

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

EDUCATION

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

ACADEMIC TRAINING

- **Postdoctoral Research Scientist** **Jan 04 - Jan 05**
Linguistics Department & Natural Language Processing Group
Stanford University, Palo Alto CA, USA
Mentor: Daniel Jurafsky. (MacArthur Fellow 02-03)
- **Research Associate** **Jul 03 - Dec 03**
Center for Spoken Language Research
University of Colorado Boulder, Boulder CO, USA
Mentor: Daniel Jurafsky. (MacArthur Fellow 02-03)
- **Graduate Research Assistant** **Jun 97 - Jun 03**
Institute for Advanced Computer Studies. (UMIACS)
University of Maryland College Park,
College Park MD, USA
Advisor: Philip Resnik
- **Graduate Teaching Instructor** **Fall 95 - Spr 97**
Department of Computer Science
The George Washington University, Washington DC, USA

ACADEMIC POSITIONS

- **Associate Professor** **Jan 13 - present**
Department of Computer Science
School of Engineering & Applied Science

- George Washington University. (GWU)
Washington DC, USA*
- **Director of CARE4Lang NLP Lab** **Jan 13 - present**
Comprising 20 members
Graduate Students, Postdocs, Research Scientists & Engineers, and Visiting scholars
*Department of Computer Science
School of Engineering & Applied Science
George Washington University. (GWU)
Washington DC, USA*
 - **Co-Director of CADIM Group** **Nov 05 -present**
Comprising 15 members
Joint Collaboration between GWU, Columbia University, NYU-Abu Dhabi
 - **Research Scientist (PI)** **Feb 09 – Dec 12**
*Center for Computational Learning Systems (CCLS)
Columbia University
New York NY, USA*
 - **Adjunct Associate Professor** **Feb 09 – Dec 12**
*Department of Computer Science
Columbia University
New York NY, USA*
 - **Associate Research Scientist (PI)** **Feb 05 – Jan 09**
*Center for Computational Learning Systems (CCLS)
Columbia University
New York NY, USA*
 - **Adjunct Assistant Professor** **Sep 07 – Jan 09**
*Department of Computer Science
Columbia University
New York NY, USA*

PROFESSIONAL EXPERIENCE

- **Visiting Researcher** **Sep 02 – Dec 02**
*Laboratoire Perception, Systemes, et Information
Université de Rouen
Rouen, France
Host: 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*
- **Software Engineer** **Apr 93 – Feb 95**
*Regional Inform. Tech. Software Engineering Center. (RITSEC)
United Nation Development Program
Cairo, Egypt*
- **Licensed Tourist Guide** **Jun 91 – Jun 95**

Free Lance

Cairo, Egypt

- **English News Broadcast Anchor** Jul 92 - Sep 92
*Saudi Arabia National Radio
Riyadh, Saudi Arabia*

BOARD MEMBERSHIP

- **Journal of Computer Speech and Language** Apr 16 - present
Editorial Board Member
- **Journal of Language Resources and Evaluation** Oct 11 – present
Editorial Board Member
- **Springer Publisher: NLP Briefs** Jul 11 – present
Editorial Board Co-Chair
- **Journal of Computational Linguistics** Jul 13 – present
Editorial Board Member
- **Qordoba Inc.** Sep 11 – Sep 13
*Advisory Board Member
Dubai, United Arab Emirates*
- **(ISO) Semantic Annotation Framework –Named Entities** Sep 11 – Sep 12
*Advisory Board Member
International Standards Organization*
- **Cairo Microsoft Research Innovation Center (CMIC)** May 10 – Jun 11
*Advisory Board Member
One of 6 Microsoft Research Innovation Centers World Wide
Cairo, Egypt*
- **Human Language Technology Center of Excellence, JHU** Jun 09 – Aug 13
*Editorial Board Member
SCALE
Johns Hopkins University
Baltimore MD, USA*

ELECTED POSITIONS

- **GWU Elected Faculty Senator** Apr 14 – Fall 15
- **President of ACL Special Interest Group on the Lexicon. (SIGLEX)** Oct 13 – Sep 16
- **President of ACL Special Interest Group on Semitic Languages. (SIG Semitic)** Nov 11 – Oct 16
- **NAACL Executive Board Member** Jan 12 – Jun 14
- **Columbia University Senator** Aug 09 – Dec 12
One of 6 elected members campus-wide representing >3000 Research Officers
- **Secretary of ACL SIGLEX** Jun 10 – Sep 13
- **Secretary of ACL SIG Semitic** Jun 09 – Oct 10
- **Executive Board Member of ACL SIGLEX** Oct 07 – Jul 10

ACADEMIC INITIATIVES

- **Data Science Track within Computer Science Dept. (GWU)** May 14
Joint effort with other Faculty members in Computer Science Department targeting both undergraduate and graduate students
- ***SEM: First Comprehensive Semantics Conference** Jul 11

Founding Member. (With Eneko Agirre, Johan Bos)

*Creating a consolidated Unified Top Tier Semantics Conference. (current acceptance rate 31% for *SEM 1 and *SEM 2, on par with NAACL and EACL, now in its 5th edition)*

- **CALIMAT: Columbia Arabic Language Institute for Modeling and Technology** **Jul 07**

Founding Member. (With Hatim Diab, Nizar Habash, Owen Rambow, David Waltz)

Seeking Joint Institutional Collaborations with the Arab World

Columbia University, New York NY, USA

- **CADIM Group** **Jun 05**

Founding Member. (With Nizar Habash and Owen Rambow)

Columbia Arabic Dialects Modeling Group

*Currently **Computational Approaches for Arabic Dialect Modeling***

Center for Computational Learning Systems. (CCLS)

Columbia University, New York NY, USA

<http://www.ccls.columbia.edu/oldweb/cadim/>

KEYNOTE ADDRESSES

1. Opening Address, **16th International Conference on Intelligent Text Processing and Computational Linguistics. (CICLING 15)**, Cairo, Egypt, Apr 15.
2. **1st Arabic Computational Linguistics Conference. (ARCLing15)**, Cairo, Egypt, Apr 15.
3. **9th IEEE International Conference on Computer Engineering and Systems. (ICCES 14)**, Dec 14.
4. **3rd Palestinian Symposium on Computational Linguistics and Arabic Content**, Mar 14.
5. **Joint Symposium on Semantic Processing, Organized by the European Union funded projects EXCITEMENT and B-CROCE hosted by FBKI**, Trento, Italy, Nov 13.
6. **27th Arabic Linguistic Symposium**. (ALS 13), Indiana IN, USA, Feb 28- Mar 2, 13
7. Plenary Invited Address. (Main Plenary Speech), **11th International Language Engineering Conference**, Cairo, Egypt, Dec 11.
8. **IEEE 5th International Conference on Semantic Computing. (ICSC)**, Palo Alto CA, USA, Sep 11.
9. **IEEE Information and Communication Technologies International Symposium. (ICTIS07)**, Fez, Morocco, Apr 07.
10. **6th Conference on Language Engineering**, Cairo, Egypt, Nov 06
11. **EACL workshop on Cross Language Knowledge Induction**, Trento, Italy, Apr 06
12. **Arabic Linguistics Society (ALS) Conference**, Kalamazoo MI, USA, Mar 06

INVITED TALKS/EVENT/PANELIST

1. Invited Talk. (2016) **King Saud University**, Riyadh, Kingdom of Saudi Arabia. Apr.
2. Invited Talk. (2016) **Al Imam University**, Riyadh, Kingdom of Saudi Arabia, Apr.
3. Invited Talk. (2016) **Princess Noura University**, Riyadh, Kingdom of Saudi Arabia, Apr.
4. Invited Talk. (2016) **King Abdullah University for Science & Technology**, Thuwal, Kingdom of Saudi Arabia, Mar.
5. Mutliword Expressions Special Event. (2015) **CICLING 15**, Cairo, Egypt, Apr 15.
6. Invited Talk. (2015) **IBM Watson and Cognitive Computing Institute**, New York NY, May 14.
7. Invited Talk. (2013) **MITRE Corporation**, Tyson's Corner VA, USA, Mar.
8. Invited Talk. (2013) Computer Science Symposium, **Simon Fraser University**, British

- Columbia, Vancouver BC, Canada, Mar.
9. Invited Talk. (2013) Computer Science Symposium, **CLIP Lab, UMIACS, University of Maryland College Park**, College Park MD, USA, Feb.
 10. Invited Panelist. (2012) Computational lexical resources and interoperability, **Language Resources and Evaluation Conference (LREC)**, Istanbul, Turkey.
 11. Invited Arbiter/Referree. (2012) One of 6 Arab Expat Scientists to decide on Best Scientific Research in Qatar, **Qatar Foundation, Arab Expat Society & Scientific Research Forum**, Doha, Qatar, Oct.
 12. Invited Participant. (2011) 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**, Nov.
 13. Invited Panelist. (2011) Future of Semantics in Machine Translation, **DARPA BOLT PI Kick-off Meeting**, Tampa FL, USA, Oct.
 14. Invited Talk. (2011) Upper Administration, **King Abdullah University for Science and Technology (KAUST)**, Thuwal, Saudi Arabia, Sep.
 15. Invited Talk. (2011) Mathematical & Computer Sciences School, **King Abdullah University for Science and Technology. (KAUST)**, Thuwal, Saudi Arabia, Sep.
 16. Invited Participant. (2011) **NSF-sponsored Workshop on Collaborative Resource Development and Delivery**. Organized by Nancy Ide. (Vassar College), New York NY, USA, Oct.
 17. Invited Participant. (2011) **NSF-funded Adjectives Workshop**, Organized by Christiane Fellbaum. (Princeton University), Arlington VA, USA, Sep 11
 18. Invited Panelist. (2011) **Empirical Methods For Natural Language Processing (EMNLP) 11 Workshop on Geometrical Methods in Semantics (GEMS)**, Edinburgh, UK, Jul.
 19. Invited Talk. (2011) **Distinguished Lecture Symposium Series, University of Washington**, Seattle WA, USA, Apr.
 20. Invited Talk. (2011) **Microsoft Research**, Redmond WA, USA, Apr 21.
 21. Invited Participant. (2011) **NSF-funded SILT Workshop Interoperable Semantic Annotation**. Organized by Nancy Ide. (Vassar College) and James Pustejovsky. (Brandeis University), Waltham MA, USA, Apr.
 22. Invited Participant. (2011) **QCRI Social Networks Roundtable**, Qatar Computing Research Institute, Qatar Foundation, Doha, Qatar, Apr.
 23. Invited Participant. (2011) **QCRI Arabic Language Technologies Roundtable**, Qatar Computing Research Institute, Qatar Foundation, Doha, Qatar, Jan.
 24. Invited Talk. (2010) **Computer Science Department, University of Alabama**, Birmingham AL, USA, Dec.
 25. Invited Panelist. (2010) **COLING 10 Workshop on Multiword Expressions. (MWE 10)**, Beijing, China, Aug.
 26. Invited Panelist. (2010) **ACL 10 Workshop on Semantic Evaluations. (SEMEVAL 10)**, Upsala, Sweden, Jul.
 27. Invited Talk and Participant. (2010) **Workshop on the Future of Arabic Language Development**, organized by the **UAE Crown Prince Court for Strategic Planning**, Abu Dhabi, UAE, Jul.
 28. Invited Talk. (2010) **IBM TJ Watson**, Hawthorne NY, USA, Jun.
 29. Invited Talk. (2010) **School of Computing and Mathematical Sciences, King Abdullah University for Science and Technology, KAUST**, Thuwal, Saudi Arabia, May.
 30. Invited Talk. (2010) **Computer Science Department, University of Trento**, Trento, Italy, Apr.
 31. Invited Panelist. (2009) **NAACL Workshop on Semantic Evaluations. (SEMEVAL**

- 09)**, Boulder CO, USA, Jun.
32. Invited Talk. (2009) **Departamento de Sistemas Informaticos y Computacion, Universidad Politecnica de Valencia**, Valencia, Spain, May.
 33. Invited Talk. (2009) **IRST & Computer Science Department, University of Trento**, Trento, Italy, Apr.
 34. Invited Talk. (2008) **Department of Informatics, University of Sussex**, Brighton, England, Aug.
 35. Invited Talk. (2008) **Computer Science Department, Faculty of Science, University of Mohamed V**, Rabat, Morocco, May.
 36. Invited Talk. (2008) **Departamento de Sistemas Informaticos y Computacion, Universidad Politecnica de Valencia**, Valencia, Spain, May.
 37. Invited Talk (2008). **Linguistics Department, Georgetown University**, Washington DC, USA, Apr.
 38. Invited Talk. (2006) **Cairo Microsoft Innovation Center. (CMIC), MSR**, Smart Village, Cairo, Egypt, Dec.
 39. Invited Talk. (2006) **Computer Science Department, University of North Texas**, Dallas TX, USA, Sep.
 40. Invited Talk. (2005) with Nizar Habash and Owen Rambow. **Linguistic Data Consortium, University of Pennsylvania**, Philadelphia PA, USA, Dec 05
 41. Invited Talk. (2005) **Computer Science Department, New York University**, New York NY, USA, Oct.
 42. Invited Talk. (2005) **Computer Science Department, Columbia University**, New York NY, USA, Mar.
 43. Invited Talk. (2005) **Linguistics Department, Ohio State University**, Columbus OH, USA, Feb.
 44. Invited Talk. (2005) **Linguistics Department, University of Texas Austin**, Austin TX, USA, Feb.
 45. Invited Talk. (2005) **Yahoo Research Labs, Overture**, Pasadena CA, USA, Jan 05
 46. Invited Talk. (2004) **Language Computer Corporation. (LCC)**, Dallas TX, USA, Dec.
 47. Invited Talk. (2004) **Microsoft Research (MSR), Natural Language Group**, Redmond WA, USA, Dec.
 48. Invited Talk. (2004) **Department of Informatics, University of Edinburgh**, Edinburgh, Scotland, Aug.
 49. Invited Talk. (2004) **Linguistic Data Consortium, University of Pennsylvania**, Philadelphia PA, USA, May.
 50. Invited Talk. (2004) **IBM T. J. Watson Research Center**, Hawthorne NY, USA, May.
 51. Invited Talk. (2003) **Linguistics Dept. & Institute for Cognitive Science, University of Colorado, Boulder**, Boulder CO, USA, Nov.
 52. Invited Talk. (2002) **Laboratoire Perception, Systèmes, Information. (PSI), Université de Rouen**, Rouen, France, Nov.
 53. Invited Talk. (2002) **NLP Lab, Information Sciences Institute. (ISI), University of Southern California**, Marina Del Rey CA, USA, Nov.

PROFESSIONAL CONSULTING

- **King Saud University** **Nov 15 – Mar 17**
Providing Expert Advice on Computational approaches to Readability Complexity Measures in Modern Standard Arabic for Patients' Pharmaceutical Leaflets
 Contact: Hend AlKhalifa
 Riyadh, Kingdom of Saudi Arabia
- **Qordoba Inc.** **Sep 11 – Sep 13**

Providing Expert Assessment of English Arabic MT and NLP Technologies

Contact: May Habib

Dubai, UAE

- **University of Colorado, Boulder** **Spr 10 – Dec 12**
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

1. **[Google Faculty Research Award]** *Tharwa++: Building a multidialectal Arabic Lexical Repository.* (PI) Total GWU \$75K, duration 1 yr., Sep 15 –Dec 16.
2. **[NSF CRI-P]** *A Unified Universal Multiword Expression Repository,* (PI) Total GWU \$100K, duration 1 yr., Jul 15 –Jun 17.
3. **[DARPA-LORELEI]** *Sentiment and Emotion Detection in Low Resource Languages (SEEM)* (GWU-PI; collaborative with Columbia University Julia Hirschberg & Kathy Mckeown) Total GWU \$240K, duration 4 yrs., Sep 15 – Aug 19.
4. **[Qatar National Research Fund (Q NRF)]** *Optdiac: An optimal Diacritization Scheme for Arabic,* (Lead PI; collaborative research with Carnegie Mellon University Qatar Co-PI Kemal Oflazer) Total \$1.1M. (GWU portion \$367K), duration 3 yrs., Sep 13 – Jan 17.
5. **[DARPA-DEFT]** *DRATS: Detecting Relations and Anomalies in Text and Speech,* (PI; subaward from Columbia University as Prime institution) Total for PI \$844,107. (GWU portion \$705K), duration 4.5 yrs., May 13 – May 17.
6. **[NSF-Large CRI]** *Creating Resources for Multilingual Linguistic Code Switching,* (PI; with Julia Hirschberg from Columbia University, collaborative with Thamar Solorio, UAB/UH, GWU Prime institution) Total GWU \$295K, duration 3 yrs., Apr 13 –Dec 16.

PAST RESEARCH GRANTS

1. **[Google Faculty Research Award]** *Nuanced Sentiment and Perspective Analysis for Arabic Social Media Text.* (PI) Total GWU \$72K, duration 1 yr., Dec 14 –Dec 15
2. **[DARPA-BOLT]** Subcontract to GWU from BBN – *Dialectal Tool and Resource Creation.* (PI; with Columbia University Nizar Habash and Owen Rambow) Total \$5.5M, duration 5 yrs., Oct 11 –May 15

3. [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 –Jun 13
4. [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
5. [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
6. [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
7. [DARPA-GALE] Subcontract to IBM-*Pilot Unified Framework for Semantic Textual Similarity*. (PI) Total \$25K, duration 8 months, Sep 11 –Apr 12
8. [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
9. [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
10. [NSF-CRI Planning] *Linguistic Code Switching Pilot Annotation*. (PI; with Julia Hirschberg) Total/yr. \$78K, duration 1 yr., Mar 10–Mar 11
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
12. [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
13. [NSA-TTO8] Subcontract to JHU-COE-*Committed Belief and Modality in Multigenres Text*. (Co-PI; with Owen Rambow) Total/yr. \$6K, duration 3 yrs., Aug 07 –Aug 10
14. [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
15. [NSF-SGER] *Automatic Processing of Natural Language Code-Switching*. (PI; with Owen Rambow) Total \$41K, duration 1 yr., Sep 07–Aug 08
16. [NSF-SGER] *Inducing Multiword Expressions from Comparable Corpora*. (PI) Total \$19K, duration 6 months, Sep 06–Feb 07
17. [NSF-JHU Summer Workshop] *Arabic Dialect Parsing*. (Co-PI) Total \$15K, duration 1.5 months, Jun 05–Jul 05
18. [DARPA-GALE] Subcontract to University of Colorado–OntoNotes. *Building an Arabic PropBank*. (PI) Total \$15K, duration 6 months, Sep 05–Feb 06

RESEARCH EDUCATION AND TEACHING EXPERIENCE

SEMESTER LONG CLASSES

- **Select Topics: Statistical Natural Language Processing** Fall 16, Spr 14,
Spr 13
Department of Computer Science
The George Washington University
- **Data Structures and Algorithms I in JAVA. (CSCI 1112)** Fall 13
Department of Computer Science
The George Washington University
- **Research Projects in Computer Science** Fall 07 – Fall 12

- Department of Computer Science
Columbia University*
- **Arabic Computational Linguistics** **Fall 09**
*Linguistics/Computer Science Dept.
Georgetown University*
 - **C as a Second Language** **Fall 95 – Spr 97**
*Department of Computer Science and Electrical Engineering
George Washington University*
 - **Theory of Computing** **Fall 91 – Spr 95**
 - **Algorithms and Formal Languages**
 - **Artificial Intelligence**
*Department of Computer Science
The American University in Cairo*

TUTORIALS

- **ARABIC DIALECT PROCESSING. (3-5 hours)**
(Co-Taught with Nizar Habash)
- Empirical Methods For Natural Language Processing (EMNLP)* **Oct 14**
Doha, Qatar
- NAACL* **Jun 12**
Montreal, Canada
- Georgetown University Round Table. (GURT)* **Apr 10**
Washington DC, USA
- 2nd International Conference on Arabic language Resources and Tools. (MEDAR)* **Apr 09**
Cairo, Egypt
- Language Resources & Evaluation Conference. (LREC)* **May 08**
Marrakech, Morocco
- Human Language Technologies–North American Association for Computational Linguistics. (HLT-NAACL)* **Jun 07**
Rochester NY, USA
- American Machine Translation Association. (AMTA)* **Aug 06**
Boston MA, USA
- **AN A-Z STATISTICAL NLP CRASH COURSE. (4 hours)** **Mar 07**
IEEE Information and Communication Technologies International Symposium. (ICTIS07)
Fez, Morocco
- **STATISTICAL NATURAL LANGUAGE PROCESSING: AN A-Z CRASH COURSE. (3 days)** **Sep 06**
*Faculty of Engineering and Computer Studies, Cairo University
Cairo, Egypt*
- **INTRODUCTION TO COMPUTATIONAL LINGUISTICS. (2 hours)** **Feb 05**
*Linguistics Department,
University of Texas Austin
Austin TX, USA*

POST DOCTORAL & RESEARCH SCIENTIST MENTORSHIP

- **Mohammed Attia** **Feb 14 – Nov 14**
*Received PhD July 08: University of Manchester, UK.
Research Scientist Focus: Tharwa Resources and Spell Checking
Current Position: Research Scientist, Google Inc., New York NY, USA*
- **Mahmoud Ghoneim** **Jul 12 – present**

Received PhD July 12: Faculty of Engineering, Cairo University, Egypt
 Postdoc Focus: Integrating Semantic Knowledge in Statistical Machine Translation
 Current Position: Research Scientist, Department of Computer Science, George Washington University, Washington DC, USA

- **Abdelati Hawwari** **Mar 12 – present**
 Received PhD August 09: Faculty of Literature, Cairo University, Egypt
 Postdoctoral Focus: Arabic Lexical Semantics and Multiword Expressions
 Current Position: Research Scholar, Department of Computer Science, George Washington University, Washington DC, USA
- **Davide Picca** **May 10 – Apr 11**
 Received PhD April 10: 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 09: Polytechnical University of Valencia, Spain
 Postdoc Focus: Shallow syntactic parsing in Arabic and integration in SMT and IE
 Current Position: Research Scientist at Luki Labs, New York NY, USA
- **Marine Carpuat** **May 09 – Apr 11**
 Received PhD April 08: HKUST, Hong Kong
 PostDoc Focus: Incorporating Lexical Semantics in Statistical Machine Translation
 Current Position: Assistant Professor, Computer Science Department, University of Maryland College Park, Maryland MD, USA

PHD THESIS COMMITTEE MEMBER

1. **Aspect Extraction for Sentiment Analysis in Arabic Dialect** (Thesis Proposal) **Spr 16**
Alawya Alawami
 Graduate Program in Information Science and Technology
 School of Information Sciences, University of Pittsburg, PA, USA
2. **Stative and Stativizing Constructions in Arabic News Reports: A corpus based study** (Thesis) **Fall 15**
Aous Mansouri
 Department of Linguistics
 University of Colorado, Boulder CO, USA
3. **Text Mining on Social Networking using NLP Techniques** (Thesis) **Fall 14**
Walaa Medhat
 Faculty of Engineering
 Ain Shams University, Cairo, Egypt
4. **Developing a language-learning model to incrementally infer syntactic structures from a continuous stream of English-language discourse.** (Thesis Proposal) **Fall 13**
John Sipple
 Department of Computer Science
 George Washington University, Washington DC, USA
5. **Imperfective Constructions in Arabic.** (Thesis Proposal) **May 12**
Aous Mansouri
 Department of Linguistics
 University of Colorado, Boulder CO, USA
6. **Reverse Feature Engineering for SVMs** (Thesis) **Apr 10**

- Daniele Pighin
Computer Science Department
University of Trento, Italy
7. **Semantic Fusion and Generation. (Thesis Proposal)** **Dec 09**
- Kapil Thadani
Department of Computer Science
Columbia University, New York NY, USA
8. **Automatic Processing of Language Frame Semantics** **Apr 09**
- Bonaventura Coppola
ICT division
University of Trento, Italy
9. **Information Retrieval: English and Arabic Documents** **May 08**
- Fadoua Ataa Allah
Faculty of Science
University of Mohamed V, Rabat, Morocco

RESEARCH SUPERVISION

GRADUATED

1. **Vinodkumar Prabhakaran, PhD** **Jul 15**
Joint Supervision with Owen Rambow
Thesis Title: Social Power in Interactions: Computational Analysis and Detection of Power Relations
Department of Computer Science
Columbia University, New York NY, USA
Current Position: Postdoctoral Scientist, Stanford University. (Mentor Daniel Jurafsky)
2. **Weiwei Guo, PhD** **Apr 15**
Thesis Title: Dimension Reduction for Short Text Similarity and its Applications
Department of Computer Science
Columbia University, New York NY, USA
Current Position: Research Scientist, Linked In, CA USA
3. **Muhammad Abdul-Mageed, PhD** **Dec 14**
Thesis Title: Sentiment Analysis in Arabic Social Media Text
Joint Supervision with Sandra Kuebler
Linguistics Department
Information Sciences Department
Indiana University, Indiana IN, USA
Current Position: Assistant Professor, University of British Columbia, Canada
4. **Pradeep Dasigi, MSc** **Dec 12**
Thesis Title: Dialectal Arabic Orthography Normalization
Department of Computer Science
Columbia University, New York NY, USA
Current Position: PhD Candidate at Carnegie Mellon University working with Ed Hovy
5. **Mahmoud Ghoneim, PhD** **Mar 12**
Thesis Title: Explicit Modeling of Semantics in Statistical Machine Translation
Joint Supervision with Professor Mohsen Rashwan
Computer Engineering Department
Faculty of Engineering, Cairo University, Egypt
Current Position: Research Scientist, Department of Computer Science, George Washington University
6. **Yassine Benajiba, PhD** **Jul 09**

Thesis Title: Arabic Named Entity Recognition

Joint Supervision with Professor Paolo Russo

Polytechnical University of Valencia, Spain

Current Position: Research Scientist Luki Labs, New York, NY, USA

CURRENT (Date Started)

1. **Hebatullah EIFardy** **Fall 10**
PhD Candidate
Thesis Topic: Perspective Identification in multilingual and multigenre contexts
Department of Computer Science
Columbia University, New York NY, USA
Expected Graduation: Fall 16
2. **Ali Elkahky** **Spr 12**
PhD Candidate
Thesis Topic: Unified models for dialectal processing
Department of Computer Science
Columbia University, New York NY, USA
Expected Graduation: Spring 18
3. **Aya Zirikly** **Spr 13**
PhD Candidate
Thesis Topic: Tracking of Named Entities in Multilingual and Multigenre Settings
Department of Computer Science
George Washington University, Washington DC, USA
Expected Graduation: Spring 17
4. **Maryam Aminian** **Fall 13**
PhD Student
Thesis Topic: Domain Independent Semantic Role Labeling
Department of Computer Science
George Washington University, Washington DC, USA
5. **Hanan AIDarmaki** **Fall 13**
PhD Student
Thesis Topic: Cross Language Semantic Textual Similarity
Department of Computer Science
George Washington University, Washington DC, USA
6. **Shabnam Tafreshi** **Spr 14**
PhD Student
Thesis Topic: Affect and Emotion Detection
Department of Computer Science
George Washington University, Washington DC, USA
7. **Fahad AIGamdi** **Spr 14**
PhD Student
Thesis Topic: Computational Approaches to Linguistic Code Switching
Department of Computer Science
George Washington University, Washington DC, USA
8. **Nada AIMarwani** **Spr 14**
PhD Student
Thesis Topic: Semantically Aware Low Resource Language Machine Translation
Department of Computer Science
George Washington University, Washington DC, USA
9. **Sawsan AIQahtani** **Spr 14**

- PhD Student*
Thesis Topic: Multilingual Optimal Diacritization for Morphologically Rich Languages
 Department of Computer Science
 George Washington University, Washington DC, USA
10. **Sardar Hamadian** **Fall 14**
PhD Student
Thesis Topic: Automatic False Information Detection
 Department of Computer Science
 George Washington University, Washington DC, USA
11. **Ali Seyfi** **Fall 14**
PhD Student
Thesis Topic: Event Detection and Tracking
 Department of Computer Science
 George Washington University, Washington DC, USA
12. **Mohammed AlBadrashiny** **Spr 15**
PhD Student
Thesis Topic: Language Identification and Spelling error correction
 Department of Computer Science
 George Washington University, Washington DC, USA

PUBLICATIONS

BOOKS

1. **Diab, Mona** and Christiane Fellbaum. (*forthcoming*) Computational Lexical Semantics: A Multilingual Perspective. *Springer Publishers*, under contract, expected Spring 17.
2. **Diab, Mona**, Nizar Habash, Owen Rambow. (*forthcoming*) Processing Arabic Dialects. *Cambridge Press*, under contract, expected Fall 17.
3. **Diab, Mona**, Aline Villavicencio, Anna Korhonen, Marianna Apidianaki, Valia Kordoni, Mark Stevenson, Preslav Nakov, *Editors*. (*forthcoming*) Essays Computational Lexical Semantics in Honor of Adam Kilgarriff. *Springer Publishers*, under contract, expected Fall 16.
4. **Diab, Mona**. (2003) Word Sense Disambiguation within a multilingual framework. *PhD Thesis, University of Maryland, College Park MD, USA*.

EDITED VOLUMES

1. **Diab, Mona**, Pascale Fung, Julia Hirschberg, Tamar Solorio, *Editors*. (2016) 2nd Workshop on Computational Approaches to Linguistic Code Switching. *In Proceedings of Empirical Methods of Natural Language Processing (EMNLP)*.
2. **Diab, Mona**, Houda Bouamor, Ahmed EIKholy, Yuval Marton, Mahmoud Ghoneim, *Editors*. (2016) Workshop on Machine Translation for Semitic Languages (SEMAT). *In Proceedings of Automatic Machine Translation in the Americas (AMTA)*.
3. **Diab, Mona**, Pascale Fung, Julia Hirschberg, Tamar Solorio, *Editors*. (2014) 1st Workshop on Computational Approaches to Linguistic Code Switching. *In Proceedings of Empirical Methods of Natural Language Processing (EMNLP)*.
4. **Diab, Mona**, Timothy Baldwin, Marco Baroni. *Editors*. (2013) 2nd International Joint Conference on Semantics (*SEM). *Proceedings of *SEM 2013*.
5. Agirre, Eneko, Johan Bos, **Mona Diab**. *Editors*. (2012) 1st International Joint Conference on Semantics (*SEM). *Proceedings of *SEM 2012*.
6. Matveeva, Irina, Chris Biemann, Monojit Choudhury, **Mona Diab**. *Editors*. (2008) Proceedings of the 3rd Textgraphs workshop on Graph-based Algorithms for Natural Language Processing. *COLING 2008*.

7. Choukri, Khalid, **Mona Diab**, Bente Maegaard, Paolo Rosso, Abdelhadi Souidi, Ali Farghaly. *Editors*. (2008) Proceedings of LREC Workshop on HLT & NLP within the Arabic world: Arabic Language and local languages processing: Status Updates and Prospects. LREC 2008.
8. Darwish, Kareem, **Mona Diab**, Nizar Habash, *Editors*. (2005) Proceedings of Computational Approaches to Semitic Languages. Association for Computational Linguistics.

JOURNAL ARTICLES

1. Al Aqeel S, Abanmy N, Abeer Aldayel, Hend S. Al-Khalifa, Maha Al-Yahya, **Mona Diab**. (Submitted). Readability of medication information materials in Saudi Arabia: expert and non-expert evaluation. Journal of Methods of Medical Information.
2. Zaghouni, Wajdi, Abdelati Hawwari, **Mona Diab**. (2016) AMPN: A Lexical Semantic Resource for Arabic Morphological Patterns. International Journal of Speech Technologies, Springer Publishers.
3. Abdul-Mageed, Mohammad, **Mona Diab** and Sandra Kuebler. (2014) SAMAR: A System for Subjectivity and Sentiment Analysis for Arabic Social Media. Computer Speech and Language, 28. (201), pp. -37.
4. Abdul-Mageed, Muhammad, and **Diab, Mona**. (2012) Linguistically-motivated subjectivity and sentiment annotation and tagging of Modern Standard Arabic. International Journal on Social Media MMM: Monitoring, Measurement, and Mining.
5. 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.
6. 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.
7. Resnik, Philip, Mari B Olsen, **Mona Diab**. (201999) The Bible as a Parallel Corpus: Annotating the 'Book of 00 Tongues' Computers and the Humanities, 33. (201-2).

INVITED ARTICLES

1. **Diab, Mona**. (2015) Tharawat: A Vision for a Comprehensive Resource for Arabic Computational Processing. Journal for Computational Linguistics and Intelligent Text Processing, pp. 85-97.
2. **Diab, Mona** and Yuval Marton. (2014) Semitic Semantics. Book Chapter in Natural Language Processing for Semitic Languages, Editor Imed Zitouni, Springer Publishers, pp. 129-159.
3. ElFardy, Hebatullah, Mohamed AlBadrashiny, **Mona Diab**. (2014) A Hybrid System for Code Switch Point Detection in Informal Arabic Text. XRDS: Crossroads, The ACM Magazine for Students 21. (201), pp. 52-57.
4. Bar, Kfir, **Mona Diab**, Abdelati Hawwari. (2013) Arabic Multiword Expressions: Resource and Tool Creation. Book Chapter in Natural Language Processing for Semitic Languages, Editors Naachum Derschowitz and Ephraim Nissim, Vol.3, in honor of Yaacov Choueka, Springer Publishers.
5. **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.
6. Katz, Graham and **Mona Diab**. (2010) Special Issue on Georgetown Round Table on

Arabic Computational Linguistics Conference. (Guest Editors), *ACM Transactions on Asian Language Information Processing*. (TALIP).

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

1. Hawwari, Abdelati, Mohammed Attia, Mahmoud Ghoneim, **Mona Diab.** (2016) Explicit Fine grained Syntactic and Semantic Annotation of the Idafa Construction in Arabic. In *Proceedings of LREC 2016, Slovenia, May*.
2. **Diab, Mona,** Mahmoud Ghoneim, Abdelati Hawwari, Fahad AlGhamdi, Nada AlMarwani, Mohamed Al-Badrashiny. (2016) Creating a Large Multi-Layered Representational Repository of Linguistic Code Switched Arabic Data. In *Proceedings of LREC 2016, Slovenia, May*.
3. Zaghouni, Wajdi, Houda Bouamor, Abdelati Hawwari, **Mona Diab,** Ossama Obeid, Mahmoud Ghoneim, Sawsan Alqahtani, Kemal Oflazer. (2016) Large Scale Arabic Diacritized Corpus: Guidelines and Framework. In *Proceedings of LREC 2016, Slovenia, May*.
4. Al-Badrashiny, Mohamed, Arfath Pasha, **Mona Diab,** Nizar Habash, Owen Rambow, Wael Salloum, and Ramy Eskander, SPLIT: Smart Preprocessing (Quasi) Language Independent Tool, In *Proceedings of LREC 2016, Slovenia, May*.
5. Prabhakaran, Vinodkumar, and Tomas By, Julia Hirschberg, Owen Rambow, Samira Shaikh, Tomek Strzalkowski, Jennifer Tracey, Michael Arrigo, Rupayan Basu, Micah Clark, Adam Dalton, **Mona Diab,** Louise Guthrie, Anna Prokofieva, Stephanie Strassel, Gregory Werner, Yorick Wilks and Janyce Wiebe. (2015) A New Dataset and Evaluation for Belief/Factuality. In *Proceedings of 4th *SEM Conference, Denver CO, USA*.
6. Elfardy, Heba, and **Mona Diab** and Chris Callison-Burch. (2015) Ideological Perspective Detection Using Semantic Features. In *Proceedings of 4th *SEM Conference, Denver CO, USA*.
7. Al-Badrashiny, Mohamed, Heba Elfardy, **Mona Diab.** (2015) AIDA2: A Hybrid Approach for Token and Sentence Level Dialect Identification in Arabic. In *Proceedings of the 9th Conference on Natural Language Learning (CoNLL), Beijing, China*.
8. Zirikly, Aya, Yassine Benajiba, **Mona Diab.** (2015) GWU-EDL English Diagnostic Short System Description. KPB Task Evaluation, *Proceedings of the TAC Conference, NIST, Maryland MD, USA*.
9. Abdul-Mageed, Muhammad, **Mona Diab.** (2014) SANA: A Large Scale Multi-Genre, Multi-Dialect Lexicon for Arabic Subjectivity and Sentiment Analysis. In *Proceedings of Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland*.
10. **Diab, Mona,** Mohamed AlBadrashiny, Maryam Aminian, Mohamed Attia, Hebatullah EIFardy, Nizar Habash, Abdelati Hawwari, Wael Salloum, Pradeep Dasigi, Ramy Eskander. (2014) Tharwa: A Large Scale Dialectal Arabic –Standard Arabic – English Lexicon. In *Proceedings of Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland*.
11. Guo, Weiwei, Wei Liu, **Mona Diab.** (2014) Fast Tweet Retrieval with Compact Binary Codes. In *Proceedings of Conference of Computational Linguistics (COLING), Dublin, Ireland*.
12. Pasha, Arfath, Mohamed AlBadrashiny, **Mona Diab,** Ahmed EIKholy, Ramy Eskander, Nizar Habash, Manoj Poolery, Owen Rambow, Ryan Roth. (2014) MADAMIRA: A Fast,

- Comprehensive Tool for Morphological Analysis and Disambiguation of Arabic. In *Proceedings of Language Resources and Evaluation Conference (LREC)*, Reykjavik, Iceland.
13. Salloum, Wael, Hebatullah EIFardy, Linda Alamir-Salloum, Nizar Habash, **Mona Diab**. (2014) Sentence Level Dialect Identification for Machine Translation System Selection. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, Baltimore MD, USA.
 14. Abdul-Mageed, Muhammad, **Mona Diab**, Sandra Kubler. (2013) ASMA: A system for Automatic Segmentation and Morpho-syntactic Disambiguation of Modern Standard Arabic. In *Proceedings of Recent Advances in Natural Language Processing (RANLP)*, Sophia, Bulgaria.
 15. Abu-Jbara, Amjad, Ben King, **Mona Diab**, Dragomir Radev. (2013) Identifying Opinion Subgroups in Arabic Online Discussions. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*, Sofia, Bulgaria.
 16. Agirre, Eneko, Daniel Cer, **Mona Diab**, Aitor Gonzalez-Agirre and Weiwei Guo. (2013) *SEM 13 shared task: Semantic Textual Similarity. In *Proceedings of *SEM*, Atlanta GA, USA.
 17. EIFardy, Hebatullah, and **Mona Diab**. (2013) Sentence-Level Dialect Identification in Arabic, In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*, Sofia, Bulgaria.
 18. EIFardy, Hebatullah, Mohamed Al-Badrashiny, **Mona Diab**. (2013) Code-Switch Point Detection in Arabic, In *Proceedings of the 18th International Conference on Application of Natural Language to Information Systems (NLDB)*, MediaCity, Manchester, UK.
 19. Ghoneim, Mahmoud and **Mona Diab**. (2013) Multiword Expressions in the context of Statistical Machine Translation. In *Proceedings of IJCNLP*, Nagoya, Japan.
 20. Guo, Weiwei, and **Mona Diab**. (2013) Improving Lexical Semantics for Sentential Semantics: Modeling Selectional Preference and Similar Words in a Latent Variable Model. In *Proceedings of North American Association for Computational Linguistics (NAACL)*, Atlanta GA, USA.
 21. Guo, Weiwei, Hao Li, Heng Ji and **Mona Diab**. (2013) Linking Tweets to News: A Framework to Enrich Online Short Text Data in Social Media. In *Proceedings of Association for Computational Linguistics (ACL)*, Sofia, Bulgaria.
 22. Hawwari, Abdelati, Wajdi Zaghouni, Tim O'Gorman, **Mona Diab**, and Ahmed Badran. (2013) Building a Lexical Semantic Resource for Arabic Morphological Patterns. In *Proceedings of ICCSPA13*, Sharjah, UAE.
 23. Tomeh, Nadi, Nizar Habash, Ryan Roth, Noura Farra, Pradeep Dasigi and **Mona Diab**. (2013) Ensemble Reranking with Linguistic and Semantic Features for Arabic Character Recognition, In *Proceedings of Association for Computational Linguistics (ACL)*, Sofia, Bulgaria.
 24. Zirikly, Aya and **Mona Diab**. (2013) ANEAR: Automatic Named Entity Aliasing Resolution. In *Proceedings of the 18th International Conference on Application of Natural Language to Information Systems (NLDB)*, MediaCity, Manchester, UK.
 25. Abdul-Mageed, Muhammad and **Mona Diab**. (2012) Toward Building a large scale Arabic Sentiment Lexicon. In *Proceedings of the 6th International Global WordNet Conference*, Matsue, Japan.
 26. Abdul-Mageed, Muhammad and **Mona Diab**. (2012) AWATIF: A multigenre corpus for MSA subjectivity and sentiment Analysis. In *Proceedings of Language Resources and Evaluation Conference (LREC)*, Istanbul, Turkey.
 27. Abu Jabara, Amjad, Pradeep Dasigi, **Mona Diab**, Dragomir Radev. (2012) Subgroup Detection in Ideological Discussions. In *Proceedings of Association for Computational Linguistics (ACL)*, Jeju Island, South Korea.

28. Dasigi, Pradeep, Weiwei Guo, and **Diab, Mona (2012)** Cross-genre subgroup detection in online discussion threads: A study of the implicit attitude using textual latent semantics. *In Proceedings of Association for Computational Linguistics (ACL)*, Jeju Island, South Korea.
29. ElFardy, Hebatullah and **Mona Diab. (2012)** Simplified guidelines for the creation of Large Scale Dialectal Arabic Annotations. *In Proceedings of Language Resources and Evaluation Conference (LREC)*, Istanbul, Turkey.
30. ElFardy, Hebatullah and **Mona Diab. (2012)** AIDA: Automatic Identification & Glossing of Dialectal Arabic. *Poster Session, In Proceedings of European Association for Machine Translation (EAMT)*, Trento, Italy.
31. Guo, Weiwei and **Mona Diab. (2012)** Modeling Semantic Textual Similarity in the Latent Space. *In Proceedings of Association for Computational Linguistics (ACL)*, Jeju Island, South Korea.
32. Guo, W. & **Diab, Mona. (2012)** Learning the latent semantics of a concept from its definition. *In Proceedings of Association for Computational Linguistics (ACL)*, Jeju Island, South Korea.
33. Habash, Nizar, **Mona Diab**, Owen Rambow. (2012) CODA: A Conventional Orthography for Dialectal Arabic. *In Proceedings of Language Resources and Evaluation Conference (LREC)*, Istanbul, Turkey.
34. Prabhakaran, Vinodkumar, Owen Rambow, **Mona Diab. (2012)** Who's. (Really) the Boss? Perception of Situational Power in Written Interactions. *In Proceedings of Conference on Computational Linguistics (COLING)*, Mumbai, India.
35. Prabhakaran, Vinodkumar, Owen Rambow, **Mona Diab. (2012)** Annotations for Power Relations on Email Threads. *In Proceedings of Language Resources and Evaluation Conference (LREC)*, Istanbul Turkey.
36. Prabhakaran, Vinodkumar, Owen Rambow, **Mona Diab. (2012)** Predicting Overt Display of Power in Written Dialogs. *In Proceedings of North American Association for Computational Linguistics (NAACL), Short Papers*, Montreal, Canada.
37. Abdul-Mageed, Muhammad, **Mona Diab**, Mohammed Korayem. (2011) Subjectivity And Sentiment Analysis Of Modern Standard Arabic. *In Proceedings of Association for Computational Linguistics (ACL)*, Portland OR, USA.
38. **Diab, Mona** and Pradeep Dasigi. (2011) CODACT: Towards Identifying Orthographic Variants in Dialectal Arabic. *In Proceedings of IJCNLP*, Chiang Mai, Thailand.
39. 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 Eleventh Conference on Language Engineering (ESOLEC11)*, Cairo, Egypt.
40. Guo, Weiwei and **Mona Diab. (2011)** Semantic Topic Models: Combining Word Distributional Statistics and Dictionary Definitions. *In Proceedings of Empirical Methods For Natural Language Processing (EMNLP)*, Edinburgh, Scotland, UK.
41. Benajiba Yassine, Imed Zitouni, **Mona Diab**, Paolo Rosso. (2010) Arabic Named Entity Recognition: Using Features Extracted from Noisy Data. *In Proceedings of Association for Computational Linguistics (ACL)*, Uppsala, Sweden.
42. Carpuat, Marine, and **Mona Diab. (2010)** Task-based Evaluation of Multiword Expressions: a Pilot Study in Statistical Machine Translation. *In Proceedings of North American Association for Computational Linguistics (NAACL)*, Los Angeles CA, USA.
43. **Diab, Mona**, Nizar Habash, Owen Rambow, Mohamed Al Tantawy, Yassine Benajiba. (2010) COLABA: Arabic Dialect Annotation and Processing. *In Proceedings of the Workshop on Semitic Language Processing, Language Resources and Evaluation Conference (LREC)*, Malta.
44. Guo, Weiwei and **Mona Diab. (2010)** Combining Orthogonal Monolingual and

- Multilingual Sources of Evidence for All Words WSD. *In Proceedings of Association for Computational Linguistics (ACL)*, Uppsala, Sweden.
45. Prabhakaran, Vinodkumar, Owen Rambow, **Mona Diab**. (2010) Automatic Committed Belief Tagging. *In Proceedings of COLING*, Beijing, China.
 46. **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*, Cairo, Egypt.
 47. **Diab, Mona** and Madhav Krishna. (2009) Unsupervised classification of Verb Noun Multiword Expression Tokens. *CICLING*, Mexico City, Mexico.
 48. Parton, Kristen, 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. *In Proceedings of ACL-IJCNLP*, Singapore.
 49. BenAjiba, Yassine, **Mona Diab**, Paolo Rosso. (2008) Arabic Named Entity Recognition: An SVM Based Approach. *International Arab Conference on Information Technology (ACIT'08)*, University of Sfax, Sfax, Tunisia.
 50. 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)*, Hawaii HI, USA.
 51. **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)*, Columbus OH, USA.
 52. Palmer, Martha, Ann Bies, Olga Babko-Malaya, **Mona Diab**, Mohamed Maamouri, Aous Mansouri, Wajdi Zaghouni. (2008) A pilot Arabic Propbank. *In Proceedings of Language Resources and Evaluation Conference (LREC)*, Marrakech, Morocco.
 53. 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)*, Columbus OH, USA.
 54. **Diab, Mona**. (2007) Towards an optimal POS tag set for Modern Standard Arabic Processing. *In Proceedings of Recent Advances in Natural Language Processing (RANLP)*, Borovets, Bulgaria.
 55. **Diab, Mona** and Alessandro Moschitti. (2007) Semantic Parsing for Modern Standard Arabic. *In Proceedings of Recent Advances in Natural Language Processing (RANLP)*, Borovets, Bulgaria.
 56. **Diab, Mona**, Mahmoud Ghoneim, Nizar Habash. (2007) Arabic Diacritization in the Context of Statistical Machine Translation. *In Proceedings of Machine Translation Summit. (MT-Summit)*, Copenhagen, Denmark.
 57. Kirchhoff, Katrin, Owen Rambow, Nizar Habash, **Mona Diab**. (2007) Semi-Automatic Error Analysis for Large-Scale Statistical Machine Translation Systems. *In Proceedings of Machine Translation Summit. (MT-Summit)*, Copenhagen, Denmark.
 58. 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)*, Trento, Italy.
 59. **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*, Cairo, Egypt.
 60. **Diab, Mona**, and Neal Snider. (2006) Modern Standard Arabic Verb Alternations. *In Proceedings of Annual Meeting of the Arabic Linguistics Society (ALS)*, Kalamazoo MI, USA.

61. 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)*, Genoa, Italy.
62. 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)*, Sydney, Australia.
63. 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)*, New York, NY, USA.
64. **Diab, Mona**. (2004) Relieving the Data Acquisition Bottleneck for Word Sense Disambiguation. *In Proceedings of Association for Computational Linguistics (ACL)*.
65. **Diab, Mona**, Kadri Hacioglu, Daniel Jurafsky. (2004) Automatic Tagging of Arabic Text: From Raw Text to Base Phrase Chunks. *In Proceedings of Human Language Technology-North American Association for Computational Linguistics (HLT-NAACL)*.
66. **Diab, Mona**. (2004) An Unsupervised Approach for Bootstrapping Arabic Sense Tagging. *In Proceedings of Arabic Workshop, Computational Linguistics (COLING)*.
67. **Diab, Mona**. (2004) Bootstrapping a WordNet Taxonomy for Arabic. *Proceedings of First Arabic Language Technology Conference. (NEMLAR)*, Cairo, Egypt.
68. **Diab, Mona** and Philip Resnik. (2002) An Unsupervised Method for Word Sense Tagging using Parallel Corpora. *In Proceedings of Association for Computational Linguistics. (ACL)*, Philadelphia PA, USA.
69. **Diab, Mona** and Steven Finch. (2000) A Statistical Word-Level Translation Model for Comparable Corpora. *In Proceedings of Conference on Content-based Multimedia Information Access (RIA00)*, Paris, France.
70. Resnik, Philip and **Mona Diab**. (2000) Measuring Verb Similarity. *In Proceedings of the conference of International Cognitive Science Society (COGSCI00)*, Philadelphia PA, USA.
71. 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. *In Proceedings of North American Association for Computational Linguistics. (NAACL)*, Seattle, WA, USA.
72. **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. *In Proceedings of 6th International Conference on Artificial Intelligence & Applications*, Cairo, Egypt.
73. Riopka, Terry, **Mona Diab**, Peter Bock. (1998) Quantifying and Interpreting the Effect of Intelligent Information Exchange. *In Proceedings of 6th International Conference on Artificial Intelligence & Applications*, Cairo, Egypt.
74. Resnik, Philip, Mari Broman Olsen, **Mona Diab**. (1997) Creating a Parallel Corpus from the Book of 00 Tongues, *Text Encoding Initiative 10th Anniversary User Conference. (TEI-10)*, Providence RI, USA.

PEER REVIEWED WORKSHOP PAPERS

1. Ossama Obeid, Houda Bouamor, Wajdi Zaghouni, Mahmoud Ghoneim, Abdelati Hawwari, **Mona Diab**, Kemal Oflazer. (2016) MANDIAC: A Web-based Annotation System For Manual Arabic Diacritization. *Proceedings of the 2nd Workshop on Arabic Corpora and Processing Tools*, LREC 2016.
2. Abdul-Mageed, Muhammad, Hassan AlHuzliyy, Duaa' Abu Elhija, **Mona Diab**. (2016)

- DINA: A Multi-Dialect Dataset for Arabic Emotion Analysis. *Proceedings of the 2nd Workshop on Arabic Corpora and Processing Tools*, LREC 2016.
3. Elfardy, Heba and **Mona Diab**. (2016) CU-GWU Perspective at SemEval-2016 Task 6: Ideological Stance Detection in Informal Text. *In Proceedings of the International Workshop on Semantic Evaluation (SemEval 2016), NAACL 2016*. San Diego, CA, USA.
 4. Elfardy, Heba and **Mona Diab**. (2016) Annotation Complexity: The Case of Annotating Ideological Perspective in Egyptian Social Media. *In Proceedings of the 10th Linguistic Annotation Workshop (LAW X), ACL 2016*. Berlin, Germany.
 5. AlDarmaki, Hanan, and **Mona Diab**. (2016) GWU-NLP at SemEval Task 1: Matrix Factorization for Cross Lingual STS. *In Proceedings of the International Workshop on Semantic Evaluation (SemEval 2016), NAACL 2016*. San Diego, CA, USA.
 6. AlDarmaki, Hanan, and **Mona Diab**. (2016) Learning Cross-Lingual Representations with Matrix Factorization. *Proceedings of Multilingual and Cross-Lingual Methods in Computational Linguistics (MLCL), NAACL*, San Diego CA, Jun.
 7. Sardar Hamidian and **Mona Diab**. (2016) Rumor Identification and Belief Investigation on Twitter. *In Proceedings of Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis (WASSA), NAACL 2016*, San Diego CA, Jun.
 8. Agirre, Eneko, Carmen Banea, Daniel Cer, **Mona Diab**, Aitor Gonzalez-Agirre, Rada Mihalcea, German Rigau, Janyce Wiebe. (2016) SemEval-16 Task 1: Semantic Textual Similarity, Monolingual and Cross Lingual Evaluation. *In Proceedings of SEMEVAL, North American Association for Computational Linguistics, NAACL 2016*, San Diego CA, Jun.
 9. Agirre, Eneko, Carmen Banea, Claire Cardie, Daniel Cer, **Mona Diab**, Aitor Gonzalez-Agirre, Weiwei Guo, Inigo Lopez-Gazpio, Montse Maritxalar, Rada Mihalcea, German Rigau, Larraitz Uria, Janyce Wiebe. (2015) SemEval-15 Task 2: Semantic Textual Similarity, English, Spanish and Pilot on Interpretability. *In Proceedings of SEMEVAL, North American Association for Computational Linguistics (NAACL)*, Denver CO, USA.
 10. Aminian, Maryam, Mahmoud Ghoneim, **Mona Diab**. (2015) Unsupervised False Friend Disambiguation Using Contextual Word Clusters and Parallel Word Alignments. *In Proceedings of Workshop 9th Semantics Syntax Statistical Translation (SSST), North American Association for Computational Linguistics (NAACL)*, Denver CO, USA. **BEST PAPER AWARD.**
 11. Werner, Gregory, and Vinodkumar Prabhakaran, **Mona Diab** and Owen Rambow. (2015) Committed Belief Tagging on the Factbank and LU Corpora: A Comparative Study. *In Proceedings of Workshop EXPROM, North American Association for Computational Linguistics (NAACL)*, Denver CO, USA.
 12. Zirikly, Ayah, **Mona Diab**. (2015) Named Entity Recognition for Arabic Social Media. *In Proceedings of Workshop on Vector Space Models for NLP, North American Association for Computational Linguistics (NAACL)*, Denver CO, USA.
 13. Bouamor, Houda, Wajdi Zaghouni, **Mona Diab**, Ossama Obeid, Kemal Oflazer, Mahmoud Ghoneim, Abdelati Hawwari. (2015) A Pilot Study on Arabic Multi-Genre Corpus Diacritization. *In Proceedings of Second Workshop on Arabic Natural Language Processing (ANLP), Association for Computational Linguistics (ACL)*, Beijing, China.
 14. Attia, Mohammed, Mohamed Al-Badrashiny, **Mona Diab**. (2015) GWU-HASP-15@\$QALB-15 Shared Task: Priming Spelling Candidates with Probability. *In Proceedings of Second Workshop on Arabic Natural Language Processing (ANLP), Association for Computational Linguistics (ACL)*, Beijing, China.
 15. Aldarmaki, Hanan and **Mona Diab**. (2015) Robust Part-of-speech Tagging of Arabic Text. *In Proceedings of Second Workshop on Arabic Natural Language Processing (ANLP), Association for Computational Linguistics (ACL)*, Beijing, China.
 16. Hamadian, Sardar, and **Mona Diab**. (2015) Improved Automatic Rumor Detection. *In Proceedings of The Fifth International Conference on Social Media Technologies*,

- Communication, and Informatics (SOTICS)*, Barcelona, Spain.
17. Agirre, Eneko, Carmen Banea, Claire Cardie, Daniel Cer, **Mona Diab**, Aitor Gonzalez-Agirre, Weiwei Guo, Rada Mihalcea, German Rigau, Janyce Wiebe. (2014) SemEval-14 Task 10: Multilingual Semantic Textual Similarity. In *Proceedings of SEMEVAL, Conference of Computational Linguistics (COLING)*, Dublin, Ireland.
 18. Aminian, Maryam, Mahmoud Ghoneim, **Mona Diab**. (2014) Handling OOV Words in Dialectal Arabic to English Machine Translation. In *Proceedings of Workshop on Language Technology for Closely Related Languages and Language Variants (LT4CloseLang), Empirical Methods For Natural Language Processing (EMNLP)*, Doha, Qatar.
 19. Attia, Mohamed, Mohamed AlBadrashiny, **Mona Diab**. (2014) GWU-HASP: Hybrid Arabic Spelling and Punctuation Correction. In *Proceedings of Workshop on Arabic Natural Language Processing (ANLP), Empirical Methods For Natural Language Processing (EMNLP)*, Doha, Qatar.
 20. ElFardy, Hebatullahtullah, Mohamed AlBadrashiny, **Mona Diab**. (2014) AIDA: Identifying Code Switching in Informal Arabic Text. In *Proceedings of First Workshop on Computational Approaches to Linguistic Code Switching (CodeSwitch), Empirical Methods For Natural Language Processing (EMNLP)*, Doha, Qatar.
 21. Hawwari, Abdelati, Mohamed Attia, **Mona Diab**. (2014) A Framework for the Classification and Annotation of Multiword Expressions in Dialectal Arabic. In *Proceedings of Workshop on Arabic Natural Language Processing (ANLP), Empirical Methods For Natural Language Processing (EMNLP)*, Doha, Qatar.
 22. Solorio, Tamar, E. Blair, S. Maharjan, S. Bethard, M. Diab, M. Ghoneim, A. Hawwari, F. AlGhamdi, J. Hirschberg, A. Chang, P. Fung. (2014) Overview for the First Shared Task on Language Identification in Code Switched Data. In *Proceedings of First Workshop on Computational Approaches to Linguistic Code Switching (CodeSwitch), Empirical Methods For Natural Language Processing (EMNLP)*, Doha, Qatar.
 23. Zirikly, Ayah, **Mona Diab**. (2014) Named Entity Recognition for Dialectal Arabic. In *Proceedings of Workshop on Arabic Natural Language Processing (ANLP), Empirical Methods For Natural Language Processing (EMNLP)*, Doha, Qatar.
 24. 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 (ACL)*, Jeju Island, South Korea.
 25. Agirre, Eneko; Daniel Cer; **Mona Diab**; Aitor Gonzalez-Agirre. (2012) SemEval-12 Task 6: A Pilot on Semantic Textual Similarity. In *Proceedings of SEMEVAL Workshop, North American Association for Computational Linguistics (NAACL)*, Montreal, Canada.
 26. Guo, W. & Diab, Mona. (2012) A simple unsupervised latent semantics based approach for sentence similarity. In *Proceedings of the 6th International Workshop on Semantic Evaluation (SEMEVAL), North American Association for Computational Linguistics (NAACL)*, Montreal, Canada.
 27. Hawwari, Abdelati; Kfir Bar; **Mona Diab**. (2012) Building an Arabic Multiword Expressions Repository. In *Proceedings of Workshop on Computational Linguistics for Literature, North American Association for Computational Linguistics (NAACL)*, Montreal, Canada.
 28. 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 Workshop on Extra-propositional aspects of meaning in computational linguistics (ExProM)* Jeju Island, South Korea.
 29. Zaghouani, Wajdi, Abdelati Hawwari and **Mona Diab**. (2012) A Pilot PropBank Annotation

- for Quranic Arabic. *In Proceedings of the computational linguistics for literature workshop, North American Association for Computational Linguistics (NAACL)*, Montreal, Canada.
30. Abdul-Mageed, Muhammad and **Mona Diab**. (2011) Subjectivity and Sentiment Annotation of Modern Standard Arabic Newswire. *In Proceedings of workshop LAW V, Association for Computational Linguistics (ACL)*, Portland OR, USA.
 31. Dasigi, Pradeep and **Mona Diab**. (2011) Noisy Channel Modeling for Named Entity Transliteration. *In Proceedings of NEWS 11 Workshop, IJCNLP*, Chiang Mai, Thailand.
 32. **Diab, Mona** and Ankit Kamboj. (2011) An Annotation Scheme Leveraging Crowd Sourcing for Linguistic Code Switching in Hinglish. *In Proceedings of ALR9 Workshop, IJCNLP*, Chiang Mai, Thailand.
 33. Benajiba, Yassine and **Mona Diab**. (2010) A Web Application for Dialectal Arabic Text Annotation. *In Proceedings of the Workshop on Semitic Language Processing, Language Resources and Evaluation Conference (LREC)*, Malta.
 34. Guo, Weiwei and **Mona Diab**. (2010) COLEPL and COLSLM: An Unsupervised WSD Approach to Multilingual Lexical Substitution, Tasks 2 and 3 SemEval 10. *In Proceedings of SEMEVAL, Association for Computational Linguistics (ACL)*, Uppsala, Sweden.
 35. Zaghouni, Wajdi, **Mona Diab**, Aous Mansouri, Sameer Pradhan, Martha Palmer. (2010) The Revised Arabic PropBank. *In Proceedings of Linguistic Annotation Workshop. (LAW) IV, Association for Computational Linguistics (ACL)*, Uppsala, Sweden.
 36. **Diab, Mona**, and Pravin Bhutada. (2009) Verb Noun Construction MWE Supervised Token Classification. *In Proceedings of ACL-IJCNLP 09 Workshop on Multiword Expressions: Identification, Interpretation, Disambiguation and Applications, ACL-IJCNLP*, Singapore.
 37. **Diab, Mona**, Lori Levin, Teruko Mitamura, Owen Rambow, Vinodkumar Prabhakaran, Weiwei Guo. (2009) Committed Belief Annotation and Tagging. *In Proceedings of workshop on Third Linguistic Annotation Workshop (LAW III), ACL-IJCNLP*, Singapore.
 38. **Diab, Mona** and Madhav Krishna. (2009) Handling Sparsity for Unsupervised MWE Classification. *GEMS Workshop, European Association for Computational Linguistics (EACL)*, Athens, Greece.
 39. Guo, Weiwei, and **Mona Diab**. (2009) Improvements to Monolingual English Word Sense Disambiguation. *In Proceedings of the NAACL 09 workshop: Semantic Evaluations: Recent Achievements and Future Directions. (SEW-09), North American Association for Computational Linguistics (NAACL)*, Boulder CO, USA.
 40. BenAjiba, Yassine, **Mona Diab**, Paolo Rosso. (2008) Arabic Named Entity Recognition using Conditional Random Fields. *In 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 Conference (LREC)*, Marrakech, Morocco.
 41. Habash, Nizar, Owen Rambow, **Mona Diab**, Reem Farraj. (2008) Guidelines for Annotating Arabic Dialect, *Proceedings of Workshop on Arabic and its local languages, Language Resources and Evaluation Conference (LREC)*, Marrakech, Morocco.
 42. **Diab, Mona**. (2007) Improved Arabic Base Phrase Chunking with a new enriched POS tag set. *In Proceedings of workshop on Computational Approaches to Semitic Languages (CASL), Association for Computational Linguistics (ACL)*, Prague, Czech Republic.
 43. **Diab, Mona**, Musa Alkhalifa, Sabry ElKateb, Christiane Fellbaum, Aous Mansouri, Martha Palmer. (2007) SemEval-07 Task 18: Arabic Semantic Labeling. *In Proceedings of SEMEVAL, Association for Computational Linguistics (ACL)*, Prague, Czech Republic.
 44. **Diab, Mona**, Alessandro Moschitti, Daniele Pighin. (2007) CUNIT: A Semantic Role Labeling System for Modern Standard Arabic. *In Proceedings of SEMEVAL, Association for Computational Linguistics (ACL)*, Prague, Czech Republic.
 45. **Diab, Mona**. (2000) An Unsupervised Method for Word Sense Tagging using Parallel

Corpora: A Preliminary Investigation. *In Proceedings of Special Interest Group in Lexical Semantics (SIGLEX) Workshop, Association for Computational Linguistics (ACL), Hong Kong.*

TECHNICAL REPORTS

1. Habash, Nizar, **Mona Diab**, Owen Rambow. (2014) Conventional Orthography for Dialectal Arabic (CODA) Principles and Guidelines- Egyptian Arabic – Version 0.7- March 14. *Technical Report*, <http://dx.doi.org/10.7916/D83X8562>, Center for Computational Learning Systems. (CCLS), Columbia University
2. **Diab, Mona**, Nizar Habash, Owen Rambow, Ryan Roth. (2013) LDC Arabic Treebanks and Associated Corpora: Data Divisions Manual. *Technical Report*, <http://hdl.handle.net/10022/AC:P:21761>, Center for Computational Learning Systems. (CCLS), Columbia University
3. 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>, Center for Computational Learning Systems. (CCLS), Columbia University
4. 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*
5. **Diab, Mona**. (2000) Exploiting Translations for Semantic Annotation. *Candidacy Exam Thesis, Linguistics Department, University of Maryland, College Park MD, USA.*
6. 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-00-38/CS-TR-4147, University of Maryland College Park*
7. Resnik, Philip and **Mona Diab**. (2000) Measuring Verb Similarity. *Technical Report: LAMP-TR-047/ UMIACS-TR-00-40/CS-TR-4149, University of Maryland College Park*
8. **Diab, Mona** and Steve Finch. (2000) A Statistical Word-Level Translation Model for Comparable Corpora. *Technical Report: LAMP-TR-048/ UMIACS-TR-00-41/CS-TR-4150, University of Maryland College Park*
9. **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-00-39/CS-TR-4148, University of Maryland College Park*
10. **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*

TUTORIALS

1. **Diab, Mona** and Nizar Habash. (2006) Arabic Dialect Processing.
2. **Diab, Mona**. (2005) Statistical NLP: An A-Z Crash Course.

RELEASED SOFTWARE

ROBUST RELEASED TECHNOLOGY

1. [DIRA] Dialectal Arabic Information Retrieval Assistant. (2013)
Demo: <http://nlp.ldeo.columbia.edu/dira>
2. [AIDA] Automatic Identification of Dialectal Arabic. (2013)
Demo: <http://nlp.ldeo.columbia.edu/aida>
3. [MADAMIRA] A system for morphological analysis and disambiguation, base phrase chunking and named entity recognition. (2013)

Demo: <http://nlp.ldeo.columbia.edu/madamira>.

RESEARCH PROTOTYPE TECHNOLOGY

4. **Diab, Mona** and Pradeep Dasigi. (2011) **CODACT** A Tool for Arabic Dialect Orthographic Conventionalization. <https://nlp.ccls.columbia.edu/CODACT>
5. Benajiba, Yassine and **Mona Diab**. (2010) **NIDA** Arabic Named Entity Recognition. Current Version 2.0. <https://nlp.ccls.columbia.edu/NIDA>
6. **Diab, Mona**, Yassine Benajiba, Hebatullah EIFardy. (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>
7. Benajiba, Yassine, **Mona Diab**, Hebatullah EIFardy. (2011) **COLANN** COLABA Annotation Interface v.1.0: An annotation interface for Arabic Dialects. Current Version 1.0. <https://nlp.ccls.columbia.edu/COLANN>
8. **Diab, Mona** and Weiwei Guo. (2009) **SALAAM-CONT** Automatic Multilingual Online All words Word Sense Disambiguation. Current Version 1.0. <https://nlp.ccls.columbia.edu/COLANN>
9. Prabhakaran, Vinodkumar, **Mona Diab**, Owen Rambow. (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)

1. *Emotion, Intensity, and Urgency*. (2016) **DARPA LORELEI Site Visit**, Columbia University, New York NY, USA, Jul.
2. *Committed Belief Tagging*. (2016) **DARPA DEFT Site Visit**, Columbia University, New York NY, USA, Jul.
3. *Named Entity Linking & Event Detection*. (2016) **DARPA DEFT Site Visit**, Columbia University, New York NY, USA, Jul.
4. *Arabic Dialect Tools*. (2014) **DARPA BOLT Site Visit**, March, Columbia University, New York NY, USA, Mar.
5. *Perspective, Novelty and Belief in Streaming data and informal genres*. (2014) **DARPA DEFT Site visit**, George Washington University, Washington DC, USA, Feb.
6. *Arabic Dialect Processing under BOLT*. Invited Presentation. (2011) **DARPA BOLT Kick-off Meeting**, Tampa FL, USA, Oct.
7. *Arabic Language Modeling for OCR*. (With Nizar Habash) Invited Presentation. (2011) **DARPA MADCAT PI Meeting**. Boston MA, USA, Jul.
8. *Arabic Subjectivity and Sentiment Analysis: System & Resources*. Invited Poster. (2011) **DARPA BOLT Kick-off Meeting**, Tampa FL, USA, Oct.
9. *CADIM Arabic MSA NLP Tools*. (With Nizar Habash and Owen Rambow) Invited Poster, (2011) **DARPA BOLT Kick-off Meeting**, Tampa FL, USA, Oct.
10. *COLABA Tools and Resources*. (With Nizar Habash and Owen Rambow) Invited Presentation (2011) **DARPA BOLT Kick-off Meeting**, Tampa FL, USA, Oct.
11. *COLABA Tools and Resources*. (With Nizar Habash and Owen Rambow) Invited Presentation. (2011) **Brigham Young University**, Salt Lake City UT, USA, Aug 11.
12. *Conventionalized Dialectal Arabic. (CODA): Guidelines and Tools*. (With Nizar Habash and Owen Rambow) Invited Poster. (2011) **DARPA BOLT Kick-off Meeting**, Tampa FL, USA, Oct.
13. *Multilayered Semantic Annotation and challenges of Interoperability*. Invited Presentation. (2011) **NSF-funded SILT/FLaReNet Forum on Sustainable Interoperability for Language Technologies**. Brandeis University, Boston MA, USA, Apr.

14. *Power Shifts & Rifts in Social Communications*. (With Kathleen McKeown, Owen Rambow, Dragomir Radev) Invited Presentation. (2011) **IARPA SCIL PI Meeting**. Washington DC, USA, May.
15. *Power Shifts & Rifts in Social Communications*. (With Kathleen McKeown, Owen Rambow, Dragomir Radev) Invited Presentation. (2011) **IARPA SCIL PI Meeting**. Washington DC, USA, Nov 11.
16. *Arabic Dialect Processing within the Context of Social Media*. (With Nizar Habash and Owen Rambow) Invited Presentation. (2010) **COLABA Review Day, MITRE Corporation**. McLean VA, USA, Apr 10.
17. *Arabic Language Modeling for OCR*. (With Nizar Habash) Invited Presentation. (2010) **DARPA MADCAT PI Meeting**. Boston MA, USA, Sep.
18. *COLABA Annotation, Tools, Resources*. (With Nizar Habash and Owen Rambow) Invited Presentation. (2010) **Government Day AMTA 10**. Denver CO, USA, Oct.
19. *Power Shifts & Rifts*. (With Kathleen McKeown, Owen Rambow, Dragomir Radev) Invited Presentation. (2010) **IARPA SCIL PI Meeting**. Washington DC, USA, May.
20. *Arabic Language Modeling for OCR*. (With Nizar Habash) Invited Presentation. (2009) **DARPA MADCAT PI Meeting**. Boston MA, USA, Sep.
21. *Power Shifts & Rifts*. (With Kathleen McKeown, Owen Rambow, Dragomir Radev) Invited Presentation. (2009) **IARPA SCIL PI Meeting**. Washington DC, USA, Oct.
22. *CADIM: Natural Language Processing for Arabic and its dialects*. (With Nizar Habash and Owen Rambow) Invited Presentation. (2008) **CASL, University of Maryland**. College Park MD, USA, Jun.
23. *Columbia CCLS: Committed Belief and Dialog Acts*. (With Owen Rambow and Becky Passaneau) Invited Presentation. (2008) **Human Language Technology Center of Excellence, University of Maryland**. College Park MD, USA, Sep.
24. *Factivity Annotations*. (With Owen Rambow, Lori Levin, and Teriku Mitamura) Invited Presentation. (2008) **JHU Human Language Technology Center of Excellence, University of Maryland**. College Park MD, USA, May.
25. *SALAMAT: Sense Assignment Leveraging Alignments, Monolingual And Translation*. (With Weiwei Guo) Invited Demo/Poster. (2008) **NSF sponsored symposium on Semantic Knowledge Discovery, Organization and Use, NYU**. New York NY, USA, Nov.
26. *Tools for Arabic Morphological Analysis, Disambiguation and Tokenization*. (With Nizar Habash and Owen Rambow) Invited Presentation. (2008) **Technology Forum, DIA and TSWG**. McLean VA, Jul.
27. *Impact of Partial Arabic Diacritization on Statistical Machine Translation*. (With Mahmoud Ghoneim and Nizar Habash) Invited Presentation. (2007) **DARPA GALE PI Meeting**. San Francisco CA, USA, May.
28. *Semi-Automatic Error Analysis of the NIGHTINGALE Machine Translation System*. (With Katrin Kirchhoff, Nizar Habash, Owen Rambow, Evgeny Matusov) Invited Presentation. (2007) **DARPA GALE PI Meeting**. San Francisco CA, USA, May.
29. *Modern Standard Arabic Verb Alternations*. (With Neal Snider) (2006) Paper presented at the **Annual Meeting of the Arabic Linguistics Society. (ALS) Conference**. Kalamazoo MI, USA, Mar.
30. *NLP Tools for Arabic*. (With Nizar Habash and Owen Rambow) Invited Presentation (2006) **DARPA GALE PI Meeting**. Boston MA, USA, May.
31. *Treebank Transduction. Arabic Dialect Parsing*. (With Owen Rambow, Nizar Habash, Khalil Simaan, David Chiang, Rebecca Hwa, Roger Levy, Vincent Lacey, Saif-Allah Shareef) (2005) **Johns Hopkins Summer Workshop Final Presentation Symposium**. Baltimore MD, USA, Aug.

32. *An Unsupervised method for Word Sense Tagging using Parallel corpora.* (2001) Poster Session: **UMIACS Research Day, University of Maryland.** College Park MD, USA, Mar.
33. *A Statistical Model of Word-Level Mapping for Comparable Corpora.* (2000) **UMIACS Computational Linguistics Colloquium, University of Maryland.** College Park MD, USA, Spr.
34. *An Unsupervised Method for Word Sense Tagging using Parallel corpora.* (2000) **North East Scholars NLP workshop: IBM T.J. Watson Research Center,** Hawthorne NY, USA, Oct.
35. *MADLIBS: An LCS based Translingual Information Retrieval System.* (With Maria Katsova) (2000) **Demonstration: NAACL2000.** Seattle WA, USA, May.
36. *Statistical Methods and Linguistics.* (With Elixabete Murguia) (2000) **LingFest, Linguistics Department University of Maryland.** College Park MD, USA, Nov.
37. *Cross Language Information Retrieval: An LCS based approach for Spanish queries and English Documents.* (With Maria Katsova) (1999) **Poster Session: UMIACS Research Day, University of Maryland.** College Park MD, USA, Feb.
38. *Word Sense Disambiguation: Lexical Ambiguity Resolution within a Selection Preference Model.* (1999) **Linguistics Student Conference, Linguistics Dept., University of Maryland.** College Park MD, USA, Fall.
39. *A Preliminary Statistical Investigation into the Impact of Using an N-Gram Analysis for the Purposes of Text Author Identification/Classification.* (1997) **Linguistics Student Conference, Linguistics Dept., University of Maryland.** College Park MD, USA. Fall.

PROFESSIONAL ACTIVITIES

ACADEMIC SERVICES WITHIN GWU

- **Member**, *Dept. Committee Grad. Admissions & Support. (GAS)*, Fall 14- present
- **Member**, *Dept. Committee on Awards*, Fall 13 –present
- **Member**, *Dept. Promotion & Tenure Committee for Prof. Tim Wood*, Sum 16 -present
- **Member**, *Steering Committee overall IBM collaboration and Point Person for any work or education related to Watson*, Fall 14 -- present
- **Elected Senator**, representing *SEAS on Faculty Senate*, Spr 14 – Fall 15
- **Co-chair**, *Big Data Initiative Task Force*, Fall 13 – Spr 14
- **Co-chair**, *Big Data Chair Search Committee*, Fall 14-Spr 15
- **Member**, *Dept. Committee on Budget*, Fall 14- Spr 15
- **Member**, *Dept. Committee on Student Relations*, Fall 14- Spr 15
- **Member**, *Dept. Committee on Research*, Fall 13 – Spr 14
- **Member**, *Dept. Committee on Graduate Student Admission*, Spr 13 – Spr 14.
- **Member**, *Dept. Committee on Curriculum*, Spr 13
- **Member**, *Dept. Committee on Faculty Retreat*, Spr 13
- **Member**, *Dept. Committee on Target of Opportunity Hires*, Spr 13

ACADEMIC SERVICES WITHIN COLUMBIA UNIVERSITY

- **Member** of *Columbia Senate Committee for Research Officers*, Fall 05–Fall 12
 - 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– Fall 12
- **Member** of the *Columbia Senate Commission on the Status of Women*, Spr 10–Fall 12

- **Member** of *SEAS Dean's Committee for Research Scientists*, Fall 09– Fall 12
- **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– Fall 13
- **Organizer** of *CCLS-Yahoo Distinguished Lecture Series*, Spr 10– Fall 12
- **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–Fall 12
- **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

- Guest Editor of Special Issue on Georgetown Round Table on Arabic Computational Linguistics Conference. **Oct 09 – Mar 10**
ACM Transactions on Asian Language Information Processing

Grant Proposal Reviewer

- *NSF RI CI Panel* **Dec 15**
- *NSF Semantics Panel* **Mar 13**
- *KACST Research Fund* **Jan 12 –Fall 15**
- *NSF OISE Program* **Feb 12**
- *Qatar National Research Fund* **Spr 10 – Sep 12**
- *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

- **Co-Chair** *2nd Workshop on Computational Approaches to Linguistic Code Switching (CALCS), Proceedings of EMNLP 2016, Austin TX, USA* **Nov 16**
- **Chair** *Workshop on Machine Translation for Semitic Languages (SeMAT), AMTA 2016, Austin TX, USA* **Nov 16**
- **Co-Chair** *Semeval 16 Task 1, Semantic Textual Similarity, NAACL 16, San Diego CA, USA.* **Fall 15- Spr 16**
- **Co-Chair** *Semeval 15 Task 2, Semantic Textual Similarity, NAACL 15, Boulder CO, USA.* **Fall 14- Spr 15**
- **Co-Chair**, *Town Hall Meeting for Big Data efforts at GWU, Washington DC, USA* **Spr 14**
Fall 14
- **Co-Chair**, *First Workshop on Computational Approaches to Linguistic Code Switching, Empirical Methods For Natural Language Processing (EMNLP), Doha, Qatar.* **Spr – Fall 14**
- **Publicity Co-Chair**, *Empirical Methods For Natural Language Processing (EMNLP), Doha, Qatar.* **Spr -- Sum 14**
- **Co-Chair** *Semeval 14 Task 10, Multilingual Semantic Textual Similarity, International Conference of Computational Linguistics (COLING), Dublin, Ireland.* **Sum 14**
- **General Chair**, **SEM 2nd Joint Semantics Conference* **Spr 13**
- **Chair**, *DARPA Funded Unified Semantic Framework for Textual* **Feb 12**

- Similarity Workshop*
- **Area Co-Chair** for ACL 12 Multilinguality **Aug 11 –Jul 12**
 - **Program Co-Chair** of 1st International Joint Conference on Semantics *SEM 12 **Aug 11 –Jun 12**
 - **Workshops Co-Chair**, NAACL 12 **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 Empirical Methods For Natural Language Processing (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 Task on Arabic Semantic Labeling **Jun 07**
 - **Co-Chair** ACL Workshop on Computational Approaches to Semitic Languages **Jun 05**

PROGRAM COMMITTEE MEMBER

Journals

Transactions on Computational Linguistics; 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 11); Text Graphs (ACL 11); LAWS V. (ACL 11); LREC Workshop on HLT for Semitic Languages (LREC 10); Statistical Machine Translation. (ACL 08);TextGraphs. (COLING 08); Student Research Workshop (ACL 08); Arabic and Local Languages (LREC 08); Computational Approaches to Semitic Languages (ACL 07); SEMEVAL 07 (ACL 07); Computational Approaches to Arabic Script based Languages (2007); Student Research Workshop (HLT-NAACL 07); Cross Language knowledge Induction (EACL 06); Student Research Workshop (ACL 05); Parallel Text (ACL 05); Computational Approaches to Semitic Languages (ACL 05); Cross-Language Knowledge Induction Workshop (EUROLAN 05); Computational Approaches to Arabic Script based Languages (COLING 04)

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)