

History of the Ancient Near East / Monographs – X

CAMSEMUD 2007

PROCEEDINGS OF THE 13TH ITALIAN MEETING OF AFRO-ASIATIC LINGUISTICS

Held in Udine, May 21st–24th, 2007

Edited by

FREDERICK MARIO FALES & GIULIA FRANCESCA GRASSI



**S.A.R.G.O.N. Editrice e Libreria
Padova 2010**

HANE / M – Vol. X

History of the Ancient Near East / Monographs

Editor-in-Chief: Frederick Mario Fales

Editor: Giovanni B. Lanfranchi

ISBN 978-88-95672-05-2
4227-204540



© S.A.R.G.O.N. Editrice e Libreria
Via Induno 18B I-35134 Padova
SAR.GON@libero.it
I edizione: Padova, aprile 2010
Proprietà letteraria riservata

Distributed by:

Eisenbrauns, Winona Lake, Indiana 46590-0275 USA
<http://www.eisenbrauns.com>

Stampa a cura di / Printed by:
Centro Copia Stecchini – Via S. Sofia 58 – I-35121, Padova

S.A.R.G.O.N. Editrice e Libreria
Padova 2010

TABLE OF CONTENTS

F.M. Fales – G.F. Grassi, <i>Foreword</i>	v
I. SAILING FROM THE ADRIATIC TO ASIA/AFRICA AND BACK	
G.F. Grassi, <i>Semitic Onomastics in Roman Aquileia</i>	1
F. Aspesi, <i>A margine del sostrato linguistico “labirintico” egeo-cananaico</i>	33
F. Israel, <i>Alpha, beta ... tra storia–archeologia e fonetica, tra sintassi ed epigrafia</i>	39
E. Braida, <i>Il Romanzo del saggio Ahiqar: una proposta stemmatica</i>	49
F.A. Pennacchietti, <i>Il tortuoso percorso dell’antroponimo Asia tra omofoni e sviste</i>	65
G. Cifoletti, <i>Venezia e l’espansione dell’italiano in Oriente: problemi connessi con la storia della lingua franca del Mediterraneo</i>	69
II. GENERAL AND COMPARATIVE AFROASIATIC LINGUISTICS	
G. Del Olmo Lete, <i>Phonetic Distribution in Semitic Binary Articulation Bases</i>	79
M. Franci, <i>Estensione della radice nella comparazione egitto-semitica</i>	87
P. Marrassini, <i>South Semitic Again</i>	103
G. Hudson, <i>Klimov’s Active-language Characteristics in Ethiopian Semitic</i>	111
O. Kapeliuk, <i>Some Common Innovations in Neo-Semitic</i>	123
H. Jungraithmayr, <i>Mubi and Semitic — Striking Parallels</i>	133
A. Zaborski, <i>‘Afar-Saho and the Position of Cushitic within Hamitosemitic/ Afroasiatic</i>	139
V. Blažek, <i>On Application of Glottochronology to South Berber (Tuareg) Languages</i>	149
A. Mettouchi, D. Caubet, M. Vanhove, M. Tosco, Bernard Comrie, Sh. Izre’el, CORPAFROAS. <i>A Corpus for Spoken Afroasiatic Languages: Morphosyntactic and Prosodic Analysis</i>	177
III. NORTHWEST SEMITIC	
A. Gianto, <i>Guessing, Doubting, and Northwest Semitic YAQTUL-U</i>	181
F.M. Fales, <i>New Light on Assyro-Aramaic Interference: The Assur Ostrakon</i>	189
A. Faraj, <i>An Incantation Bowl of Biblical Verses and a Syriac Incantation Bowl for the Protection of a House</i>	205
I. Zatelli, <i>Performative Utterances in the Later Phase of Ancient Hebrew: the Case of Ben Sira’</i>	213

S. Destefanis, <i>I Proverbi di Ahiqar nella versione neoaramaica di Rubeyl Muhattas. Un'analisi comparativa delle sue fonti</i>	221
R. Kim, <i>Towards a Historical Phonology of Modern Aramaic: The Relative Chronology of Turoyo Sound Changes</i>	229
IV. EGYPTIAN	
H. Satzinger, <i>Scratchy Sounds Getting Smooth: the Egyptian Velar Fricatives and Their Palatalization</i>	239
G. Takács, <i>The Etymology of Egyptian $\sqrt{m}3\text{f}$</i>	247
F. Contardi, <i>Egyptian Terms Used to Indicate the Act of Reading: An Investigation about the Act of Reading in the Egyptian Society</i>	261
A. Roccati, <i>Sono dei Re quelli specificati per nome (hq3w pw mtrw rnw)</i>	271
V. ARABIC	
A.Gr. Belova, <i>Études étymologiques du lexique arabe préislamique: correspondances sémitiques et le cas de la spécification</i>	275
J. Lentin, <i>Sur quelques préformantes utilisées dans la morphogénèse de la racine: l'exemple de l'arabe</i>	281
A. Mengozzi, <i>The History of Garshuni as a Writing System: Evidence from the Rabbula Codex</i>	297
R. Contini, <i>Travel Literature as a Linguistic Source: Another Look at Doughty's Najdi Arabic Glossary</i>	305
W.C. Young, T. Rockwood, <i>Explaining Variation in Demonstrative Morphology and Syntax in Peninsular Colloquial Arabic: An Argument Based on Anaphoric and Exophoric Reference</i>	315
J. Guardi, <i>Il 'āmil nella linguistica araba moderna</i>	339
B. Airò, <i>Aspetti e tendenze degli studi di linguistica araba in Tunisia (1985–2005)</i>	349
VI. CHADIC	
O. Stolbova, <i>Chadic Lateral Fricatives (Reconstruction and Parallels in Semitic, Cushitic and Egyptian)</i>	355
R. Leger, A. Suzzi Valli, <i>The Lexeme "eye" in Chadic Reconsidered</i>	369
S. Baldi, R. Leger, <i>North versus South. Typological Features of Southern Bole-Tangale Languages</i>	375
VII. CUSHITIC	
M. Tosco, <i>Semelfactive Verbs, Plurative Nouns: On Number in Gawwada (Cushitic)</i>	385
VIII. BERBER	
V. Brugnatelli, <i>Problème de la négation en berbère: à propos de l'origine d'ulac, ula, ula d</i>	401

CORPAFROAS. A CORPUS FOR SPOKEN AFROASIATIC LANGUAGES: MORPHOSYNTACTIC AND PROSODIC ANALYSIS

Amina Mettouchi – Dominique Caubet – Martine Vanhove
Mauro Tosco – Bernard Comrie – Shlomo Izre'el

1. Introduction

This short paper presents the main lines and aims of a current project funded by the Agence Nationale de la Recherche (ANR) in France. The project¹ is a pilot corpus, which will be put on-line for the community at the end of 2010.

The background for the project is the current lack of transcribed and sound-indexed spoken data, in Afroasiatic languages.² In a technical context where it is becoming easier to compile databases, and to provide long-term archives for sound files, it appeared urgent to propose a methodology, a series of tools, and scientific analyses that may help the community of researchers compile their data. To that effect, we have been working on segmentation and glossing for the past two years, and will do so for the next two.

The final purpose of this pilot corpus is not so much the amount of data it will provide, as the full set of methods and electronic tools that will be developed and made available freely to the community.

Ultimately, the pilot corpus is designed to grow and become a reference corpus, as well as to inspire initiatives for other language phyla.

2. The Languages and the partners

The project is coordinated by Amina Mettouchi, professor at the University of Nantes, France, and researcher in the LLING laboratory. It involves two other French laboratories, the LLACAN (CNRS Villejuif), headed by Martine Vanhove, and the CREAM-LacNad (Inalco Paris), headed by Dominique Caubet. Other researchers from the Netherlands and Italy participate in the project, as associate members of the LLING laboratory.

¹ <http://web.me.com/aminamettouchi/CORPAFROAS/Abstract.html>.

² See however the Semitisches Tonarchiv of the University of Heidelberg (<http://www.semarch.uni-hd.de/index.php4>), and the Corpus of Spoken Israeli Hebrew at Tel-Aviv University (<http://www.tau.ac.il/humanities/semitic/cosih.html>).

As we deemed it necessary to have an external control over the project, as well as expertise, we asked two renowned scholars to be our experts on the two main axes of the project: professor Bernard Comrie (Max Planck Institute in Leipzig, and UCSB, Santa Barbara) for typological glossing, and professor Shlomo Izre'el (Tel Aviv University) for prosodic segmentation.

The members of the project are the following:

- LLING & Associates: Amina Mettouchi, Jean-Pierre Angoujard, Hamida Demirdache, Azeb Amha, Mauro Tosco, Graziano Sava, Stefano Manfredi, Il-il Malibert-Yatziv, Cécile Lux, Shlomo Izre'el, Bernard Comrie.
- LLACAN: Martine Vanhove, Bernard Caron, Marie-Claude Simeone-Senelle, Raymond Boyd, Christian Chanard.
- LACNAD-CREAM: Dominique Caubet, Christophe Pereira, Maha Abourahim, Elgar-Paul Magro, Alexandrine Barontini, Angeles Vicente.

The languages represented in the project are: Kabyle, Tamashek (Berber), Beja, Gawwada, Ts'amakko, Ongota, Afar (Cushitic), Wolaitta (Omotic), spoken Moroccan, Lybian and Sudanese Arabic, Maltese, spoken Hebrew, Dahalik (Semitic), Hausa, Bata, and Zaa (Chadic).

One hour per language will be thoroughly transcribed, glossed, translated into English, and sound-indexed. The sound will be available in .wav format.

In its pilot form, the corpus is not designed to present a balanced sample of languages. It covers all branches, and different types of languages, so that we have an opportunity to provide technical and scientific solutions for all potential types: tonal and intonational, concatenative and non-concatenative, endangered and well-described languages, with or without codeswitching, *etc.*

3. The scientific analysis

The project is organized along two axes, linked to the nature of the materials and to the aim of the project, which is typological comparability among languages.

Prosody

The body of data is spoken, and we have decided to take into account this oral dimension by working on segmentation. We do not use the punctuation system of written texts, because it is not adapted to the specificities of the spoken language (Wichmann 2000). Instead, we are adapting the widely accepted system of boundary-marking used for instance in the C-ORAL-Rom developed by Cresti & Moneglia.³

We therefore analyze the prosodic units of our languages into minor (non-terminal) and major (terminal) units, using the software Praat.⁴ No other specification (tones, contours *etc.*) is given to those boundaries, but the fact that the transcription is indexed to the sound, itself available in .wav format, will allow more in-depth prosodic studies on the available data.

The outcome of our analyses will be both the segmented pilot corpus, and a series of publications dealing with the nature of minor and major units and their universal or language-dependent definition.

³ <http://lablita.dit.unifi.it/coralrom/>.

⁴ Paul Boersma & David Weenink, <http://www.fon.hum.uva.nl/praat/>.

Morphosyntax

The corpus is not only translated, but also interlinearly glossed. For this purpose, we have developed a format allowing several annotation tiers. Those tiers are aimed at the automatic retrieval of all the relevant queries concerning Afroasiatic languages: pronominal systems, case systems, nominal predicates, aspect, ideophones, demonstratives, verbal derivation, *etc.* The questions underlying the glossing process are numerous, among them: how will the non-concatenative morphology of certain Semitic and Berber varieties be dealt with? And how will suprasegmental morphemes, as found in Chadic and Cushitic, be annotated and retrieved?

The theoretical question underlying our project, is that of the tension between comparability and language-internal coherence: what is the optimal degree of unification of the annotations, in order to both respect the specificities of languages, and provide a comparative basis for typology?

The annotation system we chose is based on the Leipzig Glossing Rules,⁵ developed jointly by the Department of Linguistics of the Max Planck Institute for Evolutionary Anthropology (Bernard Comrie, Martin Haspelmath) and by the Department of Linguistics of the University of Leipzig (Balthasar Bickel), in order to promote convergence in glossing systems. The list of morphemes being open, and the rules being devised more for readability than for automatic retrieval, one of the tasks currently under way is the establishment of a complete list, with rules adapted to computer retrieval.

The outcome of our analyses will be both the glossed pilot corpus, and a series of publications dealing with the tension between language-specific categories and configurations, and typological comparability and generalizations.

4. The technical aspects

The corpus will be put on-line in the CRDO (Centre de Ressources pour la Description de l'Oral),⁶ with the XML software ELAN and associated tools.

As the data will be made available online to the community, a thorough reflection process was engaged before data collection, concerning the deontological aspects of the project. Thus, anonymization procedures, as well as control over sensitive data, have been implemented. In this process, we followed international recommendations, stated in *Corpus Oraux, Guide des Bonnes Pratiques*⁷ (Baude 2006).

At the same time, all the relevant information was listed, in order to provide rich metadata on the recordings. These metadata follow the requirements of OLAC⁸ (Open Language Archives Community).

The recordings were all digital, with strict requirements as to the format: non-compressed, .wav files, recorded at 44,1 khz / 16 bits, with high-quality microphones and pre-amplifiers. The high quality of the recording is necessary, not only because one of our scientific aims is to conduct a prosodic analysis of the data, but also for conservation purposes.

The software used for the analysis of the data is: Praat⁹ for segmentation, and Toolbox¹⁰ for primary annotation. They are widely used by their respective communities, prosodicists, and field linguists. The data is then exported into Elan.¹¹ The version of Elan we are developing is part of the deliverables

⁵ http://www.eva.mpg.de/lingua/tools-at-lingboard/glossing_rules.php.

⁶ <http://crdo.risc.cnrs.fr/exist/crdo/>.

⁷ http://www.culture.gouv.fr/culture/dglf/Guide_Corpus_Oraux_2005.pdf.

⁸ <http://www.language-archives.org>.

⁹ <http://www.fon.hum.uva.nl/praat/>.

¹⁰ <http://www.sil.org/computing/toolbox/>.

¹¹ <http://www.lat-mpi.eu/tools/elan/>.

of the project: it will allow isolated researchers, and other teams, to compile data in a dedicated technical environment, with all the requirements provided by our scientific teamwork.

The technical dimension of the project implies of course the participation of an engineer on a permanent basis (Christian Chanard, of the LLACAN laboratory), as well as a developer specially hired for the project.

5. Conclusion

CORPAFROAS is a unique endeavour, in that it will provide the first corpus of Afroasiatic languages that will present sound-indexed data, with elaborated annotations. It is also unique in that it will be freely accessible, and will be accompanied by software, tools, and publications that will make it easier for field linguists to contribute to CORPAFROAS, or compile their own corpus, without having to go through the complex preparatory theoretical and technical procedures that we are working on.

Its release will be advertised in the relevant media, in 2010. For information, contact Professor Amina Mettouchi, at the University of Nantes, France.

BIBLIOGRAPHY

O. Baude (éd.) 2006, *Corpus Oraux, Guide des bonnes pratiques*, Paris 2006.

A. Wichmann 2000, *Intonation in Text & Discourse: Beginnings, Middles and Ends*, Harlow 2000.