

Mona T. Diab, PhD
Associate Professor
Department of Computer Science
School of Engineering and Applied Science
The George Washington University
mtdiab@gwu.edu
<http://www.cs.gwu.edu/people/faculty>
<http://www.seas.gwu.edu/~mtdiab>
Office: +1(202) 994.8109

EDUCATION

- **Ph.D. Computational Linguistics** **Aug '03**
Linguistics Department
University of Maryland Institute for Advanced Computer Studies (UMIACS)
University of Maryland, College Park, USA
Thesis: Word Sense Disambiguation within a Multilingual Framework
Advisor: Philip Resnik
- **M.Sc. Computer Science (Machine Learning and Artificial Intelligence)** **Jun '97**
The George Washington University, USA
Advisor: Peter Bock
- **B.Sc. Computer Science & Software Engineering** **Jun '92**
The American University in Cairo, Egypt
- **B.Sc. Tourist Guidance (Egyptology & Archaeology)** **Jun '90**
Faculty of Tourism and Hotel Management, Helwan University, Egypt

PROFESSIONAL EXPERIENCE

- **Research Scientist** **Feb '09 – Dec '12**
*Center for Computational Learning Systems (CCLS), **Columbia University**, NY, USA*
- **Adjunct Associate Professor** **Feb '09 – Dec '12**
*Computer Science Department, **Columbia University**, NY, USA*
- **Associate Research Scientist** **Jul '05 – Jan '09**
*Center for Computational Learning Systems (CCLS), **Columbia University**, NY, USA*
- **Postdoctoral Research Scientist** **Feb '05 – Jun '05**
*Center for Computational Learning Systems (CCLS), **Columbia University**, NY, USA*
- **Postdoctoral Research Scholar/Scientist** **Jan '04 – Jan '05**
*Linguistics Department and NLP Group, **Stanford University**, CA, USA*
Mentor: Daniel Jurafsky (MacArthur Award Fellow, 2002-2003)
- **Research Associate** **Jul '03 – Dec '03**
*Center for Spoken Language Research, **University of Colorado Boulder**, CO, USA*
Mentor: Daniel Jurafsky (MacArthur Award Fellow, 2002-2003)
- **Graduate Research Assistant** **Jul '97 – Jun '03**

*Institute for Advanced Computer Studies (UMIACS), **University of Maryland College Park**, MD, USA*

Advisor: Philip Resnik

- **Visiting Researcher** **Sep '02 – Dec '02**
*Laboratoire Perception Systemes Information, **Université de Rouen**, France*
Visiting with Thierry Paquet
- **Research Assistant** **May '98 – Aug '98**
Thomson Publishing NLP Labs, Rockville, MD, USA
Mentor: Stephen Finch
- **Systems Analyst** **Mar '97 – Aug '97**
Solutions Consultants International, Arlington, VA, USA
- **Graduate Teaching Instructor** **Fall '95 – Spr '97**
Computer Science Department, The George Washington University, Washington, DC, USA
- **Graduate Teaching Assistant** **Fall '93 – Spr '95**
Computer Science Department, The American University in Cairo, Egypt
- **Software Engineer** **Apr '93 – Feb '95**
Regional Inform. Tech. Software Engineering Center (RITSEC), UNDP, Cairo, Egypt
- **Licensed Tourist Guide** **Jun '91 – Jun '95**
Free Lance, Cairo, Egypt
- **English News Broadcast Anchor** **Sep '92 – Feb '93**
Saudi Arabia National Radio, Riyadh, Saudi Arabia

BOARD MEMBERSHIP

- **Qordoba Inc.** **Sep '11 –present**
Advisory Board Member
Dubai, UAE
- **Cairo Microsoft Research Innovation Center (CMIC)** **May '10 –Jun '11**
Advisory Board Member
One of 6 Microsoft Research Innovation Centers World Wide, Cairo, Egypt
- **Springer Publisher: NLP Briefs** **Jul '11 – present**
Advisory Board Member
Berlin, Germany
- **(ISO) Semantic Annotation Framework –Named Entities** **Apr '11 –present**
Advisory Board Member
International Standards Organization
- **Human Language Technology Center of Excellence, JHU** **Jun '09 –present**
Advisory Board Member
SCALE
Johns Hopkins University, MD, USA

ELECTED POSITIONS

- **NAACL Executive Board Member** **Jan '12 --present**
- **Columbia University Senator** **Aug '11 –Dec '12**
One of 6 elected members campus wide representing >3000 Research Officers in University Senate

- **Secretary of ACL Special Interest Group on the Lexicon (SIGLEX)** Jun '10 –present
- **President of ACL Special Interest Group on Semitic Languages (SIG Semitic)** Nov '11 –present
- **Secretary of ACL Special Interest Group on Semitic Language Processing (SIG Semitic)** Jun '09 – Oct '10
- **Executive Board Member of ACL SIGLEX** Oct '07 – Jul '10

ACADEMIC INITIATIVES

- ***SEM: First Comprehensive Semantics Conference** Jul '11
Founder (with Eneko Agirre, Johan Bos)
Creating a consolidated Unified Semantics Community
- **CALIMAT: Columbia Arabic Language Institute for Modeling and Technology** Jul '07
Founder (with Hatim Diab, Nizar Habash, Owen Rambow, David Waltz)
Seeking Joint Institutional Collaborations with the Arab World
Columbia University, NY, USA
- **CADIM Group** Jun '05
Founder (with Nizar Habash and Owen Rambow)
Columbia Arabic Dialects Modeling Group
Center for Computational Learning Systems (CCLS), Columbia University, NY, USA
<http://www.ccls.columbia.edu/oldweb/cadim/>

KEYNOTE SPEECHES

- Invited Talk/Keynote Speaker. 27th Arabic Linguistic Symposium (ALS 2013), Indiana, USA, Feb 28th- Mar 2nd, 2013 **2013**
- Plenary Invited Talk (Main Plenary Speech), '11th International Language Engineering Conference, Cairo, Egypt, 14 Dec '11. **2011**
- *Plenary talk: A Unified Semantic Perspective, the *SEM Initiative.* **IEEE 5th International Conference on Semantic Computing (ICSC)**, Palo Alto, CA, USA, Sep '11.
- *Advances and Challenges in Arabic Natural Language Processing.* **IEEE Information and Communication Technologies International Symposium (ICTIS'07)**, Fez, Morocco. Apr '07. **2007**
- *Automatic Processing of Arabic(s).* **6th Conference on Language Engineering**, Cairo, Egypt. Nov '06. **2006**
- *Building Automatic Resources.* **EACL workshop on Cross Language Knowledge Induction**, Trento, Italy. Apr '06.
- *Automatic Ambiguity resolution in MSA and dialectal Arabic.* **Arabic Linguistics Society (ALS) Conference**, Kalamazoo, MI, USA, Mar '06.
- *Bootstrapping an Arabic WordNet Automatically.* **MEANING Workshop, European Union & IRST**, Trento, Italy. Feb '05. **2005**

INVITED TALKS/PARTICIPATION/PANELIST

2013

- Invited Talk, Computer Science Symposium, CLIP Lab, UMIACS, University of Maryland College Park, Maryland, USA, Feb 13th 2013.

2012

- Invited Panelist on Computational lexical resources and interoperability, **LREC 2012**, Istanbul Turkey
- Invited Arbiter/Referree (One of 6 Arab Expat Scientists to decide on Best Scientific Research in Qatar), **Qatar Foundation, Arab Expat Society & Scientific Research Forum**, Qatar, 19 -23rd Oct '12.

2011

- Invited Participant (One of 50 Arab Expat Scientists to discuss and advise on future of Scientific Research Strategy in Qatar), **Qatar Foundation, Arab Expat Society & Scientific Research Forum**, 19 -23rd Nov '11.
- Invited Panelist, Future of Semantics in Machine Translation, **DARPA BOLT PI Kick-off Meeting**, Tampa, Florida, USA, Oct '11.
- Invited Talk (to the President & Deans), **King Abdullah University for Science and Technology (KAUST)**, Thuwwal, Saudi Arabia, 13 Sep '11.
- Invited Talk (Mathematical & Computer Sciences School), **King Abdullah University for Science and Technology (KAUST)**, Thuwwal, Saudi Arabia, 13 Sep '11.
- Invited Participant in **NSF-sponsored Workshop on Collaborative Resource Development and Delivery**. Organized by Nancy Ide (Vassar College), Held in New York, NY, USA, 7-8 Oct '11.
- Invited Participant in **NSF-funded Adjectives Workshop**, Organized by Christiane Fellbaum (Princeton University), Held in Arlington, VA, USA, 29-30 Sep '11.
- Invited Panelist at **EMNLP 2011 Workshop on Geometrical Methods in Semantics (GEMS)**, Edinburgh, UK, Jul '11.
- Invited Talk, **Distinguished Lecture Symposium Series, University of Washington**, Seattle, WA, USA, 30 Apr '11.
- Invited Talk, **Microsoft Research**, Redmond, WA, USA, 21 Apr '11.
- Invited Participant in **NSF-funded SILT Workshop Interoperable Semantic Annotation**. Organized by Nancy Ide (Vassar College) and James Pustejovsky, (Brandeis University), Held in Waltham, MA, USA, 11-14 Apr '11.
- Invited Participant in **QCRI Social Networks Roundtable**, Qatar Computing Research Institute, Qatar Foundation, Doha, Qatar, 3-4 Apr '11.
- Invited Participant in **QCRI Arabic Language Technologies Roundtable**, Qatar Computing Research Institute, Qatar Foundation, Doha, Qatar, '16-17 Jan '11.

2010

- Invited Talk, **Computer Science Department, University of Alabama**, Birmingham, AL, USA, Dec '10.
- Invited Panelist at **COLING 2010 Workshop on Multiword Expressions (MWE 2010)**, Beijing, China, Aug '10.
- Invited Panelist at **ACL 2010 Workshop on Semantic Evaluations (SEMEVAL 2010)**, Upsala, Sweden, 17 Jul '10.
- Invited Talk and Participant in **Workshop on the Future of Arabic Language Development**, organized by the **UAE Crown Prince Court for Strategic Planning**, Abu Dhabi, UAE, Jul '10.

- Invited Talk, **IBM TJ Watson**, Hawthorne, NY, USA, Jun '10.
- Invited Talk, **School of Computing and Mathematical Sciences, King Abdullah University for Science and Technology, KAUST**, Thuwwal, Saudi Arabia, May '10.
- Invited Talk, **Computer Science Department, University of Trento**, Trento, Italy, Apr '10.

2009

- Invited Panelist, **NAACL Workshop on Semantic Evaluations (SEMEVAL '09)**, Boulder, CO, USA, Jun '09.
- Invited Talk, **Departamento de Sistemas Informaticos y Computacion, Universidad Politecnica de Valencia**, Valencia, Spain, May '09.
- Invited Talk, **IRST & Computer Science Department, University of Trento**, Trento, Italy, Apr '09.

2008

- Invited Talk, **Department of Informatics, University of Sussex**, Brighton, England, Aug '08.
- Invited Talk, **Computer Science Department, Faculty of Science, University of Mohamed V**, Rabat, Morocco, May '08.
- Invited Talk, **Departamento de Sistemas Informaticos y Computacion, Universidad Politecnica de Valencia**, Valencia, Spain, May '08.
- Invited Talk, **Linguistics Department, Georgetown University**, Washington, DC, USA, Apr '08.

2006

- Invited Talk, **Cairo Microsoft Innovation Center (CMIC), MSR**, Smart Village, Cairo Egypt, Dec '06.
- Invited Talk, **Computer Science Department, University of North Texas**, Dallas, TX, USA, Sep '06.

2005

- Invited Talk, (with Nizar Habash and Owen Rambow), **Linguistic Data Consortium, University of Pennsylvania**, Philadelphia, PA, USA, Dec '05.
- Invited Talk, **Computer Science Department, New York University**, New York, NY, USA, Oct '05.
- Invited Talk, **Computer Science Department, Columbia University**, New York, NY, USA, Mar '05.
- Invited Talk, **Linguistics Department, Ohio State University**, Columbus, OH, USA, Feb '05.
- Invited Talk, **Linguistics Department, University of Texas Austin**, Austin, TX, USA, Feb '05.
- Invited Talk, **Yahoo Research Labs, Overture**, Pasadena, CA, USA, Jan '05.

2004

- Invited Talk, **Language Computer Corporation (LCC)**, Dallas, TX, USA, Dec '04.
- Invited Talk, **Microsoft Research (MSR), Natural Language Group**, Redmond, WA, USA, Dec '04.
- Invited Talk, **Department of Informatics, University of Edinburgh**, Edinburgh, Scotland, Aug '04.
- Invited Talk, **Linguistic Data Consortium, University of Pennsylvania**, Philadelphia, PA, USA, May '04.
- Invited Talk, **IBM T. J. Watson Research Center**, Hawthorne, NY, USA, May '04.

2003

- Invited Talk, **Linguistics Dept. & Institute for Cognitive Science, University of Colorado, Boulder**, Boulder, CO, USA, Nov '03.

2002

- Invited Talk, **Laboratoire Perception, Systèmes, Information (PSI), Université de Rouen**, Rouen, France, Nov '02.
- Invited Talk, **NLP Lab, Information Sciences Institute (ISI), University of Southern California**, Marina Del Rey, CA, USA, Nov '02.

PROFESSIONAL CONSULTING

- **Qordoba Inc.** Sep '11 – present
Providing Expert Assessment of English Arabic MT and NLP Technologies
Contact: May Habib
Dubai, UAE
- **University of Colorado, Boulder** Spr '10 – present
Arabic Propbank, Expert advice predicate argument structures for Arabic
Contact: Martha Palmer
Boulder, CO, USA
- **Cairo Microsoft Research Innovation Center (CMIC)** Oct '07 – Jun '08
Expert advice on Natural Language Processing and Arabic Language Technologies
Contact: Tarek AlAbbady
Cairo, Egypt
- **Egyptian Center of Excellence for Data Mining** Apr '06 – Dec '07
Expert advice on Statistical Natural Language Processing
Contact: Aly Fahmy, Nayer Wanas
Cairo, Egypt
- **Princeton University** Sep '05 – Dec '07
Arabic WordNet Project, Expert advice on Arabic Lexical Semantics
Contact: Christiane Fellbaum
Princeton, NJ, USA
- **Meaningful Machines Inc.** Feb '06 – Jun '06
Providing Expert advice on Arabic Lexical Semantics for Machine Translation
Contact: Jaime Carbonnell
New York, NY, USA

RESEARCH EXPERIENCE

Current Research Grants

- **[DARPA-DEFT] DRATS: Detecting Relations and Anomalies in Text and Speech, (PI;** subaward from Columbia University as Prime institution). Total \$5.7M, duration 4.5 yrs., Nov '12 – May '18.
- **[NSF-Large CRI] Creating Resources for Multilingual Linguistic Code Switching, (PI;** with Julia Hirschberg, collaborative with Tamar Solorio, UAB, Columbia Prime institution). Total \$400K, duration 3 yrs., Sep '12 – Aug '15.
- **[TSWG-DOD/ODNI-NADIA] Natural Language Processing for Dialectal Arabic–Industrialization and Insertion (PI;** with Owen Rambow and Nizar Habash). Total \$898K, duration 1.5 yrs., Apr '12 – Aug '13.
- **[Google Faculty Research Awards] Nuanced Sentiment and Perspective Analysis for Arabic Social Media Text (PI).** Total/yr. \$71,339K, duration 1 yr., Jan '13 – Dec '13.
- **[DARPA-MADCAT] Subcontract to BBN–Multilingual Automatic Document**

Classification Analysis and Translation for Arabic OCR (Co-PI; with Nizar Habash). Total/yr. (Apr '08-Mar '12, \$750K) \$153K, Duration 5 yrs., Apr '08 –Aug '13.

- **[DARPA-BOLT]** Subcontract to GWU from BBN – *Dialectal Tool and Resource Creation (PI; with at Columbia University Nizar Habash and Owen Rambow). Total/yr. \$1.1M, duration 5 yrs., Oct '11 –Dec '15.*
- **[DARPA-BOLT]** Subcontract to SRI–*Semantic Machine Translation for Arabic Dialects (CCLS-PI; with Kathy Mckeown, Michael Collins, Julia Hirschberg). Total/yr. \$1.7M, duration 1 yrs., Nov '11 –Aug '13.*

Past Research Grants

- **[DARPA-BOLT]** Subcontract to LDC–*Resource Guideline Creation for Arabic Dialects (Co-PI; with Nizar Habash and Owen Rambow). Total/yr. \$100K, duration 5 yrs., Nov '11 –Dec '12.*
- **[DARPA-GALE]** Subcontract to IBM–*Pilot Unified Framework for Semantic Textual Similarity (PI). Total \$25K, duration 8 months, Sep '11 –Apr '12.*
- **[NSA-TTO8]** Subcontract to JHU-COE–*Power and Modality Analysis and Prediction in Social Communication (Co-PI; with Owen Rambow). Total/yr. \$115K, duration 3 yrs., Sep '09 –Dec '12.*
- **[IARPA-SCIL]** *Power Rifts and Shifts in Social Communication (Co-PI; with Kathy McKeown and Owen Rambow). Total/yr. \$642K, duration 5 yrs., Aug '09 –Oct '12.*
- **[NSF-CRI Planning]** *Linguistic Code Switching Pilot Annotation (PI; with Julia Hirschberg). Total/yr. \$78K, duration 1 yr., Mar '10–Mar '11.*
- **[TSWG-DOD/ODNI-COLABA]** Subcontract to ACXION–*Cross language blog alerts–Arabic Blog Processing (PI; with Nizar Habash and Owen Rambow). Total \$780K+\$50K (CAT/CAIM support), duration 3 yrs., Aug '09–Apr '11.*
- **[DARPA-GALE]** Subcontract to IBM-Rosetta–*Arabic to English Machine Translation (Co-PI; with Owen Rambow and Nizar Habash). Total/yr. \$390K, duration 3 yrs., May '08–Apr '11.*
- **[NSA-TTO8]** Subcontract to JHU-COE–*Committed Belief and Modality in Multigenres Text (Co-PI; with Owen Rambow). Total/yr. \$620K, duration 3 yrs., Aug '07 –Aug '10.*
- **[DARPA-GALE]** Subcontract to SRI-Nightingale–*Arabic to English Machine Translation (Co-PI; with Owen Rambow and Nizar Habash). Total/yr. ~\$500K, duration 3 yrs., Sep '05–Apr '08.*
- **[NSF-SGER]** *Automatic Processing of Natural Language Code-Switching (PI; with Owen Rambow). Total \$41K, duration 1 yr., Sep '07–Aug '08.*
- **[NSF-SGER]** *Inducing Multiword Expressions from Comparable Corpora (PI). Total \$19K, duration 6 mo., Sep '06–Feb '07.*
- **[NSF-JHU Summer Workshop]** *Arabic Dialect Parsing (Co-PI). Total \$15K, duration 1.5 mo., Jun '05–Jul '05.*
- **[DARPA-GALE]** Subcontract to University of Colorado–*OntoNotes. Building an Arabic PropBank. (PI). Total \$15K, duration 6 mo., Sep '05–Feb '06.*

RESEARCH EDUCATION AND TEACHING EXPERIENCE

Semester Long Classes

- **CSCI 3907/6907 Statistical NLP**
Department of Computer Science

Spr '13

- The George Washington University*

▪ **Projects in Computer Science**

Computer Science Dept.

Columbia University

Fall '07 – Dec '12
- **Arabic Computational Linguistics**

Linguistics/Computer Science Dept.

Georgetown University

Fall '09
- **C as a Second Language**

Computer Science Dept.

George Washington University

Fall '95 – Spr '97
- **Theory of Computing**

Fall '91 – Spr '95
- **Algorithms and Formal Languages**
- **Artificial Intelligence**

Computer Science Dept.

The American University in Cairo

Tutorials

- **ARABIC DIALECT PROCESSING (3-5 hours)**

(Co-Taught with Nizar Habash)

NAACL 2012

Montreal, Canada

Jun '12
- Georgetown University Round Table (GURT)**

Washington DC, USA

Apr '10
- 2nd International Conference on Arabic language Resources and Tools (MEDAR)**

Cairo, Egypt

Apr '09
- Language Resources & Evaluation Conference (LREC)**

Marrakech, Morocco

May '08
- Human Language Technologies–North American Association for Computational Linguistics (HLT-NAACL)**

Rochester, NY, USA

Jun '07
- American Machine Translation Association (AMTA)**

Boston, MA, USA

Aug '06
- **AN A-Z STATISTICAL NLP CRASH COURSE WITH A FOCUS ON ARABIC LANGUAGE PROCESSING (4 hours)**

IEEE Information and Communication Technologies International Symposium (ICTIS'07)

Fez, Morocco

Mar '07
- **STATISTICAL NATURAL LANGUAGE PROCESSING: AN A-Z CRASH COURSE (3 days)**

Faculty of Engineering and Computer Studies, Cairo University

Cairo, Egypt

Sep '06
- **INTRODUCTION TO COMPUTATIONAL LINGUISTICS (2 hours)**

Linguistics Department,

University of Texas Austin

Austin, TX, USA

Feb '05

Post Doctoral Mentorship

- **Davide Picca**

May '10 – Apr '11

Received PhD April 2010: University of Lausanne, Switzerland

Postdoc Focus: Unsupervised semantic modeling for comparable corpora

Current Position: Research Technical Liaison, University of Lausanne, Switzerland

- **Yassine BenAjiba** **Fall '09 – Sep '10**

Received PhD May 2009: Polytechnical University of Valencia, Spain

Postdoc Focus: Shallow syntactic parsing in Arabic and integration in SMT and IE

Current Position: Research Scientist at Phillips Corporation, NY, USA

- **Marine Carpuat** **May '10 – Apr '11**

Received PhD April 2008: HKUST, Hong Kong

PostDoc Focus: Incorporating Lexical Semantics in Statistical Machine Translation

Current Position: Research Scientist at National Research Council of Canada

PhD Thesis Committee Member

- **Integrating Semantics in Statistical Machine Translation** **Mar '12**
Mahmoud Ghoneim
Computer Engineering Department
Cairo University. Cairo Egypt
Served as co-supervisor
- **Reverse Feature Engineering for SVMs** **Apr '10**
Daniele Pighin
Computer Science Department
University of Trento, Italy
- **Semantic Fusion and Generation** **Dec '09**
Kapil Thadani
Computer Science Department
Columbia University
- **Arabic Named Entity Recognition** **May '09**
Yassine Benajiba
Departamento de Sistemas Informaticos y Computacion
Universidad Politecnica de Valencia, Spain
- **Automatic Processing of Language Frame Semantics** **Apr '09**
Bonaventura Coppola
ICT division
University of Trento, Italy
- **Information Retrieval: English and Arabic Documents** **May '08**
Fadoua Ataa Allah
Faculty of Science
University of Mohamed V, Rabat, Morocco

Student Thesis & Research Supervision

- **Ali Elkahky** **Spr '12 – present**
PhD, Computer Science Department, Columbia University
Thesis Topic: Adaptive Semantic Role Labeling for Machine Translation in Morphologically rich languages (with a focus on Informal Genres)
- **Aya Zirikly** **Spr '12 – present**
PhD, Computer Science Department, George Washington University
Joint Supervision
Thesis Topic: Tracking of Named Entities in Multilingual and Multigenre Settings

- **Heba Elfardy** Fall '10 – present
PhD, Computer Science Department, Columbia University
Thesis Topic: Identifying Perspective in Social Media
- **Vinodkumar Prabhakaram** Fall '09 – present
PhD, Computer Science Department, Columbia University
Thesis Topic: Power Structures and Modality in Multilingual Social Communication
- **Weiwei Guo** Fall '07 – present
PhD, Computer Science Department, Columbia University
Thesis Topic: Semantically Nuanced Models for Natural Language Processing
- **Muhammad AbdulMageed** Spr '10 – present
PhD, Ling. Dept., Information Sciences Dept., Indiana University
Joint Supervision
Thesis Topic: Sentiment Analysis and Perspective in Arabic Social Media
- **Mahmoud Ghoneim** Fall '06 – Mar '12
PhD, Computer Engineering Department, Faculty of Engineering, Cairo University
Joint Supervision
Thesis Topic: Explicit Modeling of Semantics in Statistical Machine Translation
- **Pradeep Dasigi** Fall '10 – present
MSc, Computer Science Department, Columbia University
Thesis Topic: Dialectal Arabic Orthography Normalization

PUBLICATIONS

Books

- **Diab, Mona** and Christiane Fellbaum. (*forthcoming*) Computational Lexical Semantics: A Multilingual Perspective. Springer Publishers, expected Fall 2013
- **Diab, Mona**, Nizar Habash, Owen Rambow. (*forthcoming*) Processing Arabic Dialects. Springer Publishers, expected Fall 2014

Journal Articles

- 2013**
1. Abdul-Mageed, M., **Diab, M.**, & Kuebler, S. (To appear, 2013). SAMAR: A System for Subjectivity and Sentiment Analysis for Arabic Social Media. *Computer Speech and Language*.
- 2012**
2. Abdul-Mageed, M., and **Diab, M.** (Submitted, 2012). Linguistically-motivated subjectivity and sentiment annotation and tagging of Modern Standard Arabic. *International Journal on Social Media MMM: Monitoring, Measurement, and Mining*.
- 2009**
3. BenAjiba, Yassine, **Mona Diab**, Paolo Rosso. (2009) Arabic Named Entity Recognition: A Feature-driven Study. *In the special issue on Processing Morphologically Rich Languages In IEEE Transactions on Audio, Speech and Language Processing*. '17(5), 926-934, Jul '09.
4. BenAjiba, Yassine, **Mona Diab**, Paolo Rosso. (2009) Using Language Independent and Language Specific Features to Enhance Arabic NER. *International Arabic Journal of Information Technology*. 6(5), 464-473, Jun '09.

1999

5. Resnik, Philip, Mari B Olsen, **Mona Diab**. (1999) The Bible as a Parallel Corpus: Annotating the 'Book of 2000 Tongues' *Computers and the Humanities*, 33 (1-2), 1999.

Invited Papers

2012

6. **Diab, Mona** and Yuval Marton. (2013) Semitic Semantics. Book Chapter in Natural Language Processing for Semitic Languages, Editor Imed Zitouni, Springer Publishers, forthcoming.

2011

7. **Diab, Mona**, Nizar Habash, Owen Rambow, Ryan Roth. (2011) CADIM Arabic Tools: Morphological Analysis, Disambiguation and Generation, Tokenization, Diacritization, Lemmatization, POS Tagging and Base Phrase Chunking. Book Chapter in the Handbook of Natural Language Processing and Machine Translation, Editors Joseph Olive, Caitlin Christianson and John McCary. Springer Publishers.

2010

8. Katz, Graham and **Mona Diab** (Guest Editors) (2010) Special Issue on Georgetown Round Table on Arabic Computational Linguistics Conference. ACM Transactions on Asian Language Information Processing (TALIP), December.

2008

9. **Diab, Mona**. (2008) It's About Time. *CMIC Newsletter*, Volume '1, issue #2. Feb 15th. <http://www.microsoft.com/middleeast/Egypt/CMIC/Newsletter.aspx>

2007

10. **Diab, Mona**, Kadri Hacioglu, Daniel Jurafsky. (2007) Automated Methods for Processing Arabic Text: From Tokenization to Base Phrase Chunking. In Arabic Computational Morphology: Knowledge-based and Empirical Methods. Editors Antal van den Bosch and Abdelhadi Soudi. Kluwer/Springer Publications.

Peer Reviewed Papers

2013

11. Hawwari, Abdelati, Wajdi Zaghoulani, Tim O'Gorman, **Mona Diab**, and Ahmed Badran. (2013) Building a Lexical Semantic Resource for Arabic Morphological Patterns. Proceedings of ICCSPA13, Sharjah, UAE. February 2013.

2012

12. Abdul-Mageed, Muhammad and **Mona Diab**. (2012) Toward Building a large scale Arabic Sentiment Lexicon. Proceedings of the 6th International Global WordNet Conference, January, Matsue, Japan.
13. Abdul-Mageed, Muhammad and **Mona Diab**. (2012) AWATIF: A multigenre corpus for MSA subjectivity and sentiment Analysis. Proceedings of LREC, May, Istanbul Turkey.
14. Abdul-Mageed, M., Kuebler, S. & **Diab, M.** (2012). SAMAR: A system for subjectivity and sentiment analysis of social media Arabic. In Proceedings of the 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA), held in conjunction with the 50th Annual Meeting on Association for Computational

Linguistics, July 12, ICC Jeju, Republic of Korea.

15. Abu Jabara, Amjad, Pradeep Dasigi, **Mona Diab**, Dragomir Radev. (2012) Subgroup Detection in Ideological Discussions. *Proceedings of ACL 2012*, July, Jeju Island, Korea.
16. Agirre, Eneko; Daniel Cer; **Mona Diab**; Aitor Gonzalez-Agirre. (2012) SemEval-2012 Task 6: A Pilot on Semantic Textual Similarity. In *Proceedings of SEMEVAL Workshop, held in conjunction with *SEM and NAACL 2012, Montreal, Canada, June 2012*.
17. Dasigi, P, Guo, W. & **Diab, M.** (2012). Cross-genre subgroup detection in online discussion threads: A study of the implicit attitude using textual latent semantics. In *Proceedings of ACL 2012*, July, Jeju Island, Korea.
18. Elfardy, Heba and **Mona Diab** (2012). Token Level Identification of Linguistic Code Switching. *Proceedings of COLING, Mumbai, India, December 2012*.
19. Elfardy, Heba and **Mona Diab** (2012). Simplified guidelines for the creation of Large Scale Dialectal Arabic Annotations. *Proceedings of LREC, Istanbul, Turkey, May 2012*
20. Elfardy, Heba and **Mona Diab** (2012). AIDA: Automatic Identification & Glossing of Dialectal Arabic. *Poster Session, Proceedings of European Association for Machine Translation (EAMT 2012)*, Trento Italy, May 2012.
21. Guo, Weiwei and **Mona Diab**. (2012) Modeling Semantic Textual Similarity in the Latent Space. *Proceedings of ACL 2012, Jeju Island, Korea, July 2012*.
22. Guo, W. & **Diab, M.** (2012). A simple unsupervised latent semantics based approach for sentence similarity. In *Proceedings of the 6th International Workshop on Semantic Evaluation (SEMEVAL 2012), in conjunction with *SEM and NAACL 2012, Montreal, Canada, June 2012*.
23. Guo, W. & **Diab, M.** (2012). Learning the latent semantics of a concept from its definition. In *Proceedings of ACL 2012, Jeju Island, Korea, July 2012*.
24. Habash, Nizar, **Mona Diab**, Owen Rambow (2012). CODA: A Conventional Orthography for Dialectal Arabic. *Proceedings of LREC, Istanbul Turkey, May 2012*.
25. Hawwari, Abdelati; Kfir Bar; **Mona Diab** (2012). Building an Arabic Multiword Expressions Repository. *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*, Montreal, Canada, June 2012.
26. Prabhakaran, Vinodkumar, Owen Rambow, **Mona Diab** (2012) Who's (Really) the Boss? Perception of Situational Power in Written Interactions. *Proceedings of COLING 2012, Mumbai, India, December 2012*.
27. Prabhakaran, Vinodkumar, Owen Rambow, **Mona Diab**. (2012) Annotations for Power Relations on Email Threads. *Proceedings of LREC, May, Istanbul Turkey*.
28. Prabhakaran, Vinodkumar, Owen Rambow, **Mona Diab**. (2012) Predicting Overt Display of Power in Written Dialogs. *Proceedings of NAACL 2012, Short Papers, Montreal, Canada, June 2012*.
29. Prabhakaran, Vinodkumar, Michael Bloodgood, **Mona Diab**, Bonnie J. Dorr, Lori Levin, Christine Piatko, Owen Rambow and Benjamin Van Durme. (2012) Statistical Modality Tagging from Rule-based Annotations and Crowdsourcing. In *Proceedings of ACL Workshop on Extra-propositional aspects of meaning in computational linguistics (ExProM)*. Jeju, South Korea, July, 2012.
30. Zaghouni, Wajdi, Abdelati Hawwari and **Mona Diab**. (2012) A Pilot PropBank Annotation for Quranic Arabic. In *Proceedings of the computational linguistics for literature workshop, NAACL 2012, Montreal, Canada, June 2012*.

2011

31. Abdul-Mageed, Muhammad, **Mona Diab**, Mohammed Korayem. (2011) Subjectivity

And Sentiment Analysis Of Modern Standard Arabic. *Proceedings of ACL, June, Portland, OR, USA.*

32. Abdul-Mageed, Muhammad and **Mona Diab**. (2011) Subjectivity and Sentiment Annotation of Modern Standard Arabic Newswire. *Proceedings of ACL Workshop LAW V, June, Portland, OR, USA.*
33. Dasigi, Pradeep and **Mona Diab**. (2011) Noisy Channel Modeling for Named Entity Transliteration. *Proceedings of NEWS 2011 Workshop, IJCNLP 2011, November, Chiang Mai, Thailand.*
34. **Diab, Mona** and Pradeep Dasigi (2011) CODACT: Towards Identifying Orthographic Variants in Dialectal Arabic. *Proceedings of IJCNLP 2011, November, Chiang Mai, Thailand.*
35. **Diab, Mona** and Ankit Kamboj (2011) An Annotation Scheme Leveraging Crowd Sourcing for Linguistic Code Switching in Hinglish. *Proceedings of ALR9 Workshop, IJCNLP 2011, November, Chiang Mai, Thailand.*
36. Ghoneim, Mahmoud, Marine Carpuat and **Mona Diab**. (2011) Linguistically motivated reordering constraints for phrase-based SMT: base phrase chunks and predicate argument structures. *In Proceedings of the Eleventh Conference on Language Engineering ESOLEC11, Cairo, Egypt, 2011.*
37. Guo, Weiwei and **Mona Diab**. (2011) Semantic Topic Models: Combining Word Distributional Statistics and Dictionary Definitions. *Proceedings of EMNLP 2011, July, Edinburgh, Scotland, UK.*

2010

38. Carpuat, Marine, and **Mona Diab**. (2010) Task-based Evaluation of Multiword Expressions: a Pilot Study in Statistical Machine Translation. *Proceedings of NAACL, June, Los Angeles, CA, USA.*
39. **Diab, Mona**, Nizar Habash, Owen Rambow, Mohamed Al Tantawy, Yassine Benajiba. (2010) COLABA: Arabic Dialect Annotation and Processing. *Proceedings of the Workshop on Semitic Language Processing, LREC, May, Malta.*
40. Benajiba, Yassine and **Mona Diab**. (2010) A Web Application for Dialectal Arabic Text Annotation. *Proceedings of the Workshop on Semitic Language Processing, LREC, May, Malta.*
41. Benajiba Yassine, Imed Zitouni, **Mona Diab**, Paolo Rosso. (2010) Arabic Named Entity Recognition: Using Features Extracted from Noisy Data. *Proceedings of ACL, July, Uppsala, Sweden.*
42. Guo, Weiwei and **Mona Diab**. (2010) Combining Orthogonal Monolingual and Multilingual Sources of Evidence for All Words WSD. *Proceedings of ACL, July, Uppsala, Sweden.*
43. Zaghoulani, Wajdi, **Mona Diab**, Aous Mansouri, Sameer Pradhan, Martha Palmer. (2010) The Revised Arabic PropBank. *Proceedings of ACL Linguistic Annotation Workshop (LAW) IV, July, Uppsala, Sweden.*
44. Guo, Weiwei and **Mona Diab**. (2010) COLEPL and COLSLM: An Unsupervised WSD Approach to Multilingual Lexical Substitution, Tasks 2 and 3 SemEval 2010. *Proceedings of ACL SEMEVAL, July, Uppsala, Sweden.*
45. Prabhakaran, Vinodkumar, Owen Rambow, **Mona Diab**. (2010) Automatic Committed Belief Tagging. *Proceedings of COLING 2010, August, Beijing, China.*

2009

46. Kristen Parton, Kathleen R. McKeown, Robert Coyne, **Mona T. Diab**, Ralph Grishman, Dilek Hakkani-Tür, Mary Harper, Heng Ji, Wei Yun Ma, Adam Meyers, Sara Stolbach, Ang Sun, Gokhan Tur, Wei Xu and Sibel Yaman Who. (2009) What, When, Where, Why? Comparing Multiple Approaches to the Cross-Lingual 5W Task.

Proceedings of ACL-IJCNLP, August, Singapore.

47. **Diab, Mona**, and Pravin Bhutada. (2009) Verb Noun Construction MWE Supervised Token Classification. *Proceedings of ACL-IJCNLP 2009 Workshop on Multiword Expressions: Identification, Interpretation, Disambiguation and Applications, ACL-IJCNLP, August, Singapore.*
48. **Diab, Mona**, Lori Levin, Teruko Mitamura, Owen Rambow, Vinodkumar Prabhakaran, Weiwei Guo. (2009) Committed Belief Annotation and Tagging. *Proceedings of ACL-IJCNLP 2009 Workshop on The Third Linguistic Annotation Workshop (The LAW III), ACL-IJCNLP, August, Singapore.*
49. Guo, Weiwei, and **Mona Diab**. (2009) Improvements to Monolingual English Word Sense Disambiguation. *Proceedings of the NAACL 2009 workshop: Semantic Evaluations: Recent Achievements and Future Directions (SEW-2009), NAACL, June, Boulder, CO, USA.*
50. **Diab, Mona**. (2009) Second Generation Tools (AMIRA 2.0): Fast and Robust Tokenization, POS tagging, and Base Phrase Chunking. *MEDAR 2nd International Conference on Arabic Language Resources and Tools, April, Cairo, Egypt.*
51. **Diab, Mona** and Madhav Krishna. (2009) Unsupervised classification of Verb Noun Multiword Expression Tokens. *CICLING, March, Mexico City, Mexico.*
52. **Diab, Mona** and Madhav Krishna. (2009) Handling Sparsity for Unsupervised MWE Classification. *GEMS Workshop, EACL, March, Athens, Greece.*

2008

53. BenAjiba, Yassine, **Mona Diab**, Paolo Rosso. (2008) Arabic Named Entity Recognition: An SVM Based Approach. *International Arab Conference on Information Technology (ACIT'2008), December, University of Sfax, Tunisia.*
54. BenAjiba, Yassine, **Mona Diab**, Paolo Rosso, (2008). Arabic Named Entity Recognition using Optimized Feature Sets. *In Proceedings of Empirical Methods for Natural Language Processing (EMNLP), Oct, Hawaii, HI, USA.*
55. BenAjiba, Yassine, **Mona Diab**, Paolo Rosso. (2008). Arabic Named Entity Recognition using Conditional Random Fields. *Proceedings of Workshop on HLT & NLP within the Arabic world. Arabic Language and local languages processing: Status Updates and Prospects, 6th Int. Conf. on Language Resources and Evaluation, LREC, May, Marrakech, Morocco.*
56. **Diab, Mona**, Alessandro Moschitti, Daniele Pighin, (2008) Semantic Role Labeling Systems for Arabic Language using Kernel Methods *In Proceedings of Association for Computational Linguistics (ACL), July, Columbus, OH, USA.*
57. Habash, Nizar, Owen Rambow, **Mona Diab**, Reem Farraj. (2008) Guidelines for Annotating Arabic Dialect, *Proceedings of Workshop on Arabic and its local languages, LREC, May, Marrakech, Morocco.*
58. Palmer, Martha, Ann Bies, Olga Babko-Malaya, **Mona Diab**, Mohamed Maamouri, Aous Mansouri, Wajdi Zaghouni. (2008) A pilot Arabic Propbank *In Proceedings of LREC, May, Marrakech, Morocco.*
59. Roth, Ryan, Owen Rambow, Nizar Habash, **Mona Diab**, Cynthia Rudin. (2008) Arabic Morphological Tagging, Diacritization, and Lemmatization Using Lexeme Models and Feature Ranking. *In Proceedings of Association for Computational Linguistics (ACL), July, Columbus, OH, USA.*

2007

60. **Diab, Mona**. (2007) Towards an optimal POS tag set for Modern Standard Arabic Processing *Recent Advances in Natural Language Processing (RANLP), August, Borovets, Bulgaria.*
61. **Diab, Mona**. (2007) Improved Arabic Base Phrase Chunking with a new enriched POS tag set. *Proceedings of ACL Workshop on Computational Approaches to*

Semitic Languages (CASL), ACL '07, July, Prague, Czech Republic.

62. **Diab, Mona**, Musa Alkhalifa, Sabry ElKateb, Christiane Fellbaum, Aous Mansouri, Martha Palmer. (2007) SemEval-2007 Task '18: Arabic Semantic Labeling. *Proceedings of ACL Semantic Evaluation Workshop (SEMEVAL), ACL '07, July, Prague, Czech Republic.*
63. **Diab, Mona**, Alessandro Moschitti, Daniele Pighin. (2007) CUNIT: A Semantic Role Labeling System for Modern Standard Arabic. *Proceedings of ACL Semantic Evaluation Workshop (SEMEVAL), ACL '07, July, Prague, Czech Republic.*
64. **Diab, Mona** and Alessandro Moschitti. (2007) Semantic Parsing for Modern Standard Arabic. *Recent Advances in Natural Language Processing (RANLP), August, Borovets, Bulgaria.*
65. **Diab, Mona**, Mahmoud Ghoneim, Nizar Habash. (2007) Arabic Diacritization in the Context of Statistical Machine Translation. *Proceedings of Machine Translation Summit (MT-Summit), September, Copenhagen, Denmark.*
66. Kirchhoff, Katrin, Owen Rambow, Nizar Habash, **Mona Diab**. (2007) Semi-Automatic Error Analysis for Large-Scale Statistical Machine Translation Systems. *Proceedings of Machine Translation Summit (MT-Summit), September, Copenhagen, Denmark.*

2006

67. **Diab, Mona**. (2006) Developing a Mentoring Culture in Egypt. *Position paper at the Conference for the Association of Egyptian American Scholars (AEAS), National Research Council of Egypt, December, Cairo, Egypt.*
68. **Diab, Mona**, and Neal Snider. (2006) Modern Standard Arabic Verb Alternations. *Annual Meeting of the Arabic Linguistics Society (ALS) Conference, March, Kalamazoo, USA.*
69. Chiang, David, **Mona Diab**, Nizar Habash, Owen Rambow, Safi-Allah Sharif. (2006) Parsing Arabic Dialects. *In Proceedings of the '11th Conference of the European Chapter of the Association for Computational Linguistics (EACL), April, Trento, Italy.*
70. Snider, Neal and **Mona Diab**. (2006) Automatic Discovery of Verb Classes in Modern Standard Arabic. *In Proceedings of the Joint Conference of the International Committee on Computational Linguistics and the Association for Computational Linguistics (ACL-COLING'06), July, Sydney, Australia.*
71. Snider, Neal and, **Mona Diab**. (2006) Unsupervised Induction of Arabic Verb Classes. *In Proceedings of Human Language Technology Conference /North American chapter of the Association for Computational Linguistics Annual Meeting (HLT-NAACL'06). June, New York, NY, USA.*
72. Maamouri, Mohamed, Ann Bies, Tim Buckwalter, **Mona Diab**, Nizar Habash, Owen Rambow, Dalila Tabessi. (2006) Developing and Using a Pilot Dialectal Arabic Treebank. *In Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06), May, Genoa, Italy.*

2005

73. Darwish, Kareem, **Mona Diab**, Nizar Habash, Editors. (2005) Computational Approaches to Semitic Languages. *Proceedings of ACL Workshop on Computational Approaches to Semitic Languages (CASL), June, Ann Arbor, Michigan, MI, USA.*

2004

74. **Diab, Mona**. (2004) Relieving the Data Acquisition Bottleneck for Word Sense Disambiguation. *Proceedings of Association for Computational Linguistics (ACL).*
75. **Diab, Mona**, Kadri Hacioglu, Daniel Jurafsky. (2004) Automatic Tagging of Arabic Text: From Raw Text to Base Phrase Chunks. *Proceedings of Human Language Technology-North American Association for Computational Linguistics (HLT-NAACL).*
76. **Diab, Mona**. (2004) An Unsupervised Approach for Bootstrapping Arabic Sense

- Tagging. *Proceedings of Arabic Workshop, Computational Linguistics (COLING)*.
77. **Diab, Mona.** (2004) Bootstrapping a WordNet Taxonomy for Arabic, *Proceedings of First Arabic Language Technology Conference (NEMLAR), September, Cairo, Egypt.*
2003
78. **Diab, Mona.** (2003) Word Sense Disambiguation within a multilingual framework. *PhD Thesis, University of Maryland, College Park, August, MD, USA.*
2002
79. **Diab, Mona** and Philip Resnik. (2002) An Unsupervised Method for Word Sense Tagging using Parallel Corpora. *Proceedings of Association for Computational Linguistics (ACL), July, Philadelphia, PA, USA.*
2000
80. **Diab, Mona.** (2000) An Unsupervised Method for Word Sense Tagging using Parallel Corpora: A Preliminary Investigation, *Special Interest Group in Lexical Semantics (SIGLEX) Workshop, Proceedings of ACL, October, Hong Kong.*
81. **Diab, Mona.** (2000) Exploiting Translations for Semantic Annotation. *Candidacy Exam Thesis, Linguistics Department, University of Maryland, College Park, June, MD, USA.*
82. **Diab, Mona** and Steven Finch. (2000) A Statistical Word-Level Translation Model for Comparable Corpora, *Proceedings of Conference on Content-based Multimedia Information Access (RIAO'00), May, Paris, France.*
83. Resnik, Philip and **Mona Diab.** (2000) Measuring Verb Similarity, *Proceedings of the conference of International Cognitive Science Society (COGSCI'00), July, Philadelphia, PA, USA.*
84. Dorr, Bonnie, Gina Levow, Douglas Oard, Philip Resnik, Amy Weinberg, **Mona Diab**, Maria Katsova. (2000) MADLIBS: An Event Translingual Lexical Conceptual Structure Based Information Retrieval System. *Proceedings of North American Association for Computational Linguistics (NAACL), June, Seattle, WA, USA.*
1998
85. **Diab, Mona**, John Schuster, Peter Bock. (1998) A Preliminary Statistical Investigation into the impact of an N-Gram Analysis Approach based on Word Syntactic Categories toward Text Author Classification. *Proceedings of 6th International Conference on Artificial Intelligence & Applications, December, Cairo, Egypt.*
86. Riopka, Terry, **Mona Diab**, Peter Bock. (1998) Quantifying and Interpreting the Effect of Intelligent Information Exchange. *Proceedings of 6th International Conference on Artificial Intelligence & Applications, December, Cairo, Egypt.*
1997
87. Resnik, Philip, Mari Broman Olsen, **Mona Diab.** (1997) Creating a Parallel Corpus from the Book of 2000 Tongues, *Text Encoding Initiative '10th Anniversary User Conference (TEI-10), November, Providence, RI, USA.*

Technical Reports

88. Habash, Nizar, **Mona Diab**, Owen Rambow. (2011) Conventional Orthography for Dialectal Arabic (CODA) V.1.0. *Technical Report '137382, <http://academiccommons.columbia.edu/catalog/ac:137382>, Columbia University, New York, NY, USA*
2005
89. Rambow, Owen, David Chiang, **Mona Diab**, Nizar Habash, Rebecca Hwa, Khalil

Sima'an, Vincent Lacey, Roger Levy, Carol Nichols, and Safi-Allah Shareef. (2005) Parsing Arabic Dialects. Final Report, JHU Summer Workshop, December, Baltimore, MD, USA.

2000

90. Riopka, Terry, **Mona Diab**, Peter Bock. (2000) Quantifying and Interpreting the Effect of Intelligent Information Exchange Between Chromosomes in a Human Simulation of a Genetic Algorithm. Technical Report: LAMP-TR-045/UMIACS-TR-2000-38/CS-TR-4147, University of Maryland College Park, June, College Park, MD, USA.
91. Resnik, Philip and **Mona Diab**. (2000) Measuring Verb Similarity. Technical Report: LAMP-TR-047/ UMIACS-TR-2000-40/CS-TR-4149, University of Maryland College Park, June, College Park, MD, USA.
92. **Diab, Mona** and Steve Finch. (2000) A Statistical Word-Level Translation Model for Comparable Corpora. Technical Report: LAMP-TR-048/ UMIACS-TR-2000-41/CS-TR-4150, University of Maryland College Park, June, College Park, MD, USA.
93. **Diab, Mona**, John Schuster, Peter Bock. (2000) A Preliminary Statistical Investigation into the Impact of an N-Gram Analysis Approach Based on Word Syntactic Categories Toward Text Author Classification. Technical Report: LAMP-TR-046/UMIACS-TR-2000-39/CS-TR-4148, University of Maryland College Park, June, College Park, MD, USA.
94. **Diab, Mona**, Douglas W. Oard, Tapas Kanungo, Mark Ames, Russell Ball. (2000) Using Arabic Examples to find English Documents. Technical Report, University of Maryland College Park, June, College Park, MD, USA.

Tutorials

2006

95. **Diab, Mona** and Nizar Habash. Arabic Dialect Processing.

2005

96. **Diab, Mona**. Statistical NLP: An A-Z Crash Course

PATENTS

- **OPTDIAC**: An Optimal Diacritization System for Arabic Writing. Columbia University, Submitted, Aug '09/11
- **ROAH**: Automatic Talking Heads Leveraging Traditional & Social Media. Columbia University, Submitted, Aug '11

RELEASED SOFTWARE

- **Diab, Mona**, Heba Elfardy, Yassine Benajiba. (2009-2011) **AIDA** Automatic Identification of Arabic Dialectal Text. A Tool for Dialect Identification & Classification, Named Entity Recognition, English and Modern Standard Arabic Glossing and Normalization. Current Version '1.0. <https://nlp.ccls.columbia.edu/AIDA>
- **Diab, Mona** and Pradeep Dasigi. (2011) **CODACT** A Tool for Arabic Dialect Orthographic Conventionalization. Current Version '1.0. <https://nlp.ccls.columbia.edu/CODACT>
- Benajiba, Yassine and **Mona Diab**. (2008-2010) **NIDA** Arabic Named Entity Recognition. Current Version 2.0. <https://nlp.ccls.columbia.edu/NIDA>

- **Diab, Mona**, Yassine Benajiba, Heba Elfardy. (2005-2011) **AMIRA** Automatic Segmentation/Tokenization, Lemmatization, Diacritization, POS Tagging, Base Phrase Chunking/Shallow Parsing of Arabic Text. Current Version 2.0. <https://nlp.ccls.columbia.edu/AMIRA>
- Benajiba, Yassine, **Mona Diab**, Heba Elfardy. (2009-2011) **COLANN** COLABA Annotation Interface v.1.0: An annotation interface for Arabic Dialects. Current Version '1.0. <https://nlp.ccls.columbia.edu/COLANN>
- **Diab, Mona** and Weiwei Guo. (2007-2009) **SALAAM-CONT** Automatic Multilingual Online All words Word Sense Disambiguation. Current Version '1.0. <https://nlp.ccls.columbia.edu/COLANN>
- Prabhakaran, Vinodkumar, **Mona Diab**, Owen Rambow. (2009-2010) **COMBEL** Automatic Committed Belief Tagging for Running Text. Current Version '1.0. <https://nlp.ccls.columbia.edu/COMBEL>

PRESENTATIONS & POSTERS **(Not Associated with Publications)**

2011

- *Power Shifts & Rifts in Social Communications*. (with Kathleen McKeown, Owen Rambow, Dragomir Radev). Invited Presentation, **IARPA SCIL PI Meeting**. Washington DC, USA, Nov '11.
- *Arabic Dialect Processing under BOLT*. (with Nizar Habash and Owen Rambow). Invited Presentation, **DARPA BOLT Kick-off Meeting**, Tampa Florida, USA, Oct '11.
- *COLABA Tools and Resources*. (with Nizar Habash and Owen Rambow). Invited Presentation, **DARPA BOLT Kick-off Meeting**, Tampa Florida, USA, Oct '11.
- *Arabic Subjectivity and Sentiment Analysis: System & Resources*. Invited Poster, **DARPA BOLT Kick-off Meeting**, Tampa Florida, USA, Oct '11.
- *CADIM Arabic MSA NLP Tools*. (with Nizar Habash and Owen Rambow). Invited Poster, **DARPA BOLT Kick-off Meeting**, Tampa Florida, USA, Oct '11.
- *Conventionalized Dialectal Arabic (CODA): Guidelines and Tools*. (with Nizar Habash and Owen Rambow). Invited Poster, **DARPA BOLT Kick-off Meeting**, Tampa Florida, USA, Oct '11.
- *COLABA Tools and Resources*. (with Nizar Habash and Owen Rambow). Invited Presentation, **Brigham Young University**, Utah, USA, Aug '11.
- *Arabic Language Modeling for OCR*. (with Nizar Habash). Invited Presentation, **DARPA MADCAT PI Meeting**. Boston, MA, USA, Jul '11.
- *Power Shifts & Rifts in Social Communications*. (with Kathleen McKeown, Owen Rambow, Dragomir Radev). Invited Presentation, **IARPA SCIL PI Meeting**. Washington DC, USA, May '11.
- *Multilayered Semantic Annotation and challenges of Interoperability*. Invited Presentation, **NSF-funded SILT/FLaReNet Forum on Sustainable Interoperability for Language Technologies**. Brandeis University, MA, USA, Apr '11.

2010

- *COLABA Annotation, Tools, Resources*. (with Nizar Habash and Owen Rambow). Invited Presentation, **Government Day AMTA 2010**. Denver, CO, USA, Oct '10.
- *Arabic Language Modeling for OCR*. (with Nizar Habash). Invited Presentation, **DARPA MADCAT PI Meeting**. Boston, MA, USA, Sep '10.
- *Power Shifts & Rifts*. (with Kathleen McKeown, Owen Rambow, Dragomir Radev). Invited Presentation, **IARPA SCIL PI Meeting**. Washington DC, USA, May '10.

- *Arabic Dialect Processing within the Context of Social Media*. (with Nizar Habash and Owen Rambow). Invited Presentation, **COLABA Review Day, MITRE Corporation**. McLean, VA, USA, Apr '10.

2009

- *Arabic Language Modeling for OCR*. (with Nizar Habash). Invited Presentation, **DARPA MADCAT PI Meeting**. Boston, MA, USA, Sep '09.
- *Power Shifts & Rifts*. (with Kathleen McKeown, Owen Rambow, Dragomir Radev). Invited Presentation, **IARPA SCIL PI Meeting**. Washington DC, USA, Oct '09.

2008

- *SALAMAT: Sense Assignment Leveraging Alignments, Monolingual And Translation*. (with Weiwei Guo). Invited Demo/Poster, **NSF sponsored symposium on Semantic Knowledge Discovery, Organization and Use, NYU**. New York, NY, USA, Nov '08.
- *Columbia CCLS: Committed Belief and Dialog Acts*. (with Owen Rambow and Becky Passaneau). Invited Presentation, **Human Language Technology Center of Excellence, University of Maryland**. College Park, MD, USA, Sep '08.
- *Tools for Arabic Morphological Analysis, Disambiguation and Tokenization*. (with Nizar Habash and Owen Rambow). Invited Presentation, **Technology Forum, DIA and TSWG**. McLean, VA, Jul '08.
- *CADIM: Natural Language Processing for Arabic and its dialects*. (with Nizar Habash and Owen Rambow). Invited Presentation, **CASL, University of Maryland**. College Park, MD, USA, Jun '08.
- *Factivity Annotations*. (with Owen Rambow, Lori Levin, and Teriku Mitamura). Invited Presentation, **JHU Human Language Technology Center of Excellence, University of Maryland**. College Park, MD, USA, May '08.

2007

- *Impact of Partial Arabic Diacritization on Statistical Machine Translation*. (with Mahmoud Ghoneim and Nizar Habash). Invited Presentation, **DARPA GALE PI Meeting**. San Francisco, CA, USA, May '07.
- *Semi-Automatic Error Analysis of the NIGHTINGALE Machine Translation System*. (with Katrin Kirchhoff, Nizar Habash, Owen Rambow, Evgeny Matusov). Invited Presentation, **DARPA GALE PI Meeting**. San Francisco, CA, USA, May '07.

2006

- *Modern Standard Arabic Verb Alternations*. (with Neal Snider). Paper presented at the **Annual Meeting of the Arabic Linguistics Society (ALS) Conference**. Kalamazoo, MI, USA, Mar '06.
- *NLP Tools for Arabic*. (with Nizar Habash and Owen Rambow). Invited Presentation, **DARPA GALE PI Meeting**. Boston, MA, USA, May '06.
- *Developing a Mentoring Culture in Egypt*. Position paper presentation at the **Conference for the Association of Egyptian American Scholars, National Research Council of Egypt**. Cairo, Egypt. Dec '06.

2005

- *Treebank Transduction. Arabic Dialect Parsing*. (with Owen Rambow, Nizar Habash, Khalil Simaan, David Chiang, Rebecca Hwa, Roger Levy, Vincent Lacey, Saif-Allah Shareef). **Johns Hopkins Summer Workshop Final Presentation Symposium**. Baltimore, MD, USA, Aug '05.

2001

- *An Unsupervised method for Word Sense Tagging using Parallel corpora*. Poster Session: **UMIACS Research Day, University of Maryland**. College Park, MD, USA, Mar '01.

2000

- *Statistical Methods and Linguistics*. (with Elixabete Murguia). **LingFest, Linguistics Department University of Maryland**. College Park, MD, USA, Nov '00.
- *An Unsupervised Method for Word Sense Tagging using Parallel corpora*. **North East Scholars NLP workshop: IBM T.J. Watson Research Center, Hawthorne, NY, USA**, Oct '00.
- *MADLIBS: An LCS based Translingual Information Retrieval System*. (with Maria Katsova). **Demonstration: NAACL'00**. Seattle, WA, USA, May '00.
- *A Statistical Model of Word-Level Mapping for Comparable Corpora*. **UMIACS Computational Linguistics Colloquium, University of Maryland**. College Park, MD, USA, Spr '00.

1999

- *Word Sense Disambiguation: Lexical Ambiguity Resolution within a Selection Preference Model*. **Linguistics Student Conference, Linguistics Dept., University of Maryland**. College Park, MD, USA, Fall '99.
- *Cross Language Information Retrieval: An LCS based approach for Spanish queries and English Documents*. (with Maria Katsova). **Poster Session: UMIACS Research Day, University of Maryland**. College Park, MD, USA, Feb '99.

1997

- *A Preliminary Statistical Investigation into the Impact of Using an N-Gram Analysis for the Purposes of Text Author Identification (Author Classification)*. **Linguistics Student Conference, Linguistics Dept., University of Maryland**. College Park, MD, USA. Fall '97.

PROFESSIONAL ACTIVITIES

Academic Services within Columbia University

- Member of **Columbia Senate Committee for Research Officers**. *Fall '05–present*.
 - The committee represents a ~3k constituency of Research Officers at Columbia University including: Professional Research Officers, Post-Docs, and Research Staff. It comprises '9 members.
 - Leading the effort on the creation of professorial titles for Professional Research Officers who serve as independent PIs and advisors of record for students.
- Member of the **Columbia Senate Committee on the External Affairs and Public Relations**. *Fall '11–present*.
- Member of the **Columbia Senate Commission on the Status of Women**. *Spr '10–present*.
- Member of **SEAS Dean's Committee for Research Scientists**. *Fall '09–present*.
- Member of Columbia Senate committee on University Rules of Conduct *Spr '08–Fall '09*.
- Member of **Columbia Senate committee on Alumni Relations**. *Spr '08–Fall '09*.
- Member of **CCLS Strategy Committee**. *Fall '06–present*.
- Organizer of the **CCLS-Yahoo Distinguished Lecture Series**. *Spr '10–present*
- Organizer of **CCLS Colloquium Series**. *Fall '07– Spr '08*.
- Represented CCLS in the **SEAS Executive Council Meetings**. *Fall '07–Spr '09*.
- Representing Columbia University **Bi-annual National Council on Women in Information Technology (NCWIT)** meetings. *Fall '05–present*.

- Member of **Columbia Senate Committee on Physical Development**. *Fall '05–Spr '06*.
- Organizer of **Stanford Computational Linguistics Colloquium Series**. *Spr '04–Fall '04*.

Academic Services within NLP Community

Reviewer/Editor

- Editorial Board Member for Journal of Language Resources and Evaluation **Spr '10 – present**
- Guest Editor of Special Issue on Georgetown Round Table on Arabic Computational Linguistics Conference. ACM Transactions on Asian Language Information Processing **Oct '09 – Mar '10**

Grant Proposal Reviewer

- KACST Research Fund **Jan '12 –present**
- NSF OISE Program **Feb '12**
- Qatar National Research Fund **Spr '10 – present**
- NSF Linguistics Program **Apr '11**
- NSF SoCS Program Panel **Jan '10**
- NSF HSD Awards Panel **May '08**
- NSF Career Awards **Fall '05**

Conference and Workshop Organization

- DARPA Funded Unified Semantic Framework for Textual Similarity Workshop **Feb '12**
- Area Co-Chair for ACL 2012 Multilinguality **Aug '11 –Jul '12**
- Program Co-Chair of '1st International Joint Conference on Semantics *SEM 2012 **Aug '11 –Jun'12**
- Workshops Co-Chair, NAACL 2012 **Jun '11- Jun '12**
- Co-Organizer SEMEVAL Task on Multilayer Semantic Annotation and STS **Jul '11 – Jun '13**
- Chair and co-organizer of 2nd NSF Funded Workshop on Linguistic Code Switching **Dec '10**
- Chair and co-organizer of NSF Funded Workshop on Linguistic Code Switching **Jun '10**
- Area Chair for NAACL Semantics and Machine Learning **Feb '10**
- Area Chair for EMNLP Multilinguality **May '09**
- Co-Chair COLING-HLT Workshop on TextGraphs **Aug '08**
- Co-Chair LREC Workshop on Arabic and Local Languages **May '08**
- Chair and co-organizer of NSF Funded Workshop on Linguistics Code Switching **Apr '08**
- Organizer of SEMEVAL '1 Task '18 on Arabic Semantic Labeling **Jun '07**
- Co-Chair ACL Workshop on Computational Approaches to Semitic Languages **Jun '05**

PROGRAM COMMITTEE MEMBER

▪ **Journals**

Computational Linguistics; IEEE Speech Communication; Computers and Humanities; Computers, Speech and Language; Journal of Natural Language Engineering; Egyptian Journal of Language Engineering; Journal for Language Resources and Evaluation; ACM Transactions on Speech and Language Processing; Journal of Artificial Intelligence Research; Journal of Cognitive Science; ACM Transactions on Asian Language Information Processing (TALIP); Machine Translation

▪ **Conferences**

Association for Computational Linguistics (ACL); Cognitive Science Society; Computational Linguistics (COLING); NEMLAR; Human Language technologies (HLT); North American Association for Computational Linguistics (NAACL); Empirical Methods in Natural Language Processing (EMNLP); Egyptian Conference on Language Engineering; European Association for Computational Linguistics (EACL); IJCAI; IJCNLP; RANLP; LREC; MT Summit; AMTA; CICLING, ALTIC; AIRS

▪ **Workshops**

MWE (ACL 2011); Text Graphs (ACL 2011); LAWS V (ACL 2011); LREC Workshop on HLT for Semitic Languages (LREC 2010); Statistical Machine Translation (ACL 2008); TextGraphs (COLING 2008); Student Research Workshop (ACL 2008); Arabic and Local Languages (LREC 2008); Computational Approaches to Semitic Languages (ACL 2007); SEMEVAL '1 (ACL 2007); Computational Approaches to Arabic Script based Languages (2007); Student Research Workshop (HLT-NAACL 2007); Cross Language knowledge Induction (EACL 2006); Student Research Workshop (ACL 2005); Parallel Text (ACL 2005); Computational Approaches to Semitic Languages (ACL 2005); Cross-Language Knowledge Induction Workshop (EUROLAN 2005); Computational Approaches to Arabic Script based Languages (COLING 2004)

PROFESSIONAL MEMBERSHIPS

- Association for Computational Linguistics
- Cognitive Science Society
- American Association for Artificial Intelligence
- Institute for Electrical & Electronics Engineers
- Society of Women Engineers
- National Electrical and Computer Engineering honor society (HKN)

COMPUTER SKILLS

- Extensive software development experience under Windows and Unix and Internet web design.
- Extensive Experience developing sophisticated algorithms in AI using state of the art Machine Learning technologies.
- Languages and Tools: Perl, C, Python, Java, C++, Pascal, Prolog, Basic, Assembly, MatLab, R, WEKA

LANGUAGE SKILLS

- ***Arabic, English***: native proficiency

- **French:** very good

REFERENCES

- Available upon request

SIGNATURE

A handwritten signature in blue ink, appearing to read "H. M. J. Fine", written in a cursive style.