

**Computational Approaches
to
Semitic Languages**

Proceedings of the Workshop

Sponsored by Mid Med Bank, Malta

Edited by Michael Rosner

**16th August 1998
Université de Montréal
Montreal, Quebec, Canada**

COLING-ACL'98

© 1998 Université de Montréal.

Additional copies of these proceedings may be ordered while stocks last
from:

Association for Computational Linguistics (ACL)
75 Paterson Street, Suite 9
New Brunswick, NJ 08901 USA
Tel: +1-732-342-9100
Fax: +1-732-873-0014
rasmusse@cs.rutgers.edu

Preface

Although there exists a considerable body of CL research specifically targeted to Semitic languages, much of the work to date has been the result of initiatives undertaken by individual researchers or research establishments. A direct consequence is that there is comparatively little awareness amongst practitioners of either the state of the art as practiced outside their own locality, the common challenges faced by all practitioners, or the potential for developing a coordinated approach. The aims of this workshop are therefore:

- To provide a forum where current work in a broad range of subfields can be presented, collected and diffused.
- To assess the state of the art with a view to identifying promising areas for future collaborative research.
- To examine the possibilities of supporting such research through national and international funding initiatives.

After refereeing, submitted papers, which concern Arabic, Hebrew, Maltese and Akkadian, were classified, rather coarsely, into the following subject areas: tagging Arabic text, orthography, morphology, syntax/parsing/generation. The workshop programme comprises a session for each of these themes, together with a short supplementary session for short papers. As can be seen from the programme, there will be ample time for discussion.

I am greatly indebted to colleagues in the programme committee, who also doubled as referees, for their help, patience and comments, all of which have contributed towards what promises to be a fruitful and stimulating exchange of views.

Michael Rosner
University of Malta
June 1998

PROGRAMME COMMITTEE MEMBERS:

Michael Rosner, University of Malta, Malta (Workshop Coordinator)
Ken Beesley, Xerox Research Centre Europe, France
Joe Caruana, University of Malta, Malta
Yaacov Choueka, Bar Ilan University, Israel
Khalid Choukri, ELRA/ELDA, France
Fathi Debili, CNRS-CELLMA (Paris) / IRMC, Tunisia
Ray Fabri, University of Malta, Malta
Mamoun Hattab, Arabic Textware, Amman, Jordan
George Kiraz, Bell Labs, USA
Chadia Moghrabi, University of Moncton, Canada
Mori Rimon, Hebrew University, Israel

SPONSOR:

Mid-Med Bank, Malta

Programme

08:55–09:00 Welcome – Michael Rosner

Session 1: Tagging

09:00–09:30 Saleem Abuleil and Martha Evens (Illinois Institute of Technology, USA) *Discovering Lexical Information by Tagging Arabic Newspaper Text*

09:30–10:00 John Maloney and Michael Niv (SRA International Corp, USA) *TAGARAB: A Fast, Accurate Arabic Name Recogniser Using High Precision Morphological Analysis*

10:00–10:30 Fathi Debili and Emna Souissi (CNRS-CELLMA/IRMC, Tunisia) *Etiquetage grammatical de l'arabe voyellé ou non*

10:30–10:40 Discussion

10:40–11:00 Coffee Break

Session 2: Orthography

11:00–11:30 Michel Fanton (CERTAL/INALCO, France) *Finite State Automata and Arabic Writing*

11:30–12:00 Bonnie Glover Stalls and Kevin Knight (ISI, University of Southern California, USA) *Translating Names and Technical Terms in Arabic Text*

12:00–12:30 Fathi Debili and Hadhemi Achour, (CNRS-CELLMA/IRMC, Tunisia) *Voyellation automatique de l'arabe*

12:30–12:40 Discussion

12:40–13:30 Lunch Break

Session 3: Morphology

13:30–14:00 Kenneth R. Beesley (Xerox Research Centre Europe, France) *Arabic Morphology using only Finite-State Operations*

14:00–14:30 Martine Smets (University of Sussex, UK) *Paradigmatic Treatment of Arabic Morphology*

14:30–15:00 Riyadh Al-Shalabi and Martha Evens (Illinois Institute of Technology, USA) *A Computational Morphology System for Arabic*

15:00–15:30 François Barthélemy (Galatasaray University, Turkey and CNAM, France) *A Morphological Analyser for Akkadian Verbal Forms with a Model of Phonetic Transformations*

15:30–15:40 Discussion

15:40–16:00 Coffee Break

Session 4: Syntax, Parsing and Generation

16:00–16:30 Shuly Wintner (University of Tübingen, Germany) *Towards a Linguistically Motivated Computational Grammar for Hebrew*

16:30–17:00 Yael Dahan Netzer and Michael Elhadad (Ben Gurion University, Israel) *Generating Determiners and Quantifiers in Hebrew*

17:00–17:10 Discussion

Session 5: Short Papers

17:10–17:20 Michael Rosner, Joe Caruana and Ray Fabri (University of Malta) *Maltilex: A Computational Lexicon for Maltese*

17:20–17:30 Paul Micallef (University of Malta) *Rule-Based Lexical Analysis of Maltese*

17:30–17:40 Ali M. Obaid (Technical University of Budapest, Hungary) *A New Pattern-Matching Approach to the Recognition of Printed Arabic*

17:40–18:00 Discussion/Closing Session

Table of Contents

Preface	i
Programme	ii
Table of Contents	iii
Author Index	iv

Workshop Papers

Saleem Abuleil and Martha Evens <i>Discovering Lexical Information by Tagging Arabic Newspaper Text</i>	1
John Maloney and Michael Niv <i>TAGARAB: A Fast, Accurate Arabic Name Recogniser using High-Precision Morphological Analysis</i>	8
Fathi Debili and Emna Souissi <i>Etiquetage grammatical de l'arabe voyellé ou non</i>	16
Michel Fanton <i>Finite State Automata and Arabic Writing</i>	26
Bonnie Glover Stalls and Kevin Knight <i>Translating Names and Technical Terms in Arabic Text</i>	34
Fathi Debili and Hadhémi Achour <i>Voyellation automatique de l'arabe</i>	42
Kenneth R. Beesley <i>Arabic Morphology using only Finite-State Operations</i>	50
Martine Smets <i>Paradigmatic Treatment of Arabic Morphology</i>	58
Riyad Al-Shalabi and Martha Evens <i>A Computational Morphology System for Arabic</i>	66
François Barthélemy <i>A Morphological Analyser for Akkadian Verbal Forms with a Model of Phonetic Transformations</i>	73
Shuly Wintner <i>Towards a Linguistically Motivated Computational Grammar for Hebrew</i>	82
Yael Dahan Netzer and Michael Elhadad <i>Generating Determiners and Quantifiers in Hebrew</i>	89
Michael Rosner, Joe Caruana and Ray Fabri <i>Maltilex: A Computational Lexicon for Maltese</i>	97
Paul Micallef <i>Rule Based Lexical Analysis of Maltese</i>	102
Ali M. Obaid <i>A New Pattern Matching Approach to the Recognition of Printed Arabic</i>	106

Author Index

Abuleil, S.	1
Achour, H.	42
Al-Shalabi, R.	66
Barthélemy, F.	73
Beesley, K.R.	50
Caruana, J.	97
Debili, F.	16, 42
Elhadad, M.	89
Evens, M.	1, 66
Fabri, R.	97
Fanton, M.	26
Knight, K.	34
Maloney, J.	8
Micallef, P.	102
Netzer, Y.D.	89
Niv, M.	8
Obaid, A.M.	106
Rosner, M.	97
Smets, M.	58
Souissi, E.	16
Stalls, B.G.	34
Wintner, S.	82