Mona Diab

Center for Computational Learning Systems (CCLS)
Columbia University
Email: mdiab@cs.columbia.edu
Office: (212)8701290

Education

- **Ph.D. Linguistics (Computational Linguistics)**   Aug. '03
  University of Maryland, College Park
  Title: Word Sense Disambiguation within a multilingual framework
  Advisor: Philip Resnik
- **M.Sc. Computer Science (Artificial Intelligence - Machine Learning)**   Jun. '97
  The George Washington University
- **B.Sc. Computer Science & Software Engineering**   Jun. '92
  American University in Cairo
- **B.Sc. Tourist Guidance (Egyptology & Archaeology)**   Jun. '90
  Faculty of Tourism and Hotel Management, Helwan University

Work Experience

- **Associate Research Scientist (Columbia University)**   July'05-present

  Working on semantic role parsing of Modern Standard Arabic (MSA)
  [Collaboration with Dan Jurafsky, Stanford University].
  Working on syntactic parsing of MSA
  [Collaboration with Nizar Habash & Owen Rambow, Columbia University]
  Working on tools for processing MSA.
  [Collaboration with Kadri Hacioglu, University of Colorado, Boulder]
  Developing standards for the processing of Arabic dialects.
  [Collaboration with Nizar Habash & Owen Rambow, Columbia University]
  Building Arabic Propbank
  [Collaboration with Martha Palmer, University of Colorado, Boulder]
  Building an Arabic WordNet
  [Collaboration with Christiane Fellbaum, Princeton University]
  Working on WordNet clustering.
  [Collaboration with Dan Jurafsky & Andrew Ng, Stanford University]
  Working on Arabic verb subcategorization frame induction
  [Collaboration with Neal Snider, Stanford University]
  Extending work on building lexica using comparable corpora.
  [Collaboration with Vincent Lacey, Georgia Tech]

- **Postdoctoral Research Scientist (*Columbia University*)**   Feb. '05-Jun'05

    Working on tools for processing MSA as well as developing standards for the processing of Arabic dialects.

- **Postdoctoral Scholar (*Stanford University*)**   Jan. '04-Jan. '05

    Working with Dan Jurafsky on domain and language independent semantic parsing and disambiguation. Working with Chris Manning on Arabic syntactic parsing. In general, working on developing technologies, tools and resources for Arabic language processing.

- **Research Associate (*Univ. of Colorado at Boulder, CSLR*)**   Jul. '03 – Dec. '03

    Working with Dan Jurafsky, Jim Martin and Wayne Ward on Arabic semantic parsing, bootstrapping tools and resources for Arabic language processing

- **Grad. Research Assist. (*Univ. of Maryland College Park, UMIACS*)**   Jul. '97 – Jun. '03

    Working with Philip Resnik on multilingual word sense disambiguation, verb clustering, measuring verb similarity, information theoretic similarity measures for Lexical Conceptual Structures, normalizing the text of the Bible in 17 different languages; working with Doug Oard on Arabic cross language information retrieval

- **Graduate Teaching Assist. (*George Washington University*)**   Fall '95-Spring. '97

    Taught a course *"C as a second language"* to junior and senior students in CS and Engineering depts. Duties included choosing course material, creating exams and assignments and grading.

- **Graduate Teaching Assist. (*American University in Cairo*)**   Jan. '93 –Jun. '95

    Assisted professors in teaching Theory of Computing, Formal Languages and Artificial Intelligence courses. Duties included: Conducting Q&A sessions, holding problem sessions, running lab sessions, grading exams, quizzes and assignments.

Miscellaneous Positions

o   Visiting Researcher *(Université de Rouen, Lab. PSI)*
o   Research Assistant *(Thomson Publishing NLP Labs)*
o   Systems Analyst *(Solutions Consultants International)*
o   Software Engineer *(Reg. Info. Tech. Software Engg. Center)*
o   English News Broadcast Anchor *(Saudi National Radio)*
o   Qualified Tourist Guide *(FreeLance)*

Sept.'02--Dec.'02
May '98--Aug.'98
Mar.'97-- Aug.'97
Apr.'93 – Feb. '95
Sept.'92—Jan.'93
Jun.'91 – Aug.'95

Research Interests

As a bilingual speaker with a flair for new languages, I have devised solutions to several of the challenges I faced in computational linguistics from a multilingual perspective. I have a deep interest in cross language typology and the extent to which knowledge about language similarity and divergence may help in understanding and formulating interesting research questions about natural language and its automatic processing. My research interests span several areas and the interdisciplinary interactions between them: computational linguistics, lexical semantics, and psycholinguistics. Being a computational linguist with a background in machine learning, I am very interested in the extent to which both unsupervised and reinforcement learning methods can be utilized for natural language applications; how do such models help shed light on our linguistic cognitive system. In the scope of lexical semantics, I am very interested in issues of relating word representations (structure) to word meaning (senses). Moreover, I like to explore issues of ontological representation and concept organization, in particular, with regard to underspecification in as much as it accounts for the creative use of language and language change. Finally, in the scope of psycholinguistics, I am interested in building predictive computational models informed by and highly correlated with human judgments and behaviours.

Publications

Journal Articles

- Diab, Mona. "**Sense Assignment Leveraging Alignment And Multilinguality: SALAAM**". *in preparation*.

- Diab, Mona. "**An Optimality Theoretic Comparative View of Stress in Three Arabic Dialects: Cairene Arabic, Jordanian and Saudi Bedouin Arabic**." *in preparation*.

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

Refereed Conference Papers

- Diab, Mona. "**Relieving the Data Acquisition Bottleneck for Word Sense Disambiguation**." *Proceedings of Association for Computational Linguistics, 2004.*

- Diab, Mona, Kadri Hacioglu and Daniel Jurafsky. "**Automatic Tagging of Arabic Text: From Raw Text to Base Phrase Chunks**." *Proceedings of Human Language Technology-North American Association for Computational Linguistics, 2004.*

- Diab, Mona. "**An Unsupervised Approach for Bootstrapping Arabic Sense Tagging**." *Proceedings of Arabic Workshop, Computational Linguistics (COLING), 2004.*

- Diab, Mona. "**Bootstrapping a WordNet Taxonomy for Arabic**", *Proceedings of First Arabic Language Technology Conference (NEMLAR), Cairo Egypt, Sept. 2004.*

- Diab, Mona and Philip Resnik, *"**An Unsupervised Method for Word Sense Tagging using Parallel Corpora**", Proceedings of Association for Computational Linguistics*, 2002.

- Diab, Mona. **"An Unsupervised Method for Word Sense Tagging using Parallel Corpora: A Preliminary Investigation***", Special Interest Group in Lexical Semantics (SIGLEX) Workshop, Association for Computational Linguistics**,* 2000.

- Diab, Mona and Steven Finch. *"**A Statistical Word-Level Translation Model for Comparable Corpora***", Proceeding of Conference on Content-based Multimedia Information Access (RIAO2000)*, 2000.

- Resnik, Philip and Mona Diab, *"**Measuring Verb Similarity***", Proceedings of the conference of International Cognitive Science Society (COGSCI2000)*, 2000.

- Dorr, Bonnie, Gina Levow, Douglas Oard, Philip Resnik, Amy Weinberg, Mona Diab, Maria Katsova. "**MADLIBS: An Event Translingual Lexical Conceptual Structure Based Information Retrieval System**", *Proceedings of North American Association for Computational Linguistics*, 2000.

- Diab, Mona, John Schuster and Peter Bock. *"**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*, Egypt 1998.

- Riopka, Terry, Mona Diab and Peter Bock. *"**Quantifying and Interpreting the Effect of Intelligent Information***", Proceedings of 6th International Conference on Artificial Intelligence & Applications*, Egypt 1998.

Invited Talks

- Computer Science Department, New York University, USA — Oct. '05
- Center for Computational Learning Systems, Columbia University, USA — Apr. '05
- Computer Science Department, Columbia University, USA — Mar. '05
- Dept. of Linguistics, Ohio State University, Columbus Ohio, USA — Feb. '05
- Dept. of Linguistics, UT Austin, Austin, Texas, USA — Feb. '05
- Guest lecturer of Intro. To Computational Linguistics, UT Austin, Texas, USA — Feb. '05
- Yahoo research Labs, Overture, Pasadena, California, USA — Jan. '05
- Language Computer Corporation, LCC, Dallas, Texas, USA — Dec. '04
- Microsoft Research, Natural Language Group, Redmond, Washington, USA — Dec. '04
- Department of Informatics, University of Edinburgh, Scotland — Aug. '04
- Linguistic Data Consortium, University of Pennsylvania, USA — May '04
- IBM T. J. Watson Research Center, New York — May '04
- Linguistics Dept. & Institute for Cognitive Science, Univ. of Colorado, Boulder — Fall '03
- Université de Rouen, PSI Lab (France) — Fall '02
- University of Southern California, Information Sciences Institute — Fall '00
- IBM T.J. Watson Research Center, New York — Fall '00

Special Awards

- *Invited Speaker: Meaning Workshop, European Union, IRST, Italy* — *Feb. '05*
- *Senior Research member for Johns Hopkins University Summer Workshop on Arabic Dialect Parsing, Awarded by NSF* — *Jul.'05- Aug.'05*

Language Skills

Arabic, English: native proficiency
French: very good
Hebrew, Italian, Japanese: basic knowledge

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)

Professional Activities

Panelist

NSF Career Awards Fall 2005
ACL 2005 Student Research Panelist (Reviewed papers as well as gave public feedback to students on their research)

Paper Reviewer

Journals:
Computational Linguistics,
IEEE Speech Communication,
Computers and Humanities,
Computer Speech and Language

Conferences:
Association for Computational Linguistics (ACL),
Cognitive Science Society,
Computational Linguistics (COLING),
NEMLAR
Human Language technologies (HLT)
Empirical Methods in Natural Language Processing (EMNLP)

Workshops:
ACL 05 Parallel Text workshop
ACL 05 Cross Linguistic data induction

Organizer

Co-organized ACL 2005 Workshop on Computational Approaches to Semitic Language Processing

Committee Member

Columbia University Research Officer Committee
Columbia University Committee on Physical Development (Representing Research Officers)
Columbia University Task Force on Physical Development (Representing Research Officers)

References

- Philip Resnik, University of Maryland College Park, (resnik@umiacs.umd.edu)
- Daniel Jurafsky, Stanford University, (jurafsky@stanford.edu)
- Amy Weinberg, University of Maryland College Park, (weinberg@umiacs.umd.edu)
- Bonnie Dorr, University of Maryland College Park, (bonnie@umiacs.umd.edu)
- Owen Rambow, Columbia University, (rambow@cs.columbia.edu)