which in itself unmistakably conforms to the criteria suggested above for a unipartite clause. For similar and related structures see further Izre'el forthcoming.

It should be pointed out that each utterance, which has been defined as a stretch of speech encapsulated by a prosodic set, is by definition delimited by a major prosodic boundary, which accordingly indicates its terminal point. As such, an utterance is the largest discourse unit that can contain a clause. Looking at it from a different angle, a major prosodic boundary always indicates the end of a clause and therefore also the beginning of a new clause in the following utterance (prosodic set). As it is exemplified in (2), each utterance includes a predicate, i.e., the informative, new, focused component in the clause that carries with it the clause's modality. Thus, the interface between prosodic units (prosodic sets), discourse units (utterances), and syntactic ones (clauses) is established via the integrative approach that guides us in our endeavor to search for the basic units of spoken language.

In (3), IMs [1], [5], [7]–[8], [9], [10], [11], [12] are examples of unipartite clauses.

- (3) [1] sp1:  
*ma* |  
 what  
 “What?”
- [2] *jaani hem baim –*  
 That\_is they-M come.PTCP.MPL  
 “You mean, they come”
- [3] *baim bæεε ʃalʃ babəkεε ε* |  
 come.PTCP.MPL in\_which three in\_the\_morning uh  
 “come like 3 a.m. uh”
- [4] *aχβε miklaχat* |  
 after shower  
 “after shower?”
- [5] sp2:  
*mɔvu* |  
 Moru  
 “Moru,”
- [6] *aχβε* |  
 after  
 “after ...”
- [7] *lɔ kak bε* |  
 Not only in  
 “not only at”
- [8] *laɟla* ||  
 night  
 “night.”

- [9] *gam bejam[] bejom ||*  
 also in\_da[] in\_day  
 “During the da... the day too.”
- [10] *beʔʔt jom ||*  
 in\_hours\_of day  
 “During daily hours.”
- [11] *beavʔdat jom ||*  
 in\_work\_of day  
 “During day-working time.”
- [12] sp1:  
*eze hagzama ||*  
 which exaggeration  
 “What an exaggeration!”  
 (OCD:2480”–2490”)

Sp1 opens with an interrogative particle *ma* “what?” (IM [1]), which is elaborated contents-wise in the following three IMs (IM [2]–[4]). It might be objected that IM [1] is a substantive one, as it may be seen as one that regulates the discourse flow, although independent IMs with the same content do exist, and when they express surprise they do serve to convey new contents and carry modality. By virtue of this such units can be regarded as clauses, whether in substantive IMs or in regulatory ones. This is the case here, although the utterance (encapsulated by a prosodic set) has not concluded yet (see Izre'el forthcoming).

As is the case with IM [1] of (2) above, the vocative in IM [5] comprises by itself an entire IM and can be seen as satisfying the requirements for constituting a clause. The minor boundary tone should not divert us from this conception, as the modality of this clause is indicated by the intonation contour as a whole. IM [6] is a suspended prosodic set of which the utterance has not reached conclusion. IMs [7]–[8] (forming together a single utterance), [9], [10], [11] are all complete utterances that convey each some new information and carries assertive modality, and each is prosodically focused. All these clauses are responsive to sp1's question. None of them includes a subject, only a predicate. Finally, IM [12] is an exclamative clause, reactive to sp2's message and – like all previous clauses – consists of only a predicate.

Many unipartite clauses are anchored in a previous discourse. Givón (1992) has shown a significant correlation between the occurrence of clauses without representation of the referent and referential distance, i.e., the gap between the current and last appearance of the referent in the discourse. From data collected in several languages, Givón shows that the mean distribution of clauses without an explicit representation of the referent (for him: zero anaphora) will reach up to 100% of the occurrences when they immediately follow a referential representation in a previous clause. On the other hand, referents tend to be overtly and explicitly represented

in the discourse the larger the gap from a previous occurrence of the same referent becomes (see his table on p. 21). With a somewhat different approach, I undertake a preliminary, broad classification of unipartite clauses in spoken Israeli Hebrew, aiming primarily at further establishing the criteria for determining the notion of unipartite clause.

#### 4. Classification of unipartite clauses

Every discourse takes place in a specific location, occurs at a specific time, and has its direct interlocutors, indicated in the discourse by the first and second personal pronouns. This is the point of departure for all deixis, the *ORIGO* ('origin'), to use Bühler's (1934) term (Abraham 2011: xviii). An intricate system of means is used to refer to elements in the conceptual world by linguistic signs, whether such elements are external to the discourse or occurring within it. Discourse structure uses a variety of deictic and anaphoric elements to refer to these items, notably when reference recurs in the discourse. Recurrent reference may be called for by reduced referential expressions (e.g., pronominal clitics or affixes) or may not be explicitly made at all. In fact, there are many languages which systematically avoid the use of referential expressions (Kibrik 2011: Chapter 3). Within the boundaries of a clause, reference can be made either in the subject position or in the predicative domain or in both. Of course, our interest here lies with clauses where no subject is present. I hope to show that unipartite clauses are not dependent on referential representation at the subject position.

The classification suggested below shows whether or not the predicate can be seen as anchored in referential expressions beyond the clause domain, and where it does – where that anchor will be located in the discourse structure. I shall examine whether predicate anchoring can be established only to specific referential expression, ones that are grounded in specific referents in the conceptual world, or whether anchoring to other discourse components such as predicates and therefore to whole clauses is also possible (cf. Kibrik 2011: §§ 2.1–2.2). Some predicates can be shown to have no anchor in any specific intra-discursive component but in (a part of) the discourse itself (Givón 1992: § 6.5.5). These clauses will be classified as having broad anchoring. All these types of anchoring will be classified as *INTRA-DISCURSIVE*. In other cases, the predicate cannot be shown to have an anchor in elements that have explicit linguistic expression in the discourse; rather they are anchored in elements that are external to the discourse, either within the *ORIGO* of this specific discourse or external to it (cf. Givón 1992: § 6). These latter types of anchoring will be classified as *EXTRA-DISCURSIVE*. The suggested broad classification of anchoring of predicates in unipartite clauses will accordingly be as follows:

- Anchored (§ 4.1)
  - Intra-discursive anchoring (§ 4.1.1)
    - Proximate (§ 4.1.1.1)
      - Structural (§ 4.1.1.1.1)
      - Non-structural (§ 4.1.1.1.2)
    - Remote (§ 4.1.1.2)
      - Structural (§ 4.1.1.2.1)
      - Non-structural (§ 4.1.1.2.2)
    - Broad (§ 4.1.1.3)
  - Extra-discursive anchoring (§ 4.1.2)
    - Intra-ORIGO (§ 4.1.2.1)
    - Extra-ORIGO (§ 4.1.2.2)
- Unanchored (§ 4.2)

This broad classification does not go into details as regards the function of each clause in the discourse (such as response, reaction, alleged predicate to a previously mentioned referent, etc.). Refining this classification must remain for further investigation, although some hints on such functions have already been given above and some more will be given in the following sections. Taking the point of view of the recipient, our aim here is purely descriptive, leaving aside any explanatory endeavors.

## 4.1 Anchored

### 4.1.1 *Intra-discursive anchoring*

Predicates can be anchored in the discourse either in an adjacent clause (PROXIMATE) or a more distant one (REMOTE). As mentioned above, Givón (1992) has shown correlation between the representation of a referent in a clause and its proximity to the representation of the same referent earlier in the discourse, to the extent that the mean distribution of clauses without an explicit representation of the referent (Givón's zero anaphora) will reach up to 100% of the occurrences when they immediately follow a referential representation in a previous clause. It will be interesting to check a similar type of correlation also in Hebrew. At this time, I suggest a binary classification between proximate and remote anchoring, where PROXIMATE includes only adjacent clauses and REMOTE includes all others. Either proximate or remote anchoring can be structural or non-structural. By structural anchoring I mean that the anchored predicate can be seen as having a virtual or a potential syntactic link with another element in the discourse. Finally, broad anchoring will be exemplified.

4.1.1.1 *Proximate*4.1.1.1.1 *Structural*

- (4) [1] *ata lɔkeax smɔla* ||  
 you.SGM take.PTCP.SGM leftward  
 “You turn to the left.”  
 [2] *niχnas b=a=delet* |  
 enter.PTCP.SGM in=the\_door  
 “(You) enter the door,”  
 (C842\_sp2\_150–151)

In (4), IM [2] includes a predicate, *niχnas* “enter”, but does not show an overt subject. An overt referent, to which this predicate can be ascribed, can be located in the previous clause, viz. *ata* “you-sgm”. Similar types of predicate-only clauses are well known from the linguistic literature and are usually analyzed as elliptical (Winkler 2006). One should note that IM [1] constitutes an utterance on its own, as the PM that encapsulates it ends in a major boundary. This utterance constitutes an independent clause. IM [2] opens a new utterance following a break of 1261 ms. In the framework proposed here, such clauses are not regarded as including a zero subject or as representing ellipsis of any kind. The approach taken here sees the link between the predicate *niχnas* “enter” and its referential anchor not on the syntactic level but on the discourse level (cf. Givón 1992).

As in (4), IM [1] of (5) ends in a major prosodic boundary and encapsulates an IM that constitutes a complete clause; IM [2] includes what is usually regarded as an ‘afterthought’:

- (5) [1] *at mavixa et=ha=veax fel=ha=fampɔ* ||  
 you.SGF smell.PTCP.SGF ACC=the=smell of=the=shampoo  
 “You smell the shampoo.”  
 [2] *miʃnehem* ||  
 from\_both  
 “From both of them.”  
 (OCD:2492.5”–2495.2”)

Prima facie, the term ‘afterthought’ implies only that a stretch of speech follows another one and seems not to differ from ‘right dislocation’, which may imply the same. However, Ziv & Grosz (1994: §2) have suggested that an afterthought and ‘right dislocation’ differ in function and in some formal characteristics, noting that an afterthought comes after a prosodic boundary and comprises a separate utterance. As we see in (5), a major prosodic boundary indeed separates between the two speech stretches, thus forming two distinct utterances. Complying further with the requirements of newness of information, focusing, modality (assertive in this case),

the phrase *mifnehem* “from both of them” in IM [2], standing as an utterance on its own, will be regarded from the syntactical point of view as a predicate constituting a unipartite clause. Looking at it from the point of view of parts-of-speech classification, the structure of the word that constitutes this clause is one that will be defined as an adverbial phrase. Taking this point of view, as well as looking at the semantic structure of the utterances in both IM [1] and IM [2], one can see that the utterance *mifnehem* “from both of them” in IM [2] is structurally related to the predicate nucleus *mavixa* “smell” in IM [1]. Of course, a virtual syntactic link between the predicate in IM [1] and the adverbial phrase in IM [2] can also be deduced, one that can be tested had the two occurred within the boundaries of a single utterance (or clause). In that case, the adverbial phrase would not be regarded as a predicate of a new clause but as an adjunct, since it would not carry its own modality. In Hebrew, we find any part of speech in predicate position: nominal (substantives, adjectives), participial (active or passive), pronominal (personal pronouns, demonstratives, interrogatives, and other pronouns), adverbs, and prepositional phrases, as well as larger phrases, clauses, and other types of syntactic complexes (Izre'el 2012: § 3). In the present framework, where prosody is taken as the basis for segmentation of both discourse and syntactic units, as well as on the basis of the analysis above where the adverbial phrase *mifnehem* is taken to be a predicate, the relationship between the two utterances must be seen not on the syntactic level but on the discourse level.

#### 4.1.1.1.2 Non-structural


- (6) [1] *lama* |  
 why  
 “Why?”  
 [2] *ma ka=l=i pɔ* ||  
 what bad=to=me here  
 “I am quite happy here.”  
 [3] *miklaχat=fɛl=i* ||  
 shower=of=me  
 “(I have) my shower.”  
 [4] *miklaχat=fɛl=i* ||  
 shower=of=me  
 “(I have) my shower.”  
 [5] *ani=ɔhev ɛt=fɛl=i* ||  
 I=love ACC=of=me  
 “I love mine.”

Contrary to the case presented in (5), there is no structural relationship between IM [3] (and the repetitive one IM [4]) in (6) and any other clause or component

in the preceding discourse. The related component is the locative adverb *pɔ* “here” in the previous clause (IM [2]), which is a proximate, non-structural anchor to the unipartite clause in IM [3].

#### 4.1.1.2 Remote

##### 4.1.1.2.1 Structural

- (7) [1] *jɛf ʃnɛ nɛhagim* ||  
ext two drivers  
“There are two drivers.”
- [2] *jɔʃvim al=ha=hɛgɛ* |  
sit.PTCP.PL on=the=wheel  
“They sit at the wheel,”
- [3] *hɛm lɔ mitxɔlfim* ||   
they NEG change-PTCP.PL  
“(they) do not change.”
- [4] *ad=ʃɛ=ɛxɔd nɪʒdam* |  
till=that=one falls\_asleep  
“Until one (of them) falls asleep,”
- [5] *mitxɔlfim* |  
change.PTCP.PL  
“(then) they change.”
- [6] *ɔ ʃɛ=jɛf hɔʃsaka* |  
or that=ext break  
“Or, when there is a break,”
- [7] *hɛm ɛ zɛ* |  
they uh this  
“they uh well ...”
- [8] *b=a=panʃɛʃ* |  
in=the=puncture  
“When a puncture occurs,”
- [9] *metaknim ɛt=ha=panʃɛʃ* |  
repair.PTCP.PL ACC=the=puncture  
“they take care of the puncture,”
- [10] *mamʃixim* |  
continue.PTCP.PL  
“(then they) continue.”
- [11] *zɛ maʃɛhu madhim* ||  
this something amazing  
“This is amazing.”  
(OCh\_sp1\_161-172)



The participle *mitχalfim* “change” in IM [5] is structurally, albeit not syntactically, related to the pronoun *hem* “they” in IM [3], where a similar (negated) participle *lō mitχalfim* functions as the (immediate, syntactic) predicate to this pronominal subject. One should note that IM [3], which is encapsulated by a PM ending in a major boundary tone, forms an utterance on its own, which constitutes a complete clause. In between the pronoun *hem* in IM [3] and the participial predicate of the unipartite clause in IM [5] there is a temporal adverbial clause in IM [4]. The same relationship can be detected between the pronoun in IM [7] and the participial predicates in IM [9] and in IM [10]. Let me emphasize again that in the framework proposed here, none of these clauses should be seen as including a zero subject or as representing ellipsis of any kind. In either case, the pronoun cannot be seen as subject of the respective remote predicate, since it does not belong to the same clause. The relationship between these pronouns and the respective predicates must be seen not on the syntactic level but on the discourse level.

#### 4.1.1.2.2 *Non-structural*

The text in (8) follows a short narrative discussing riding horses and its aftereffects, which in itself follows a narrative about a trip that the speaker and his friends had taken, reaching a quite remote northern spot.

- (8) [1] *kɔɤ* |  
coldness  
[2] *mavet* ||  
death  
“(It was) deadly cold.”  
[3] *hajinu* |  
we\_were  
“We were ...”  
[4] *nitkanu*      *ʃam*   *bɛ* |  
we\_got\_stuck there in  
“We got stuck there in ...”  
[5] *zɛ*   *kvaɤ*   *haja*  
this already was  
“It was already”  
[6] *ɛmtsa*   *septembɛɤ* ||  
mid September  
“mid-September.”  
(OCh\_sp1\_240–245)

The utterance *kɔɤ* | *mavet* || “(It was) deadly cold” (IMs [1]–[2]) is anchored in a narrative discourse that ended in OCh\_sp1\_234, i.e., just before the intervention

on horse riding. The word *hatsafna* “northward” occurs in OCh\_sp1\_233, while the mentioning of Siberia occurs way back in OCh\_sp1\_217. This unipartite clause therefore is discourse bound; it refers to a remote referential expression without having any (virtual or potential) structural link to it.

#### 4.1.1.3 *Broad*

The text in (9) comes after the following exchange, having taken place during a car drive going home after a wedding. Sp3 is the driver:

Sp2: I wanted you to turn right here, taking a different route.

Sp3: It doesn't matter. So, we'll take another route, OK? Why are you so stressed?  
Your husband is not home, and you are staying over with me.

- (9) [1] sp2:  
        $l\gamma \quad \varepsilon t = z\varepsilon \parallel$   
       NEG ACC=this.SGM  
       “Not this.”  
       [2]  $bifvile\chi \parallel$   
       for\_you.SGF  
       “For your sake.”  
       (OCD\_1\_sp2\_003–004)

The meaning of these two utterances is: I didn't say it because I was worried or stressed, but for your sake, so that you would go via a shorter route.

Neither the pronoun  $z\varepsilon$  in IM [1] nor the entire clause refers to any specific referent but to a discourse stretch which starts with sp2's turn (“I wanted you to turn right here, taking a different route”) and continues with the second part of sp1's turn (“Why are you so stressed? Your husband is not home, and you are staying over with me.”) (cf. Givón 1992: § 6.5.5). Each of the two utterances cited in (9) consists of a meta-discursive clause. Therefore, I propose to regard these two utterances, although discourse related, as detached from any specific intra-discourse component and classify their anchoring as BROAD. The same applies to the unipartite clause in IM [2]. Thus, both utterances cited in (9), constituting together a meta-discursive note, have their anchor in a discourse stretch rather than to a single referent, or rather a specific intra-discourse component. Therefore, I propose to classify their anchoring as BROAD.

#### 4.1.2 *Extra-discursive anchoring*

In the previous examples, elements in the discourse have been shown to have their anchors in elements that have an explicit expression in the discourse. This section includes elements that are anchored not in linguistic elements but in elements that are external to the discourse, therefore labeled EXTRA-DISCURSIVE ANCHORING.

#### 4.1.2.1 *Intra-origo*

The unipartite clause in (10) is uttered by a visitor upon entering into the study of his host, and is obviously anchored in the location where the discourse is taking place:

- (10) *makɔm nɛχmad* ||  
 place nice  
 “Nice place.” (i.e., “(This is) a nice place.”)  
 (C612\_4\_sp2\_001)

#### 4.1.2.2 *Extra-origo*

Looking at an atlas, the speaker spots locations he had visited during a trip to China and Mongolia a few years earlier while uttering their names:

- (11) [1] *hinɛ sɛʃʃuan* |  
 PRES Sichuan  
 “Here (is) Sichuan,”  
 [2] *junan* |  
 Yunnan  
 “Yunnan,”  
 [3] *ɛʃɔ junan* |  
 where Yunnan  
 “where is Yunnan?”  
 [4] *hinɛ* |  
 PRES  
 “Here (it is)!”  
 [5] *guanʃi* |  
 Guanshi  
 “Guanshi,”  
 [6] *ɛzɛ jɔʃi* ||  
 what beauty  
 “awesome!”  
 [7] *ja.ala* ||  
 wow  
 “Wow!”  
 [8] *χatgal* ||  
 Hatgal  
 “Hatgal.”  
 (OCh\_sp1\_027–034)

After discovering Guanshi on the map (IM [5]), the speaker discovers Hatgal. Before uttering the name of Hatgal (IM [8]), he calls in excitement: *ɛzɛ jɔʃi* || *ja ala* || “Awesome! Wow!” (IMs [6]–[7]). Needless to say, the Mongolian town Hatgal that

the speaker is happy to (re)discover cannot be identified as if included in the ORIGO of this discourse. Still, one might suggest that the discovery of the town takes place within the ORIGO setting, since the atlas makes part of the ORIGO, in which case it will be an expansion of the concept of ORIGO (but see further the remarks below). In any case, the utterances expressing excitement, which are identified as unipartite clauses, are directed not at the atlas but at the recollections of the town of Hatgal, thus transposing the location and the time to the time and space when the speaker was visiting Hatgal. This is, I believe, a clear case of extra-ORIGO anchoring.

Furthermore, it might be the case that all utterances containing place names or those including the presentational particle *hine* in this passage should be classified as having an extra-ORIGO anchoring, although the atlas page at which the speaker would be pointing is of course found within the location where the discourse takes place. The very beginning of this discourse includes the following remark about the atlas:

- (12) [1] *waj εε atlas* ||  
       wow what atlas  
       “Wow! What an atlas!”  
       [2] *χufiling ze* ||  
       fantastic it  
       “It is fantastic.”  
       [3] *a /*  
       isn’t\_it  
       “Isn’t it?”  
       (OCh\_sp1\_001–003)

The first clause expresses the speaker’s attitude towards the atlas by its exclamative modality. The referent (viz., the atlas) is anchored within the time and space of the discourse (thus classified as intra-ORIGO anchoring of this predicate). It is thus being introduced into the discourse. Mentioning the atlas seems to establish a new setting for those units of the discourse which will refer to the atlas and its maps, and thus all presentational clauses and other references to the atlas have their origin in a different setting, viz. the atlas map, representing another ORIGO.

#### 4.2 Unanchored

The final category of anchoring unipartite clauses presents clauses of which the predicates are not anchored in the discourse at all and have no obvious, direct anchors – either internal or external. The most conspicuous cases of unanchored clauses are those introducing a brand new topic – or referent – into the discourse via a presentational construction (cf., inter alia, Lambrecht 1994: § 4.4). One way

of introducing brand-new referents into the discourse in Hebrew is by using the so-called existential particle *jɛf*, as in (13).

- (13) [1] *tifmɛu davar* ||  
hear.IMP.PL thing  
“Listen to this:”
- [2] *jɛf makom* |  
EXT place  
“There is a (certain) place”
- [3] *bɛ=ɛχɔv* |  
in=street
- [4] *levinski* |  
Levinsky  
“in Levinsky Street”
- [5] *bɛ* |  
in
- [6] *tɛlaviv* |  
Tel-Aviv  
“in Tel-Aviv;”
- [7] *miʃei* |  
someone.F
- [8] *jɛ* |  
that
- [9] *ʔsa* |  
make.SGF
- [10] *tavlinim* |  
spices  
“(There is) someone (there) who makes spices,”
- [11] *jɛ* |  
that
- [12] *ʔkaχat* |  
concoct.SGF  
“who concocts –”
- [13] *lo ʔkaχat* ||  
NEG concoct.SGF  
“not concocts,”
- [14] *bɛɛtsɛm maʔkiva* ||  
in\_fact composes.SGF  
“mixes –”
- [15] *kol=minɛj* |  
all=sorts\_of

- [16] *e* |  
uh
- [17] *tamhilim* |  
mixtures
- [18] *ʃel=kəl=minej tavlinim bejaʁad* ||  
of=all=sorts\_of spices together  
“sorts of uh combinations of various kinds of spices together.”  
(Sh2c:2315”–2329.5”)

Following a discourse-regulative comment (IM [1]), the speaker introduces her new topic by an existential clause (IMs [2]–[6]). Against the consensual analysis of this type of clauses which regards the existential particle as a predicate (Glinert 1989: § 16.9; Schwarzwald 2001: 96) – it is the new referent introduced into the discourse that is to be regarded as a predicate. The existential particle *ʃel* serves in such constructions as a presentational particle very much like the presentational particle *hine* discussed above. The uses and functions (presentational, existential, locative, possessive, etc.) of the particle *ʃel* and related forms (notably its negative counterpart *en*) should be subject for further research. In any case, the type of presentational clauses like the one in (13), IMs [2]–[6], should be regarded as unipartite clauses, and if they introduce a brand new topic into the discourse, they should be classified as unanchored. In this excerpt, after the initial reference to “a place in Levinsky street in Tel-Aviv” is made, the speaker introduces another referent, this time not making use of the existential particle, perhaps because now the newly referential expression is anchored in the already presented location. The utterance in IMs [7]–[18] is an expanded unipartite clause, which includes two subordinate clauses which are unipartite clauses all the same (IMs [9]–[10]; [12]–[18]), each embedded by the element *ʃe* “that” (IM [8], IM [11]) with an inserted parenthesis (IMs [12]–[14]).

## 5. Conclusion

Spoken language shows a remarkably different structure from written language. The most prominent feature of spoken language is its prosodic structure, upon which segmentation of the spoken string is actualized. Adopting a framework of an integrative approach to the structure of spoken language that includes prosody, information structure, and syntax has resulted in our ability to account for what has been termed here UNIPARTITE CLAUSES, syntactic units consisting of only a predicate domain, i.e., a nuclear or an extended predicate. The term PREDICATE is preferred here over terms from other areas of investigation (e.g., ‘rheme’, ‘comment’, or the like), because I wish

to stay in the realm of syntax as my basic domain of reference to linguistic structure. By default, the predicate is viewed as the element carrying the informational load of the clause, the 'new' element in the discourse and the focused component of the clause. Essentially, the predicate carries the clause modality

A broad classification of unipartite clauses in Hebrew has also been presented. This classification must be viewed as preliminary and rather tentative, aimed first and foremost at establishing the criteria for determining unipartite clauses in spoken Israeli Hebrew. This classification was based on rather small collection of data, which now forms The Corpus of Spoken Israeli Hebrew (CoSIH). Further research, based on this corpus and on a larger collection of texts, will surely enhance our understanding of both the nature and the functions of unipartite clauses. It is my hope that research following the lines suggested here will be taken on other languages than Hebrew, Afro-asiatic and beyond. As already mentioned briefly, many other languages, spoken and written alike, attest similar structures in various degrees of frequency. Among Semitic languages, I should mention especially Akkadian (Cohen 2005: § 1), as well as written Israeli Hebrew (Rubinstein 1968: Chapter 6; Sadka 1991). Dickins (2010: § 3) deals with similar structures in Sudanese Arabic, although for him unipartite clauses (in Dickins' terminology: 'monopartite') also include verbal forms, which are by definition bipartite, as they include both the subject and the predicate of a clause (Goldenberg 1998; Izre'el 2012: § 4). Both Cohen and Dickins treat such units as clauses. So does Rubinstein. Sadka vacillates between 'utterance' and 'sentence'.

I end my discussion with two examples from Biblical Hebrew, where the notion of UNIPARTITE CLAUSE might raise a few eyebrows, yet it forms part of the Biblical Hebrew structure as well. (14) exhibits a response to a question in the form of an independent unipartite clause; (15a) shows a subordinate unipartite clause, whose structure can be established by comparing the bipartite clause in (15b).

- (14) *wayyomru 'elw 'ayye šwr 'išteḵ*  
 "They said to him: 'Where is your wife Sarah?'"  
*wayyomer hinne ḥ'ohel*  
 and\_he\_said PRES in\_the\_tent  
 "And he said: 'In the tent.'"  
 (Genesis 18:9)

- (15a) *wayyar(') 'elohim 'et ḥ'or ki ṭob*  
 and\_he\_saw God ACC the\_light that good  
 "And God saw that the light was good (lit: the light that (it) is good)."  
 (Genesis 1:4)

- (15b) *wayyar(') š'ul 'ašer hu(') maškil m'od*  
 and\_he\_saw Saul that he successful very  
 "And Saul saw that he was very successful."  
 (Genesis 1:4)

In this spirit, let me express my wish as follows:

wayyar(ʿ)   habbalšən   ʿet   hammišpəṭ   haḥad-ʿeḇori   ki   toḇ  
 and\_he\_saw the\_linguist ACC the\_clause the\_unipartite that good  
 “And the Linguist saw that the unipartite sentence is good.”

## References

- Abraham, Werner. 2011. “Traces for Bühler’s Semiotic Legacy in Modern Linguistics”. *Theory of Language: The Representational Function of Language*, ed. by Karl Bühler, xiii–xliii. Translated by Donald Fraser Goodwin. In collaboration with Achim Eschbach. Amsterdam: John Benjamins.
- Benayoun, Jean-Michel. 2003. “Sujet zéro, pacte référentiel et thème”. *Le sujet: Actes – augmentés de quelques articles – du colloque Le sujet organisé à l’Université de Provence, les 27 et 28 septembre 2001, avec le concours du CELA (EA 851) de l’UFR LAG-LEA et de l’Université de Provence* (Bibliothèque de Faits de Langues), ed. by Jean-Marie Merle, 173–182. Paris: Ophrys.
- Bally, Charles. 1965. *Linguistique générale et linguistique française*. Quatrième édition revue et corrigée. Berne: Éditions Francke.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finnegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Blanche-Benveniste, Claire. 2006. “Les énoncés sans verbe en français parlé”. Written version of a lecture given at Naples, 23 February 2006.
- Borochofsky Bar-Aba, Esther. 2010. *Issues in Colloquial Hebrew*. Jerusalem: The Bialik Institute.
- Bühler, Karl. 1934. *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Ullstein Taschenbuch, 1159. Stuttgart: Gustav Fischer (1982).
- Butler, Christopher S. 2003. *Structure and Function: A Guide to Three Major Structural-Functional Theories* (Studies in Language Companion Series, 63–64). Amsterdam: John Benjamins.
- Carter, Ronald, & Michael McCarthy. 2006. *Cambridge Grammar of English: A Comprehensive Guide; Spoken and Written English Grammar and Usage*. Cambridge: Cambridge University Press.
- Chafe, Wallace. 1994. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: The University of Chicago Press.
- Cohen, Eran. 2005. “Addenda to Non-Verbal Clauses in Old Babylonian”. *Journal of Semitic Studies* 50.247–279. doi:10.1093/jss/fgi037
- Cresti, Emanuela. 2005. “Notes on Lexical Strategy, Structural Strategies and Surface Clause Indexes in the C-ORAL-ROM Spoken Corpora”. *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages* (Studies in Corpus Linguistics 15), ed. by Emanuela Cresti & Massimo Moneglia. Amsterdam: John Benjamins. 209–256 (Chapter 6). doi:10.1075/scl.15.o8cre
- Cresti, Emanuela, & Massimo Moneglia (eds.). 2005. *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages* (Studies in Corpus Linguistics 15). Amsterdam: John Benjamins. doi:10.1075/scl.15



- Daniel, Michael & Andrew Spencer. 2009. "The Vocative – an Outlier Case". *The Oxford Handbook of Case*, ed. by Andrej Malchukov & Andrew Spencer, 626–634. Oxford: Oxford University Press.
- Dickins, James. 2010. "Basic Sentence Structure in Sudanese Arabic". *Journal of Semitic Studies* 55.237–261. doi:10.1093/jss/fgq046
- Foley, William A. 2006. "Topic Chaining Constructions". *Encyclopedia of Language and Linguistics*, 2nd edn. ed. by Keith Brown, 773–777. Oxford: Elsevier. doi:10.1016/B0-08-044854-2/02019-8
- Givón, Talmy (ed.). 1983. *Topic Continuity in Discourse: A Quantitative Cross-Language Study* (Typological Studies in Language 3). Amsterdam: John Benjamins. doi:10.1075/tsl.3
- Givón, Talmy. 1992. "The Grammar of Referential Coherence as Mental Processing Instructions". *Linguistics* 30.5–55. doi:10.1515/ling.1992.30.1.5
- Glinert, Louis. 1989. *The Grammar of Modern Hebrew*. Cambridge: Cambridge University Press.
- Goldenberg, Gideon. 1998. "On Verbal Structure and the Hebrew Verb". *Studies in Semitic Linguistics: Selected Writings*, 138–147. Jerusalem: Magnes.
- Hall, Alison. 2009. "Subsentential Utterances, Ellipsis, and Pragmatic Enrichment". *Pragmatics & Cognition* 17.222–250. doi:10.1075/pc.17.2.02hal
- Halliday, Michael A. K. 2004. *An Introduction to Functional Grammar*. 3rd edition revised by Christian M. I. M. Matthiessen. London: Arnold.
- Harnish, Robert M. 2009. "The Problem of Fragments: Two Interpretative Strategies". *Pragmatics & Cognition* 17.251–282. doi:10.1075/pc.17.2.03har
- Izre'el, Shlomo. 2005. "Intonation Units and the Structure of Spontaneous Spoken Language: A View from Hebrew". *Proceedings of the IDP05 International Symposium on Discourse-Prosody Interfaces*, ed. by Cyril Auran, Roxanne Bertrand, Catherine Chanet, Annie colas, Albert Di Cristo, Cristel Portes, Alain Reynier & Monique Vion. [http://www.tau.ac.il/~izreel/publications/IntonationUnits\\_IDP05.pdf](http://www.tau.ac.il/~izreel/publications/IntonationUnits_IDP05.pdf)
- Izre'el, Shlomo. 2012. "Basic Sentence Structures: A View from Spoken Israeli Hebrew". *Penser les langues avec Claire Blanche-Benveniste* ed. by Sandrine Caddéo, Marie-Noëlle Roubaud, Magali Rouquier & Frédéric Sabio, 215–227. Aix-en-Provence: Presses Universitaires de Provence. Corrected version can be downloaded at [http://www.tau.ac.il/~izreel/publications/SentenceStructure\\_BlancheMem.pdf](http://www.tau.ac.il/~izreel/publications/SentenceStructure_BlancheMem.pdf)
- Izre'el, Shlomo. Forthcoming. "The Basic Unit of spoken Language and the Interface Between Prosody, Discourse and Syntax: A View from spontaneous Spoken Hebrew."
- Izre'el, Shlomo & Amina Mettouchi. 2015. "Representation of Speech in CorpAfroAs: Transcriptional Strategies and Prosodic Units". *Corpus-based Studies of Lesser-described Languages: The CorpAfroAs Corpus of Spoken AfroAsiatic Languages*, ed. by Amina Mettouchi, Martine Vanhove & Dominique Caubet, 13–41. Amsterdam: John Benjamins.
- Izre'el, Shlomo, & Vered Silber-Varod. 2009. "OMER LENATEAX LENATEAX: Perception of Prosodic Groups in Spoken Hebrew". *Hebrew Linguistics* 63–64.13–33.
- Jespersen, Otto. 1924. *The Philosophy of Grammar*. London: George Allen & Unwin (1951).
- Johansson, Marjut, & Eija Suomela-Sahni. 2011. "Énonciation: French Pragmatic Approache(s)". *Discursive Pragmatics: Handbook of Pragmatics Highlights*, vol. 8 ed. by Jan Zienkowski, Jan-Ola Östman & Jef Verschueren, 71–101. Amsterdam: John Benjamins. doi:10.1075/hoph.8.05mar
- Kibrik, Andrej A. 2011. "Reference in Discourse". *Oxford Studies in Typology and Linguistic Theory*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199215805.001.0001
- Lambrecht, Knut. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representation of Discourse Referents* (Cambridge Studies in Linguistics 71). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511620607

- Lefeuve, Florence. 1999. "La phrase averbale en français". *Collection Langue & Parole, Recherches en Sciences du Langage*. Paris: L'Harmattan.
- Le Goffic, Pierre. 1993. *Grammaire de la Phrase Française*. Collection Hachette Université: Langue française. Paris: Hachette.
- Martin, Philippe. 2009. *Intonation du français*. Collection U, Linguistique. Paris: Armand Colin.
- Mettouchi, Amina, Anne Lacheret-Dujour, Vered Silber-Varod & Shlomo Izre'el. 2007. "Only Prosody? Perception of Speech Segmentation". *Nouveaux cahiers de linguistique française 28 : Interfaces discours – prosodie: actes du 2ème Symposium international & Colloque Charles Bally*, 207–218.
- Nariyama, Shigeko. 2007. "Ellipsis and Markedness: Examining the Meaning of Ellipsis". *6th International Discourse Anaphora and Anaphor Resolution Colloquia 2007 Proceedings*, 97–102. Porto: Centro de Linguistica da Universidade do Porto.
- Palmer, F. R. 2001. *Mood and Modality* (Cambridge Textbooks in Linguistics), 2nd edn. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139167178
- Rubinstein, Eliezer. 1968. *The Nominal Sentence: A Study in the Syntax of Contemporary Hebrew*. Tel-Aviv: Hakibbutz Hameuchad.
- Sadka, Isaac. 1991. "The Unipartite Utterance". *Shay le-Chaim Rabin: Studies on Hebrew and Other Semitic Languages Presented to Professor Chaim Rabin on the Occasion of His 75th Birthday*. Jerusalem: Akademon. 295–310. (= Isaac Sadka. 1997. *Studies in Hebrew Syntax and Semantics*, ed. by Moshe Goshen-Gottstein, Shlomo Morag & Simha Kogut, 1012–115. Beer-Sheva: Ben-Gurion University Press.)
- Schwarzwald, Ora R. 2001. *Modern Hebrew* (Languages of the World/Materials 127). München: LINCOM Europa.
- Segel, Esben. 2008. "Re-evaluating Zero: When Nothing Makes Sense". *SKASE Journal of Theoretical Linguistics* 5:2.1–20.
- Sinclair, John. 2002. Review of Biber et al. 1999. *International Journal of Corpus Linguistics* 6.339–359. doi:10.1075/ijcl.6.2.09sin
- Tao, Hongyin. 1996. *Units in Mandarin Conversation: Prosody, Discourse, and Grammar* (Studies in Discourse and Grammar 5). Amsterdam: John Benjamins. doi:10.1075/sidag.5
- Tesnière, Lucien. 1966. *Éléments de syntaxe structurale, Deuxième édition revue et corrigée*. Paris: Klincksiek.
- Vion, Robert. 2001. "Modalités, modalisations et activités langagières". *Marges Linguistiques* 2.209–231.
- Winkler, S. 2006. "Ellipsis". *Encyclopedia of Language and Linguistics*, ed. by Keith Brown, 2nd edn. Oxford: Elsevier. 109–113. doi:10.1016/B0-08-044854-2/02003-4
- Ziv, Yael, & Barbara Grosz. 1994. "Right Dislocation and Attentional State". *The Israel Association for Theoretical Linguistics: Proceedings of the 9th Annual Conference and Workshop on Discourse*, ed. by Anita Mittwoch and Rhonna Buchalla, 184–199. Jerusalem: Akademon.

**PAGE PROOFS**

© JOHN BENJAMINS PUBLISHING COMPANY

**2nd proofs**

# The Interaction of state, prosody and linear order in Kabyle (Berber)

## Grammatical relations and information structure

Amina Mettouchi

Ecole Pratique des Hautes Etudes, Paris, and CNRS-LLACAN

The aim of this paper is to show how, starting only from forms belonging to various domains (morphology, syntax, and prosody), it is possible to compute the grammatical relations and the information structure constructions in Kabyle spontaneous speech.

The study is based on recordings made in the field, transcribed, translated, and annotated with Praat and Elan-CorpA. The methodology consists of systematically retrieving the sequences containing a verb and looking for the presence of a noun (and its inflection) within the prosodic group of the utterance, or outside, as well as studying the linear orders involved. This non-aprioristic methodology reveals the close interaction between grammatical relations and information structure in Kabyle. The study provides evidence to support the claim that the encoding of grammatical relations on nouns is a by-product of information structure constraints in Kabyle.

**Keywords:** prosody, information structure, grammatical relations, Kabyle, Berber

### 1. Introduction

The aim of this paper<sup>1</sup> is to show how, starting only from forms belonging to various domains (morphology, syntax, and prosody), it is possible to compute grammatical relations and information structure constructions in Kabyle spontaneous speech.

The study is based on data recorded in the field, transcribed, translated, and annotated with Praat<sup>2</sup> and Elan-CorpA.<sup>3</sup>

---

1. The hypotheses in the present paper were first presented in June 2011 at the 14th International AfroAsiatic Linguistics Conference in Turin (Italy).

2. <http://www.fon.hum.uva.nl/praat/>

3. [http://llacan.vjf.cnrs.fr/res\\_ELAN-CorpA.php](http://llacan.vjf.cnrs.fr/res_ELAN-CorpA.php)

The methodology consists of systematically retrieving the sequences containing a verb, and looking for the presence of a noun (and its inflection) within the prosodic group of the utterance, or outside, as well as studying the linear orders involved.

From such forms, it is possible to ask fundamental questions about the nature of grammatical relations and the structures associated with various information structure values in Kabyle.

The hypothesis underlying the paper is that the marking of grammatical relations on nouns is a byproduct of the information structure constraints described in §2.

### 1.1 General information about Kabyle

Berber languages are spoken in Northern Africa, in a zone delimited by the Atlantic Ocean to the west, the Mediterranean to the north, the oasis of Siwa (Egypt) to the east, and the southern borders of Mali and Niger to the south. Those languages constitute a family within the Afroasiatic phylum. Well-known members of the family are Kabyle (spoken in Northern Algeria), Tashelhiyt (Shilha) (spoken in Southern Morocco), and Tamashek and Tahaggart (also called Tuareg), spoken in Southern Sahara. Kabyle is spoken by about four million people in the north of Algeria. The variety investigated in this paper is a western one, spoken in the village of Ait Ikhlef, close to the town of Bouzeguene. I have collected all the data on fieldwork between 2007 and 2011.<sup>4</sup>

In Kabyle, as in all Berber languages, a minimal predication consists of a verb and its bound personal pronoun, or a non-verbal predicate. In this paper I focus on verbal predicates. In addition to this core, the clause may contain noun phrases and prepositional phrases, as well as adverbs. Within noun phrases, modifiers follow the modified constituent. The language has two genders and two numbers, marked on pronominal affixes and clitics to verbs, nouns, and prepositions, as well as on adjectives (a subclass of nouns) and on nouns. It also has two states, marked on nouns.

### 1.2 Relevant coding means

This section describes the formal means that come into play for the encoding of grammatical relations and information structure.<sup>5</sup>

4. My warmest thanks go to the speakers I recorded in the village over the years.

5. Examples are transcribed morphophonologically according to the following rules: IPA symbols are used whenever they consist of a single character. Otherwise, they are replaced by

### 1.2.1 The state distinction

Nouns have two forms, the absolute (ABS, traditionally called ‘état libre’ (‘free state’)), and the annexed (ANN, traditionally called ‘état d’annexion’ (‘annexation state’)). E.g., man = *argaz* in the absolute, *wrgaz* in the annexed; girl = *taqfiɬt* in the absolute, *tqfiɬt* in the annexed.

Table 1. Illustration of state alternation in Kabyle<sup>6</sup>

	Masculine SG	Masculine PL	Feminine SG	Feminine PL
Annexed	<i>w-rgaz</i> “man”	<i>j-rgaz-n</i> “men”	<i>t-qfiɬ-t</i> “girl”	<i>t-qfiɬ-in</i> “girls”
Absolute	<i>a-rgaz</i> “man”	<i>i-rgaz-n</i> “men”	<i>t-a-qfiɬ-t</i> “girl”	<i>t-i-qfiɬ-in</i> “girls”

The state distinction in itself does not code grammatical roles (Galand 1964; Chaker 1988). Evidence for this fact is developed in Mettouchi & Frajzyngier (2013), where the states are thus defined: the annexed state indicates that the noun provides the value (in the logical sense) for the variable of the function grammaticalized in a preceding constituent (Mettouchi & Frajzyngier 2013). Such functions are diverse. The absolute state is simply the default form of the noun, and as such it has no overall function of its own (Mettouchi & Frajzyngier 2013). Within various structures, the states contribute to the creation of grammatical meaning, thus being the backbone of the grammar of Kabyle.

### 1.2.2 Prosody

An intonation unit is a segment of speech that has a coherent intonation contour (Chafe 1994), and is delimited by its boundaries (Cruttenden 1997), which bear a ‘boundary tone’ (Pierrehumbert & Hirschberg 1990). In Kabyle, Intonation Units are marked by one or more of the following cues:

conventional characters of the Berberologist tradition: pharyngealized/emphatic consonants are written with a subscribed dot (ɖ, ɣ, ʂ, ʈ, ʒ), affricates are written č, ž, ġ, and ʈ (for ʈʃ, ɖʒ, ɖʒ, and ts).

The following abbreviations are used: ABS absolute state; ABSV absolutive pronominal paradigm; ANN annexed state; AOR aorist; ASSOC associative; CAUS causative; CNS shared reference demonstrative; COM comitative; COP copula; DAT dative; EXNEG existential negative; F feminine; GEN genitive; HESIT hesitation; IDP independent pronoun; IPFV imperative; KIN kinship pronominal paradigm; M masculine; NEG negation; PFV perfective; PL plural; POS positive; POSS possessive pronominal paradigm; PREP prepositional pronominal paradigm; PROX proximal; SBJ subject pronominal paradigm; REAL realis; REL relator; RELSBJ subject relativization circumfix; SG singular; VOC vocative.

6. In this table, only underlying forms are given, so that the morphology of the state alternation is clearer. Syllabification rules which result in schwa insertion are reflected in the examples throughout the paper.

Main external cues:

1. final lengthening;
2. initial rush;
3. pitch reset;
4. pause;
5. creaky voice.

Main internal cues:

1. declination;
2. tonal parallelism, or isotony.

On the basis of those cues (see also Izre'el & Mettouchi 2015), a one-hour corpus composed of 20 minutes of dialogue and 40 minutes of monologue (narrative) was segmented into intonation units, regardless of any other cue, syntactic, semantic, or other.

### 1.2.3 *Linear orders*

Linear ordering concerns not only noun phrases and verbs, as it also comes into play for other units, for instance clauses. In this paper, I consider linear ordering as a series of formal means, crucially depending on the existence of a reference point (Frajzyngier & Shay 2003: 60–62), which is overtly and unambiguously coded for this function. This reference point differs depending on the language.

In Berber, the verb is a salient potential reference point, as it is formally always affixed by a dedicated pronominal paradigm. It cannot therefore be confused with any other element of the clause.

Another reference point that I consider here is the prosodic boundary between intonation units (see § 1.2.2). As a discontinuity in the flow of speech, it constitutes a salient potential reference point, regardless of its functional values, which depend on the boundary tone, and other prosodic cues.

In this paper, I focus on the ordering of noun phrases with respect to the verb: before or after the verb; as well as the relative order of noun phrases when this is formally computable. I also take into account the position relative to the prosodic boundary: before or after the prosodic boundary.

### 1.2.4 *Traditional analysis of the structure of the clause in Berber*

Since (Galand 1964), the bound pronoun affixed to the verb has been considered as the actual subject of the Berber clause, coreferent noun phrases being expansions of this pronoun.

- (1) *j-ħma*  
 3SG.M-be\_warm:PFV  
 “It is/was warm.”

The lexical item coreferent with the subject pronoun is either (Galand 1964) the ‘indicateur de thème’ (‘theme indicator’) (i.e., the noun is in the absolute state and appears before the verb); or the ‘complément explicatif’ (‘explanatory complement’) (i.e., the noun is in the annexed state and appears after the verb).

- (2) *aḍazin j-ħma*  
 tagine:ABS.M.SG 3SG.M-be\_warm:PFV  
 “The tagine (cooking pot) is hot”
- (3) *j-ħma uḍazin*  
 3SG.M-be\_warm:PFV tagine:ANN.M.SG  
 “The tagine is hot”

The term ‘indicateur de thème’ is not limited to the subject, as the noun preceding the verb can also be coreferent with pronouns other than the subject:

- (4) *aksum j-čča=t*  
 meat:ABS.M.SG SBJ3SG.M-eat:PFV=ABSV3SG.M  
 “He ate the meat” (‘object’)
- (5) *argaz-nni t-mmut tmṭṭut-is*  
 man:ABS.M.SG-CNS SBJ3SG.F-die:PFV woman:ANN.F.SG-POSS3SG  
 “That man, his wife died” (‘possessor’)

The ‘complément explicatif’ can also be coreferential with pronouns other than the subject in Kabyle:

- (6) *j-čča=t wəksum-nni* (‘object’)  
 SBJ3SG.M-eat:PFV=ABSV3SG.M meat:ANN.M.SG  
 “He ate the meat”

And the coreferent pronoun can be affixed to a noun, and not only to the verb:

- (7) *t-mmut tmṭṭut-is wərgaz-nni*  
 SBJ3SG.F-die:PFV woman:ANN.F.SG-POSS3SG man:ANN.M.SG-CNS  
 “His wife died, that man.” (‘possessor’)

In this traditional analysis, only two structures are taken into account:

‘Indicateur de thème’ – V  
 V – ‘Complément explicatif’



Those structures have essentially been used to argue against the attribution of the subject role to noun phrases (Galand 1964). Some publications have gone further in their endeavor to study the information structure of the constructions actually encountered in Kabyle spontaneous speech (Mettouchi 2007a, 2007b, 2008; Kuningas & Leino 2006), but no systematic study of all possible combinations has been undertaken to date.

### 1.2.5 *Constituent order analysis*

The present paper adopts a non-aprioristic methodology: I only take into account the formal means involved (state, prosodic boundary, linear order), without assumptions concerning the function of each noun before the structures are examined. Consequently, a wide array of constructions are actually investigated. The potential structures<sup>7</sup> are listed below.<sup>8</sup> Note that the abbreviation  $V_{\text{sbj}}$  represents the minimal predication composed of a verb and its obligatory personal affix (and possibly other bound pronouns).


$[V_{\text{sbj}}]$   
 $[V_{\text{sbj}} \text{ NP}_{\text{abs}}]$   
 $[V_{\text{sbj}} \text{ NP}_{\text{ann}}]$   
 $[V_{\text{sbj}} \text{ NP}_{\text{ann}} \text{ NP}_{\text{abs}}]$   
 $[V_{\text{sbj}} \text{ NP}_{\text{abs}} \text{ NP}_{\text{ann}}]$   
 $[\text{NP}_{\text{abs}} V_{\text{sbj}}]$   
 $[\text{NP}_{\text{abs}} V_{\text{sbj}} \text{ NP}_{\text{abs}}]$   
 $[\text{NP}_{\text{abs}} V_{\text{sbj}} \text{ NP}_{\text{ann}}]$   
 $\text{NP}_{\text{abs}} [V_{\text{sbj}} (\text{NP}) (\text{NP})]$   
 $[V_{\text{sbj}} (\text{NP}) (\text{NP})] \text{ NP}_{\text{ann}}$

When the NP does not bear the indices “ann” or “abs” in the list above, it means that both states are possible in this position. The state opposition is an independent coding means whose function is not to mark grammatical relations or information structure (cf. § 0.2.1). Indication of the state borne by the nouns, however, is relevant for distinguishing among some structures: for instance between  $[V_{\text{sbj}} \text{ NP}_{\text{abs}}]$  and  $[V_{\text{sbj}} \text{ NP}_{\text{ann}}]$ , which do not have the same information structure value.

Noun phrases following prepositions have not been taken into account here because they are unambiguous as far as function is concerned: indirectly affected

7. The square brackets [ and ] are used to represent the presence of prosodic boundaries, which are transcribed / or // in actual examples taken from my recording. 

8. Prepositional phrases have not been included, but another study in progress, on linear ordering of PPs, shows that their presence has no effect on the information structure values of the combinations listed in this section.

argument is introduced by preposition *i*, and instrumental complement is introduced by preposition *s*, *while locative complements are introduced by prepositions g, sag, yər, among other.* 

## 2. Information structure

In investigating the functions of those structures, five constructions were found, each consisting of one or more structures and having one function:

- $[V_{\text{sbj}} (N_{\text{abs}})]$ , which subsumes the following:  
 $[V_{\text{sbj}}]$   
 $[V_{\text{sbj}} N_{\text{abs}}]$
- $[V_{\text{sbj}} N_{\text{ann}} (N)]$ , which subsumes the following:  
 $[V_{\text{sbj}} N_{\text{ann}}]$   
 $[V_{\text{sbj}} N_{\text{ann}} N_{\text{abs}}]$   
 $[V_{\text{sbj}} N_{\text{abs}} N_{\text{ann}}]$
- $[N V_{\text{sbj}} (N)]$  which subsumes the following  
 $[N_{\text{abs}} V_{\text{sbj}}]$   
 $[N_{\text{abs}} V_{\text{sbj}} N_{\text{abs}}]$   
 $[N_{\text{abs}} V_{\text{sbj}} N_{\text{ann}}]$
- $N_{\text{abs}} [V_{\text{sbj}} (N) (N)]$
- $[V_{\text{sbj}} (N) (N)] N_{\text{ann}}$

The exact functions of those constructions will be investigated one by one, in main and independent clauses.

### 2.1 Function of $[V_{\text{sbj}} (N_{\text{ABS}})]$

This construction is the default one in terms of syntax (it is in itself a full-formed clause) and information structure. It has no marked value but builds on previous context without any shift or change of perspective. All examples involving only a verb and its bound subject pronoun, possibly followed by a noun in the absolute within the same intonation unit, were encountered inside a subtopic in a narrative or a conversation.

Subtopics are discourse-level topics that rank lower than a basic-level topic but higher than a sentence topic. A topic is “an aggregate of coherently related events, states and referents that are held together in some form in the speaker’s semi-active consciousness” (Chafe 1994: 121). Chafe implies that a subtopic should encompass more than an intonation unit, since an intonation unit is generally associated with a “focus of consciousness” related to “active” information (1994: 29).

- (8) *i-ruḥ*                      *ar wədrar* //                      *i-ruḥ*                      *ar*  
 SBJ3SG.M-go:PFV to mountain:ANN.SG.M // SBJ3SG.M-go:PFV to  
*wədrar*                      *a*    *Amina* / *i-qqaz*                      *i-qqaz*  
 mountain:ANN.SG.M VOC amina / SBJ3SG.M-dig:IPFV SBJ3SG.M-dig:IPFV  
*i-qqaz*                      *i-qqaz*                      *i-qqaz*  
 SBJ3SG.M-dig:IPFV SBJ3SG.M-dig:IPFV SBJ3SG.M-dig:IPFV  
*i-qqaz* /                      *i-xdəm*                      *lbir* /    *annstilat* //  
 SBJ3SG.M-dig:IPFV / SBJ3SG.M-make:PFV well:ABS / enormous //  
 “He went to the mountain, he went to the mountain, Amina, he dug and dug,  
 he made a well, an enormous one.”

In this discursive context, using the verb (with its obligatory subject pronoun), possibly followed by a noun in the absolute – which is never coindexed with a preceding pronoun – (such as *lbir*, “well”, the object of *i-xdəm*, “he made”) is the unmarked informational function, namely (sub-) topic continuity: the protagonist is the same, and the narrative is carried forward. This function is very well described in the general literature across languages (Givón 1983; Chafe 1994; Lambrecht 1994); for Kabyle it has been analyzed in Mettouchi 2008. Even if there are several protagonists, only bound pronouns are used (cf. Mettouchi 2005, 2007b).

Once noun phrases other than the direct object – i.e., noun phrases coreferent to a bound pronoun – appear, the information structure value is changed. This shows that the mere presence of a noun phrase coreferent to a bound pronoun, in a pronominal-argument language such as Kabyle, has information structure value. Which value this is depends on the position of that noun with respect to the verb and the prosodic boundary, as well as on the state of the noun (absolute or annexed).

## 2.2 Function of [ $V_{\text{subj}}$ NP<sub>ann</sub> (NP<sub>abs</sub>)]

This construction is realized as three different structures, depending on the presence of two NPs or one. The important factor is the presence of a noun in the annexed state in the position after the verb (possibly separated from it by another noun, in the absolute), within the same intonation unit as the verb.

- [ $V_{\text{subj}}$  NP<sub>ann</sub>]
- [ $V_{\text{subj}}$  NP<sub>ann</sub> NP<sub>abs</sub>]
- [ $V_{\text{subj}}$  NP<sub>abs</sub> NP<sub>ann</sub>]

The three structures promote an event or a state to topic status: both the main participant and the predicate are expressed, and even if the participant is known or

mentioned, the relationship it holds with the predicate is presented as new and is about to be developed in the following intonation units. This structure was analyzed as ‘sentence-focus’ or ‘thetic’ in Mettouchi 2008, because newness was not limited to the referent, but encompassed the state of affairs itself, represented by the association of a verb (and its bound subject pronoun) and a noun in the annexed state. That analysis is true at the level of the sentence, but it doesn’t provide information about the role of the construction in discourse.

An examination of my corpus show that in discourse the structure is used to introduce a new episode in a narrative or a new subtopic in a conversation. Sometimes, as in the following example, it is mentioned as a disclosure:

- (9) sp3:  
*anda lla-nt tlata tqjifin //*  
 where be:PFV-SBJ3PL.F three girls:ANN //  
 “Where are those three girls?”  
 sp2:  
*jah / tɔ-mmud lwiza //*  
 in fact SBJ3SG.F-die:PFV Louisa //  
 “actually, Louisa died.”

The preceding discourse was structured around the subtopic of a woman’s number of children. When it was clear that the two speakers hadn’t come up with the same number of children, Speaker 2 added the crucial information that a daughter had died. Since the statement is sad, it is soon replaced by another a piece of information: the names of the other two daughters, presented with the same structure:

- (10) sp2:  
*tɔ-lla nadija / faɖila ɛ::: /*  
 SBJ3SG.F-be:PFV Nadia / Fadila HESIT //  
 “There is Nadia, Fadila...”  
 sp1:  
*faɖila d nadija d tinəggura //*  
 Fadila ASSOC Nadia COP last:ABS.PL.F  
 “Fadila and Nadia are the youngest ones.”

This structure is used typically in *wh*- questions with presupposed information, as in Speaker 3’s question above, or for new information (regardless of the activation state of the referent itself).

Sometimes, a noun in the absolute also appears in the structure: either after the noun in the annexed state ([V<sub>sbj</sub> N<sub>ann</sub> N<sub>abs</sub>]) or before ([V<sub>sbj</sub> N<sub>abs</sub> N<sub>ann</sub>])

In the following example, the father, prompted by his sons, is sending his younger son to steal a beautiful carpet from an ogress:

- (11) *jə-nna=jas*                      *zra-n*                      *wajtma-s*  
 SBJ3SG.M-say:PFV=DAT3SG see:PFV=SBJ3PL.M brother:ANN.PL-KIN3.SG  
*tazarbit*                      *ar jəmma*                      *Nuʒa /*  
 carpet:ABS.SG.F to mother:ANN.F.SG Nuʒa  
 “The father said his brothers had seen a carpet at Mother Nuja’s.”

This new piece of information is the basis of his demand to his son, that he should go and steal it from the ogress.

In the following example, taken from the same tale, the brothers come to the father with the news that the ogress has a hen, whose eggs heal all sorts of illness.

- (12) *t-sʃa*                      *tajaziʃ*                      *jəmma*                      *Nuʒa /*  
 SBJ3SG.F-possess:PFV hen:ABS.SG.F mother:ANN.F.SG Nuʒa /  
 “Mother Nuja has a hen.”

All three structures belong to the same construction, as they all have the same informational value; the difference is first between intransitive and transitive predication, and second, among the latter, between the default order, which is  $[V_{\text{sbj}} N_{\text{ann}} N_{\text{abs}}]$  if both nouns have equal weight but which becomes  $[V_{\text{sbj}} N_{\text{abs}} N_{\text{ann}}]$  if the noun in the annexed state is heavier (in terms of information status or length).

Note that, apart from construction 1.1, this construction is the only one that can be found in a dependent clause (in particular in relative clauses, but also in complement clauses). This is in keeping with the fact that the construction globally construes the event or situation, it does not comment on a topic, or reactivate a referent.

### 2.3 Function of $[N_{\text{ABS}} V_{\text{sbj}} (N)]$

Within the same prosodic unit, a noun can appear before the verb. This noun is always in the absolute state (the annexed state is only used after a grammatical morpheme encoding a function, for which the noun in the annexed state is the variable (Mettouchi & Frajzyngier 2013).

This construction, which has not been described for Kabyle yet, is used as a background for further developments, when a salient preceding situation is recapitulated, so that the listener grasps the whole situation and its importance for the current discourse.

In the following example, all aspects of the situation have already been introduced, mostly through  $[V_{\text{sbj}} N_{\text{ann}} (N_{\text{abs}})]$  structures. Most referents have been previously mentioned, as is shown by the suffix *-nni*, which marks shared reference (Mettouchi 2006, 2011: 482) (often through previous mention).

- (13) *taq3unt-nni t-ssəglaf /*  
 dog:ABS.SG.F-CNS SBJ3SG.F-bark:CAUS.IPFV /  
 “The dog was barking,  
*azdduz-nni jə-ṭṭawi=t ubəḥri /*  
 big\_stick:ABS.SG.M-CNS SBJ3SG.M-bring:IPFV=ABSV3SG.M wind:ANN.SG.M /  
 the wind was moving the stick,  
*nutənti zddəm-nt /*  
 IDP3PL.F gather\_wood:IPFV-SBJ3PL.F /  
 the girls were gathering wood.”

In the first line, the noun in the absolute is the subject, as in the third line, but in the second line, it is an object and is taken up by the absolutive pronoun = *t*. What is important here is that one argument appears before the verb but not separated from it by a prosodic boundary. The construction allows the reduction of a long and complicated story into its salient characteristics.

In conversation, similar recapitulations occur, as in the following example, where the speaker takes up information scattered in the preceding context – where various brothers of her grandfather’s, as well as the grandfather himself, were said to have married a number of women – and then goes on to comment on the genealogy of the family:

- (14) *zəddi j-uy səṭṭi /*  
 grandfather:ABS.SG.M SBJ3SG.M-take:PFV grandmother:ABS.SG.F /  
 “My grandfather married my grandmother.”

This construction must not be confused with the following one, where the noun in the absolute preceding the verb is also before the prosodic boundary that precedes the verb.

## 2.4 Function of $N_{\text{ABS}} [V_{\text{sbj}} (N) (N)]$

This construction is characterized not by the internal structure of the intonation unit but by constituent ordering with respect to the verb (possibly preceded by a particle or auxiliary, as in the following example with negation),<sup>9</sup> and prosodic boundary. The argument preceding the prosodic boundary is taken up by a bound pronoun in the clause. Such constructions always imply a shift in perspective or

9. Note that preverbal particles (modal, aspectual, or negative) have fixed positions, and nothing else than a string of clitics can separate the particle from the verb. This makes the particle an alternative reference point to the verb for the calculation of linear ordering.

contrast with previous expectations, as in the following example. They are not topic-promoting devices, where a referent is promoted from non-active state to active state as argued in Kuningas & Leino (2006); indeed it is another structure which has this function in Kabyle:  $[V_{\text{subj}} (N) (N)] N_{\text{ANN}}$ , treated in § 1.5.

In the following example, what is important is the fact that the presupposition concerning the stepmother's relationship to her husband's daughters (built throughout the previous episode: she promised she would take care of them and love them dearly) is rejected.

- (15) *aj argaz tura jəssi-k-agi /*  
 VOC man:ABS.SG.M now daughter:ABS.PL-KIN2SG.M-PROX /  
 "My husband, now those daughters of yours,  
*ur zəddy-əy ara jid-sənt //*  
 NEG dwell:IPFV-SBJ1SG POSTNEG COM-PREP3PL.F //  
 I'm not living with them!"

Similarly, in the next example, the father had given his seven wives apples. They all gave birth to a normal boy, except the last one, who had eaten only half an apple:

- (16) *tin iwumi jə-fka akka nnəfš*  
 the\_one:SG.F to\_whom SBJ3SG.M-give:PFV thus half  
 "The one to whom he had given half  
*n tətəffaht / t-urw=ədd ləabd /*  
 GEN apple:ANN.SG.F / SBJ3SG.F-give\_birth:PFV=PROX human\_being:ABS.SG /  
 an apple, she gave birth to a human being,  
*qqar-n=as / ašmar nnəfš //*  
 say:IPFV-SBJ3PL.M=DAT3SG / Ašmar half //  
 whom people called Amar the Midget,  
*t-urw=it=id d nnəfš //*  
 SBJ3SG.F-give\_birth:PFV=ABSV3SG.M=PROX COP half //  
 it was half a child she had given birth to."

I suggest to call those structures 'contrastive comments', since they go against a presupposition about the topic which was built in the preceding context.

## 2.5 Function of $[V_{\text{sbj}} (N) (N)] N_{\text{ANN}}$

Those structures are characterized by the presence of a noun in the annexed state<sup>10</sup> after the prosodic boundary. Here again, the internal composition of the prosodic unit is not important. It is the state of the noun and the position after the boundary, which formally identify the construction. This type of construction is often called ‘right-dislocated’ in the literature. I will not use that term because I make no assumptions concerning possible underlying structures.

All examples involving such a construction are used to activate a referent that had lost its active (and even semi-active) status. This reactivation is generally associated with further continuation of the discourse with the activated referent as topic. In the following example, the house is reactivated after having stayed unmentioned for a few intonation units, and then the following subtopic starts (the girls set to explore the house, which is described in details).

- (17) *t-ufa d amfif n wədrar //*  
 SBJ3SG.F-find:PFV COP cat:ABS.SG.M GEN mountain:ANN.SG.M //  
 “She found it was the Mountain Cat
- i=t izədyən / wəxxam-nni //*  
 REL.REAL=ABSV3SG.M dwell:PFV:RELSBJ.POS / house:ANN.SG.M-CNS //  
 who inhabited it, the house.”

## 2.6 Synthesis on information structure

The study of formal sequences based on position relative to the verb, and to a prosodic boundary, provides a number of structures which can be grouped together on the basis of function.

Construction  $[V_{\text{sbj}} (N_{\text{abs}})]$ , without any surrounding nouns belonging to the clause outside the prosodic group of the verb, marks (sub-)topic continuation. Typically, those structures are realized as sequences of verbs with their obligatory person affix, possibly complemented by nominal direct objects.

$[V_{\text{sbj}}]$   
 $[V_{\text{sbj}} N_{\text{abs}}]$

10. If the noun is in the absolute state and after the prosodic group of the verb in Kabyle, it cannot be related to the preceding verb, but has to be construed as starting a new clause. This is due to the function of the state distinction (see Mettouchi & Frajzyngier 2013).



Construction  $[V_{\text{subj}} N_{\text{ann}} (N)]$  has no surrounding nouns belonging to the clause outside the prosodic group of the verb either, but there is at least one noun after the verb, and it bears the annexed state. Typically, those are structures where the nominal subject follows the verb within the same prosodic unit. A nominal direct object can be present too, and in that case both relative orders of nouns are possible, without any change in the function of the construction. The function of this construction is topicalization (promotion to topic status) of an event or a state, in athetic perspective. This construction is used to present situations or events as a whole as new, regardless of the activation status of the referents themselves: they inform the listener, providing a statement that triggers a new subtopic.

$[V_{\text{subj}} N_{\text{ann}}]$   
 $[V_{\text{subj}} N_{\text{ann}} N_{\text{abs}}]$   
 $[V_{\text{subj}} N_{\text{abs}} N_{\text{ann}}]$

Construction  $[N_{\text{abs}} V_{\text{subj}} (N)]$  has no surrounding nouns belonging to the clause outside the prosodic group of the verb either, and its defining feature is the presence of a noun in the absolute state just before the verb and still within the prosodic group containing the verb. Another noun can appear after the verb. This construction recapitulates salient elements of a situation which have been exposed before at some length. It is a summary, which provides backgrounding for the following discourse.

$[N_{\text{abs}} V_{\text{subj}}]$   
 $[N_{\text{abs}} V_{\text{subj}} N_{\text{abs}}]$   
 $[N_{\text{abs}} V_{\text{subj}} N_{\text{ann}}]$

Construction  $N_{\text{abs}} [V_{\text{subj}} (N) (N)]$  is characterized by the presence of a noun in the absolute before the prosodic boundary opening on the prosodic unit containing the verb. This noun has to bear the special continuative (rising) boundary tone that links it to the following sequence; otherwise it would not be interpreted as a topic but as belonging to the previous clause. This construction is binary, as it involves a topic and a comment. The topic is a referent reactivated from the previous discourse, but the important element here is the comment: it goes against a presupposition about the topic that was built in the preceding context. This is why I propose to label the function of this construction ‘contrastive comment’.

Construction  $[V_{\text{subj}} (N) (N)] N_{\text{ann}}$  is characterized by the presence of a noun in the annexed state after the right prosodic boundary of the prosodic unit containing the verb. The annexed state here is the mirror image of the continuative boundary tone on the topic in the absolute in the abovementioned construction: its role is to tie this noun to the clause and to indicate that it does not belong to the following clause. The function of this construction is the reactivation of a participant for topic promotion.

### 3. Grammatical relations<sup>11</sup>

I have just shown that constructions characterized by word order and prosodic grouping, and to some extent state alternation, had specific informational values in discourse. My hypothesis in this second part is that information structure values imply that in some cases grammatical relations on nouns should be transparent, whereas in other cases, they needn't be. Indeed, in this section, I show that grammatical relations are not systematically coded on nouns. For this, I am using the same non-aprioristic method as in Part 1, with a query in mind: to what extent are prosodic grouping, state alternation, and word order also involved in the encoding of grammatical relations in Kabyle, at the level of the clause? This way of investigating things makes it possible to show how discourse and clause-level grammar interact in spontaneous speech in Kabyle.

The bound pronouns that are obligatorily affixed to the verb form a special paradigm that codes the main participant in the situation, regardless of semantic role, animacy, topicality, etc.<sup>12</sup> This paradigm can therefore be considered as a subject paradigm. As the verb and its subject affix alone constitute a well-formed sentence, I do not consider them as agreement markers, but as a case of pronominal argument marking (Galand 1964; Mettouchi 2005). Other bound pronominal paradigms are absolutive clitics (objects of transitive verbs and main participant of non-verbal predicates) and dative clitics (affected participant).<sup>13</sup>

All pronominal paradigms in Kabyle are inflected for person, number, and gender (in the second and third person (singular and plural) for the subject and

11. A preliminary version of this section was written during a period of collaborative work with Zygmunt Frajzyngier on the state opposition in Kabyle. I am grateful to him for his comments on those hypotheses and for the role his methodology played in the elaboration of my argumentation.

12. Here is the subject pronominal paradigm, with perfective radical *-uli-*, “go up”, “climb”.

	Singular	Plural
1	uli-γ	n-uli
2 M	t-uli-dʕ	t-uli-m
2 F		t-uli-mt
3 M	j-uli	uli-n
3 F	t-uli	uli-nt

13. Clitics are characterized by their ability to climb from the default position after the verb to the position after the negative, modal, or aspectual particle preceding the verb or to the position after the relative marker in relative clauses.

absolutive paradigms, in the second (singular and plural), and third (plural) for the dative paradigm).

- (18) *tə-wwət=it*  
 3SG.F-kick.PFV=ABSV3SG.M  
 “She kicked it.”
- (19) *ulaʃ=it*  
 EXNEG=ABSV3.SG.M  
 “He is not there.”
- (20) *tə-mmut=as*                      *təqʃɪt*  
 3SG.F-die.PFV=DAT3SG.M girl.ANN.SG.F  
 “She lost a daughter (lit. a girl died on her).”

In this context, noun phrases are often used as expansions of the bound pronouns. This is true for NPs coreferent to subject affixes, or absolutive ones. Indirect objects are introduced by preposition *i*, and they won’t be considered here because their role is quite transparent given their prepositional phrase structure. Only one type of NP is not coreferent to any pronoun, and this is the nominal direct object, as will be shown below.

Formally, all we can rely on, for the computation of grammatical relations on nouns, is the state of the noun, its gender-number markers, its position with respect to the verb and the prosodic boundary, and the gender-number markers on the pronominal affixes and clitics.

I will now show the following:

- a. the state opposition in itself does not mark grammatical relations;
- b. coreference in gender and number between the noun and the bound pronoun, in itself, is not transparent for the encoding of grammatical relations; and
- c. word order in itself does not mark grammatical relations.

However, the interaction of state, word order, and prosodic grouping allows the computation of grammatical relations for nouns.

### 3.1 Grammatical relations are not marked unambiguously by one coding means

#### 3.1.1 *The state opposition in itself does not mark grammatical relations*

This first statement is demonstrated by the fact that the noun in the annexed state can be a subject or an object. If the same mark is compatible with two grammatical relations that are in principle incompatible, then its function is not to indicate grammatical relations. In the following example, the noun in the annexed state has the same referent as the subject pronoun of the clause:

- (21) *tə-mmut                      təqfiʃt //*  
 SBJ3SGF-die.PFV girl.ANN.SG.F  
 “The girl is dead.”

In the following example, the noun “house” in the annexed state has the same referent as the ‘object’ (absolutive) pronoun =*t*:

- (22) *t-ufa                      d      amʃif                      n      wədrar //*  
 SBJ3SG.F-find:PFV COP cat.ABS.M.SG GEN mountain.ANN.M.SG //  
*i=t                      izədyən /                      wəxxam-nni //*  
 REL=ABSV3SG.M inhabit:PFV:RELSBJ.POS / house.ANN.M.SG-CNS //  
 “She found it was the Mountain Cat who inhabited it, the house.”

### 3.1.2 *Coreference in gender and number in itself does not mark grammatical relations*

This second statement is demonstrated by the fact that a noun can have the same features of gender and number than the subject affix, without being coreferent with it. This shows that identity in number and gender (often called ‘agreement’) is not sufficient to mark grammatical role.

- (23) *jə-krəz                      igər*  
 SBJ3M.SG-plough.PFV field:ABS.M.SG  
 “He ploughed the field” (and not “The field is ploughed”, for which the noun needs to be in the annexed state)

According to Frajzyngier & Shay (2003: 64), agreement “must occur in any clause with a singular nominal or pronominal subject in the clause and it cannot occur if there is no nominal singular subject in the clause”. This definition of agreement clearly excludes what happens in Berber in terms of person-gender-number marking on the verb, since this marking is obligatory, regardless of the presence of a nominal argument in the clause. This brings Frajzyngier & Shay (2003: 64) to the following conclusion: “in many languages, so-called agreement phenomena are actually independent coding means in that they occur regardless of whether the argument that they code appears in the clause”.

I will not pursue in detail the role of person-number-gender marking on the verb, but observe that on the one hand this mark is necessary for the verb to become a clause, a predication, and on the other hand it plays a role in referent-tracking in discourse, and, as shown in Section 1.1, if not accompanied by a coreferential noun in the same clause (in the prosodic group of the verb or in the immediately preceding prosodic unit), has the function of marking continuing topic as far as information structure is concerned.

### 3.1.3 *Word order in itself does not mark grammatical relations*

As nouns that are computable as subjects and objects can either precede or follow the verb and as, if they follow the verb, there is no fixed ordering between them, word order in itself is not sufficient to mark grammatical relations.

In (24) the noun in the position before the verb can be interpreted as subject or object:

- (24) *ayəɾdaj-nni jə-čča=t*  
 rat.ABS.M.SG-CNS SBJ3M.SG-eat.PFV=ABSV3M.SG  
 “He ate the rat.” or “The rat ate it.”

In (25)–(26) the nouns following the verb can be interpreted as subject or object, the position just after the verb does not code subject or object exclusively. We have to additionally take into account the state marked on the noun.

- (25) *jə-swa wəmfif ajfki*  
 SBJ3M.SG-drink.PFV cat.ANN.SG.M milk.ABS.SG.M  
 “The cat drank milk.”

- (26) *jə-swa ajfki wəmfif*  
 SBJ3M.SG-drink.PFV milk.ABS.SG.M cat.ANN.SG.M  
 “The cat drank milk.”

As none of those coding means alone transparently code grammatical relations, we have to hypothesize either that grammatical relations are not relevant for nouns in Kabyle, but only for pronouns, or that they are transparently retrievable, but through the interaction of several coding means.

## 3.2 The interaction of state, position, prosodic grouping, and gender-number marking

Investigation of the various coding shows that the first distinction is between nouns that are outside of the prosodic group of the verb and nouns that are inside.

### 3.2.1 *Nouns outside the prosodic group of the verb*

In the position before the opening prosodic boundary of the prosodic group containing the verb, nouns are in the absolute state and can have any grammatical role.

In the following example, the NP *tamtɛtut n lqbaɟl* can be interpreted as subject of the following verb.

- (27) *tamɛttut*                      *n*                      *lqɛajl* /  
 woman.ABS.F.SG GEN kabyle\_tribe.ANN.PL /  
 “The Kabyle woman,  
       *ad=dd*                      *t-kkr*/                      *ad*                      *t-ruħ*  
 POT=PROX SBJ3SG.F-stand\_up:AOR POT SBJ3SG.F-go:AOR  
 she would stand up, she would go  
       *ad=dd*                      *t-zdəm* /  
 POT=PROX SBJ3SG.F-gather\_wood:AOR /  
 gather wood (...)”

However other roles also appear in this position: in the following example, the noun *ajtma* is coreferent with the dative clitic = *asn*.

- (28) *ajtma* /                      *t-uy-ɛ=asn=idd* /  
 brother:ABS.M.PL / SBJ2-buy:PFV-SBJ2SG=DAT3PL.M=PROX /  
 “My brothers, you bought them things.”

The position before the prosodic boundary opening on the prosodic group of the verb is therefore not a coding means for grammatical relations. The referents themselves can be retrieved through coreference in gender and number with one of the pronominal affixes or clitics surrounding the verb. However, ambiguity is always possible if more than one pronoun has the same features of gender and number as the initial noun:

- (29) *taqɛift* /                      *t-uy=as=idd*                      *taksiwt*  
 girl:ABS.SG.F / SBJ3SG.F-take:PFV=DAT3SG=PROX dress:ABS.SG.F  
 “The girl bought her a dress.” or “She bought a dress for the girl.”

Gender-number identity of features between pronoun and noun is therefore not a coding means for grammatical relations either.

In the position before the prosodic boundary, grammatical relations cannot be transparently computed.

In the position after the prosodic boundary closing the prosodic group of the verb, nouns are in the annexed state, and can have any grammatical role.

In the following example, the noun *wəmyar* is coreferent with the possessor pronoun *-is* on the noun; its grammatical role, if one is to be attributed to it, is possessor.

- (30) *tə-mmut*                      *tmɛttut-is* /                      *wəmyar-nni* //  
 SBJ3SG.F-die:PFV woman.ANN.F.SG-KIN3SG / old\_man.ANN.M.SG-CNS //  
 “His wife died, that man”

In the following example, the noun in the annexed state after the prosodic group of the verb is coreferent with the subject pronoun; its grammatical role, if one is to be attributed to it, is subject.

- (31) *ad=dd      ħku-y /      amk i /      tʃɪf-nt*  
 POT=PROX tell:AOR-SBJ1SG / how REL.REAL / live:IPFV-SBJ3PL.F  
*zik /      lɣalat      n      lqbaɟl-nnəy /*  
*long\_ago / woman.ANN.PL.F GEN kabyle\_tribe.ANN.PL-POSS1PL/*  
 “I will tell how they lived in the old days, the Kabyle women.”

Those are not the only grammatical roles that can be found in this position. Others are object, kinship relationship (but not indirect object, which is always preceded by a preposition).

The position after the prosodic boundary closing on the prosodic group of the verb is therefore not a coding means for grammatical relations. The referents corresponding to the pronouns can be retrieved through coreference in gender and number with one of the pronominal affixes or clitics surrounding the verb, but ambiguities in gender and number can always arise.

In the position after the prosodic boundary, grammatical relations cannot therefore be transparently computed either.

### 3.2.2 *N<sub>abs</sub> after the verb within the prosodic group of the verb*

This situation is much more constrained than the one described in § 2.2.1: the absolute state implies that the noun is not to be interpreted as the variable of a function grammaticalized on the preceding constituent (Mettouchi & Frajzyngier 2013), namely the subject role marked by the bound pronoun on the verb. By default, it is interpreted as the direct object.

- (32) *i-sʃa      taqədsit /*  
 SBJ3SGM-possess:PFV herd.ABS.F.SG  
 “He had a herd.”

Only two noun phrases can appear after the verb within its prosodic group: one in the absolute state, one in the annexed state. The latter is necessarily the nominal subject (no other grammatical interpretation is possible for such nouns in such position), and the former the nominal object. The relative ordering of the two noun phrases is irrelevant, as state is here sufficient to disambiguate the grammatical role of each NP (cf. § 3.1.3). The noun in the absolute can therefore immediately follow the verb, as above, or be separated from it by another noun, as in (33).

- (33) *i-sfa*                      *uməkṣa*                      *t-a-qəḍṣi-t* /  
 SBJ3SGM-possess:PFV shepherd.ANN.SG.M herd.ABS.F.SG  
 “The shepherd had a herd.”

If only position was involved, the computation of grammatical relations would not be possible. But since one of the nouns must be in the absolute state and the other in the annexed state, the two roles cannot be confused. The noun in the absolute appears only when the verb is transitive, and it refers to its second argument. It is therefore the object. The nominal object can therefore be defined as a noun in the absolute state following the verb inside the prosodic group of the verb.

No coreference is involved because in Kabyle the nominal object is the only direct complement in postverbal position, which does not corefer to a pronoun, within the prosodic group of the verb.

### 3.2.3 *N<sub>ann</sub> after the verb within the prosodic group of the verb*

Only one noun phrase in the annexed state can occur within the prosodic group of the verb. This noun is always coreferent with the subject affix. If the construction is intransitive, only the noun in the annexed state occurs. If it is transitive, a noun in the absolute may appear, which is computed as an object (cf 3.2.2.).

The combination of annexed state and position (following (immediately or not) the verb within the prosodic group of the verb) provides unambiguous instructions for the decoding of the grammatical relation ‘subject’.

- (34) sp1:  
 <ça fait> *t-mmūt=as*                      *təqfiṭ*  
 it\_is              SBJ3SG.F-die:PFV=DAT3SG girl.ANN.SG.F  
*i*              *Zahwa Tasliṭ* //  
 DAT Zahwa daughter\_of\_Ali //  
 “So she lost a girl, Zahwa Taalits?”
- sp2:  
*t-mmūt=as*                      *tmənzut* //  
 SBJ3SG.F-die:PFV=DAT3SG elder.ANN.SG.F //  
 “Her eldest daughter died (on her).”

### 3.2.4 *N<sub>abs</sub> before the verb within the prosodic group of the verb*

The noun in the absolute is unambiguously the subject if and only if there is no clitic pronoun in the prosodic group of the verb (only the subject affix). Indeed, in the position before the verb, the noun has to be coreferent with a pronoun affixed or cliticized to the verb.



- (35) *taq3unt-nni t-ssəglaf /*  
 dog.ABS.SG.F-CNS SBJ3SG.F-bark:CAUS.IPFV /  
 “The dog was barking.”
- (36) *tiq3ifin ur mmut-nt ara /*  
 girl:ABS.PL.F NEG die:PFV-SBJ3PL.F POSTNEG /  
 “The girls didn’t die.”

If there are one or two clitics the grammatical relation is no longer transparent and has to be disambiguated thanks to gender-number feature identity. In the following example, based only on coreference in gender and number, the noun can be computed as the affected object, but if the subject affix was feminine, ambiguity would arise and the noun *jiwāt* could be interpreted as the subject as well:

- (37) *jiwāt jə-fka=jas nnəfš //*  
 one:F SBJ3SG.M-give:PFV=DAT3SG half //  
 “He gave half (an apple) to one (of his wives).”

In the position before the verb, the noun can be transparently coded as subject if and only if there are no pronouns cliticized to the verb, only the subject affix.

### 3.4 Implications

This raises the question of what can be called a function: do we call something a function only if there is a one-to-one relationship between a coding means and a value? In that case only pronominal paradigms mark grammatical relations, and nouns can only be indirectly computed as subjects or objects.

If we accept that a function can be marked by the interaction of several coding means (i.e., a construction), then we can say that grammatical relations are coded on nouns in Kabyle but that this coding is complex and only holds within some structures.

In Kabyle, nominal subjects and objects can only be unambiguously computed within the prosodic group of the verb:

- a. A noun is a nominal subject if and only if, within the prosodic group of the verb:
  - the verb has no clitics other than the subject affix AND the noun occurs before the verb, and is in the absolute state; and
  - the noun occurs after the verb (immediately or not) and is in the annexed state.
- b. A noun is a nominal object if and only if, within the prosodic group of the verb, the noun occurs after the verb (immediately or not) and is in the absolute state.

As far as information structure is concerned, functions are clearly complex and involve the presence or absence of the noun, its state, and its position relative to the verb and prosodic boundaries.

#### 4. General conclusion

Why is it that in a language such as Kabyle, nominal subject and nominal object are only coded in specific environments? I suggest that this has to do with the fact that the language has pronominal argument marking instead of agreement. Mithun (1992: 58) already links that kind of language with pragmatically-based word order, and the appositive role nouns play in those languages, as opposed to pronouns, which bear the primary case relations to the verb. I have shown that in Kabyle nouns are not simply appositive (for instance, nominal objects can appear without an object clitic) but that their presence and ordering code a number of pragmatic functions and, in some cases, syntactic ones.

The question now is why subject and object grammatical relations are relevant only in some contexts when nouns are involved, whereas they are always unambiguously coded by bound pronouns (subject affixes and absolutive clitics). A look at those contexts may provide an answer: structures involving two nominal arguments (or one when the verb is intransitive) within the prosodic group of the verb are either topicalizations of events or states with a *thetic* perspective or recapitulations of salient elements of a previously narrated episode. In both cases, the construal of the situation is not of the 'comment on a topic', or 'topic continuation', or 'reactivation of a referent' type, but rather implies the role of each participant in the event or state. This type of context therefore calls for disambiguation of grammatical roles, which is what we see in Kabyle. Other structures conveying information of the 'aboutness' or 'topic promotion' type are less tied to their argument structure; what is important there is referential information and topical information. In those structures, grammatical relations on nouns are not unambiguously retrievable.

It would be interesting to see whether pronominal argument languages with no basic word order and a pragmatically-based ordering in the sense of Mithun (1992) tend to restrict to some constructions only the unambiguous coding of grammatical relations on nouns, leaving other constructions grammatically unspecified regarding the interpretation of nouns as subjects or objects.

## References

- Chafe, Wallace. 1994. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: The University of Chicago Press.
- Chaker, Salem. 1988. "Annexion". *Encyclopédie Berbère* 5.686–695.
- Cruttenden, Alan. 1997. *Intonation* (Cambridge Textbooks in Linguistics), 2nd edn. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139166973
- Frajzyngier, Zygmunt & Erin Shay. 2003. *Explaining Language Structures through Systems Interaction*. Amsterdam: John Benjamins. doi:10.1075/tsl.55
- Galand, Lionel. 1964. "L'énoncé verbal en berbère. Etude de fonctions". *Cahiers Ferdinand de Saussure* 21.33–53.
- Givón, Talmy (ed.). 1983. *Topic Continuity in Discourse*. Amsterdam: John Benjamins. doi:10.1075/tsl.3
- Izre'el, Shlomo & Amina Mettouchi. 2015. "Representation of Speech in CorpAfroAs: Transcriptional Strategies and Prosodic Units". *Corpus-based Studies of Lesser-described Languages: The CorpAfroAs Corpus of Spoken Afroasiatic Languages*, ed. by Amina Mettouchi, Martine Vanhove & Dominique Caubet, 13–41. Amsterdam: John Benjamins.
- Kuningas, J. & Jaakko Leino. 2006. "Word Orders and Construction Grammar". *Festschrift in Honour of Fred Karlsson. SKY Journal of Linguistics* 19.301–309.
- Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents* (Cambridge Studies in Linguistics 71). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511620607
- Mettouchi, Amina. 2005. "Discourse-Configurationality and the Encoding of Semantic Macroroles in Taqbaylit Berber". *Studi afroasiatici. Contributi presentati all'XI Incontro italiano di linguistica camito-semitica (Bergamo, 5–7 giugno 2003)*, ed. by Alessandro Mengozzi, 83–96. Milano: Franco Angeli.
- Mettouchi, Amina. 2006. "Anaphoricité et appel à l'attention partagée dans un conte oral en kabyle (berbère)". *Loquentes Linguis, Studi linguistiche e orientali in onore di Fabrizio A. Pennacchietti*, ed. by Piergiorgio Borbone, Alessandro Mengozzi & Mauro Tosco, 499–507. Wiesbaden: Harrassowitz.
- Mettouchi, Amina. 2007a. "Preferred Argument Structure in Taqbaylit Berber: A Preliminary Study". *Proceedings of the XIIth Italian Meeting of Afroasiatic Linguistics, Università di Catania, Ragusa (Italie), les 6–10 juin 2005*, ed. by Marco Moriggi & Mirella Cassarino, 295–303. Soveria Mannelli (Cosenza): Rubbettino.
- Mettouchi, Amina. 2007b. "Word Order in Conversational Taqbaylit Berber: Preposed and Postposed Subjects". "From Beyond the Mediterranean". *Akten des 7. Internationalen Semito-hamitistenkongresses (Berlin 13–15 September 2004)*, ed. by Rainer Voigt, 513–531. Aachen: Shaker.
- Mettouchi, Amina. 2008. "Case-Marking, Syntactic Domains and Information Structure in Kabyle (Berber)". *Interaction of Syntax and Morphology: Case Studies in Afroasiatic*, ed. by Zygmunt Frajzyngier & Erin Shay, 7–40. Amsterdam: John Benjamins. doi:10.1075/tsl.75.02met
- Mettouchi, Amina. 2011. "Démonstratifs et construction de la référence en kabyle". « *Parcours berbères* », *Mélanges offerts à Paulette Galand-Pernet et Lionel Galand pour leur 90e anniversaire*, ed. by Amina Mettouchi, 469–484. Köln: Rüdiger Köppe.

- Mettouchi, Amina & Zygmunt Frajzyngier. 2013. "A Previously Unrecognized Typological Category: The State Distinction in Kabyle (Berber)". *Linguistic Typology* 17:1.1–30. doi:10.1515/lity-2013-0001
- Mithun, Marianne. 1992. "Is Basic Word Order Universal?" *Pragmatics of Word Order Flexibility*, ed. by Doris Payne, 15–62. Amsterdam: John Benjamins. doi:10.1075/tsl.22.02mit
- Pierrehumbert, Janet & Julia Hirschberg. 1990. "The Meaning of Intonational Contours in the Interpretation of Discourse". *Intentions in Communications*, ed. by Philip R. Cohen, Jerry Morgan & Martha E. Pollack, 271–311. Cambridge, MA: MIT.

**PAGE PROOFS**

© JOHN BENJAMINS PUBLISHING COMPANY

**2nd proofs**