

EACL 2009

**Proceedings of the
EACL 2009 Workshop on
Computational Approaches
to Semitic Languages**

31 March 2009

Megaron Athens International Conference Centre
Athens, Greece

Production and Manufacturing by
TEHNOGRAFIA DIGITAL PRESS
7 Ektoros Street
152 35 Vrilissia
Athens, Greece

©2009 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

Preface

We are delighted to present you with this volume containing the papers accepted for presentation at the EACL 2009 Workshop on Computational Approaches to Semitic Languages, held in Athens, Greece, on March 31st, 2009.

The Workshop is an opportunity for the ACL Special Interest Group on Computational Approaches to Semitic Languages to meet and discuss current and future directions in Computational Linguistic and Natural Language Processing approaches to Semitic Languages. Previous workshops took place in 1998 (Montreal), 2002 (Philadelphia), 2005 (Ann Arbor) and 2007 (Prague). We are happy to see more, better submissions every year. This year we received 19 submissions, of which 9 were selected for presentation and are included in this volume. These papers address Arabic (both standard and dialectal), Hebrew and Ancient Syriac, as well as general issues pertaining to all Semitic languages. We are privileged to have Prof. Michael Gasser (Indiana University) as our keynote speaker this year. Professor Gasser will talk on Finite state morphology for Ethiopian Semitic languages.

We are extremely grateful to the members of the Program Committee who reviewed the submissions and helped shape up the Program: Ann Bies (LDC, University of Pennsylvania, USA), Tim Buckwalter (University of Maryland, USA), Violetta Cavalli-Sforza (Carnegie Mellon University, USA), Joseph Dichy (University of Lyon 2, France), Michael Elhadad (Ben Gurion University, Israel), Martha W. Evens (Illinois Institute of Technology, USA), Ray Fabri (University of Malta), Ali Farghaly (Oracle, USA), Alexander Fraser (University of Stuttgart, Germany), Andrew Freeman (Washington University, USA), Albert Gatt (University of Aberdeen, UK), Gregory Grefenstette, (Exalead, France), Nizar Habash (Columbia University, USA), Alon Itai (Technion, Israel), Steven Krauwer (Utrecht University, Netherlands), Mohamed Maamouri (LDC, University of Pennsylvania USA), Bente Maegaard (CST, University of Copenhagen, Denmark), Nurit Melnik (Oranim College, Israel), Uzzi Ornan (Technion, Israel), Owen Rambow (Columbia University, USA), Paolo Rosso (Universidad Politecnica Valencia, Spain), Khalil Sima'an (University of Amsterdam, Netherlands), Abdelhadi Soudi (Ecole Nationale de l'Industrie Minerale, Morocco), Adam Ussishkin (University of Arizona, USA), and Imed Zitouni (IBM Research, USA)

Above all, we would like to thank the authors and the participants of the Workshop. We hope you have a very enjoyable and productive time in Athens!

Mike Rosner and Shuly Wintner
Program Committee Chairs

Table of Contents

<i>How to Establish a Verbal Paradigm on the Basis of Ancient Syriac Manuscripts</i> Wido van Peursen	1
<i>The Karamel System and Semitic Languages: Structured Multi-Tiered Morphology</i> François Barthélemy	10
<i>Revisiting Multi-Tape Automata for Semitic Morphological Analysis and Generation</i> Mans Hulden	19
<i>A Hybrid Approach for Building Arabic Diacritizer</i> Khaled Shaalan, Hitham M. Abo Bakr and Ibrahim Ziedan	27
<i>Unsupervised Concept Discovery In Hebrew Using Simple Unsupervised Word Prefix Segmentation for Hebrew and Arabic</i> Elad Dinur, Dmitry Davidov and Ari Rappoport	36
<i>Automatic Treebank-Based Acquisition of Arabic LFG Dependency Structures</i> Lamia Tounsi, Mohammed Attia and Josef van Genabith	45
<i>Spoken Arabic Dialect Identification Using Phonotactic Modeling</i> Fadi Biadisy, Julia Hirschberg and Nizar Habash	53
<i>Structure-Based Evaluation of an Arabic Semantic Query Expansion Using the JIRS Passage Retrieval System</i> Lahsen Abouenour, Karim Bouzoubaa and Paolo Rosso	62
<i>Syntactic Reordering for English-Arabic Phrase-Based Machine Translation</i> Jakob Elming and Nizar Habash	69

Conference Program

Tuesday, March 31, 2009

9:00–10:00 Opening Remarks and Invited Talk by Michael Gasser

Session 1: Philology

10:00–10:30 *How to Establish a Verbal Paradigm on the Basis of Ancient Syriac Manuscripts*
Wido van Peursen

Session 2: Morphology

11:00–11:30 *The Karamel System and Semitic Languages: Structured Multi-Tiered Morphology*
François Barthélemy

11:30–12:00 *Revisiting Multi-Tape Automata for Semitic Morphological Analysis and Generation*
Mans Hulden

12:00–12:30 *A Hybrid Approach for Building Arabic Diacritizer*
Khaled Shaalan, Hitham M. Abo Bakr and Ibrahim Ziedan

Session 3: Applications

14:30–15:00 *Unsupervised Concept Discovery In Hebrew Using Simple Unsupervised Word Prefix Segmentation for Hebrew and Arabic*
Elad Dinur, Dmitry Davidov and Ari Rappoport

15:00–15:30 *Automatic Treebank-Based Acquisition of Arabic LFG Dependency Structures*
Lamia Tounsi, Mohammed Attia and Josef van Genabith

15:30–16:00 *Spoken Arabic Dialect Identification Using Phonotactic Modeling*
Fadi Biadisy, Julia Hirschberg and Nizar Habash

Tuesday, March 31, 2009 (continued)

Session 4: Applications

16:30–17:00 *Structure-Based Evaluation of an Arabic Semantic Query Expansion Using the JIRS Passage Retrieval System*

Lahsen Abouenour, Karim Bouzoubaa and Paolo Rosso

17:00–17:30 *Syntactic Reordering for English-Arabic Phrase-Based Machine Translation*

Jakob Elming and Nizar Habash

Closing

17:30–18:00 Discussion and final remarks